Mathematical Engineering

Andreas Rauh
Luise Senkel *Editors*

Variable-Structure Approaches

Analysis, Simulation, Robust Control and Estimation of Uncertain Dynamic Processes

Springer

Mathematical Engineering

Series editors

Claus Hillermeier, Neubiberg, Germany
Jörg Schröder, Essen, Germany
Bernhard Weigand, Stuttgart, Germany

More information about this series at http://www.springer.com/series/8445

Andreas Rauh · Luise Senkel
Editors

Variable-Structure Approaches

Analysis, Simulation, Robust Control and Estimation of Uncertain Dynamic Processes

Springer

Editors
Andreas Rauh
Universität Rostock
Rostock, Mecklenburg-Vorpommern
Germany

Luise Senkel
Universität Rostock
Rostock, Mecklenburg-Vorpommern
Germany

ISSN 2192-4732 ISSN 2192-4740 (electronic)
Mathematical Engineering
ISBN 978-3-319-31537-9 ISBN 978-3-319-31539-3 (eBook)
DOI 10.1007/978-3-319-31539-3

Library of Congress Control Number: 2016936579

© Springer International Publishing Switzerland 2016
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

In recent years, numerous variable-structure approaches have been developed for the control of nonlinear dynamic systems and for the model-based estimation of non-measurable states and parameters. These approaches typically make use of first-order as well as higher order sliding mode techniques and related procedures that are characterized by a variable-structure nature. One of their main advantages is the inherent proof of asymptotic stability. This stability proof is either performed offline during the corresponding controller and estimator design or online by the real-time evaluation of a suitable candidate for a Lyapunov function.

The methodological framework for variable-structure control and estimation approaches is quite well developed in the case of continuous-time and discrete-time systems, for which process models are accurately known.

Nevertheless, research efforts are still necessary to make the corresponding procedures applicable when only worst-case bounds are available for specific parameters. This type of uncertainty is often caused by non-negligible, however, inevitable manufacturing tolerances. Moreover, significant stochastic disturbances—for example, as a result of measurement noise—may act as further system inputs in many practically relevant applications. To enhance robustness in such cases, it is possible to combine variable-structure approaches with techniques which are for instance based on interval analysis, stochastic differential equations, or linear matrix inequalities.

This book aims at presenting current research activities in the field of robust variable-structure systems. The scope equally consists in highlighting novel methodological aspects as well as in presenting the use of variable-structure techniques in industrial applications including their implementation on hardware for real-time control.

Besides variable-structure approaches for the design of feedback control strategies and state estimation procedures, computational techniques for simulation—as included in predictive controllers—robustness and stability analysis, as well as for the identification of system models which are characterized by an inherent variable-structure behavior are included. Such models may result from a mathematical representation of state-dependent transitions between various state-space

representations, for example, due to faults of selected system components or due to different system models depending on the current operating conditions of the considered system.

Rostock
January 2016

Andreas Rauh
Luise Senkel

Acknowledgments

The editors of this special volume on *Variable-Structure Approaches for Analysis, Simulation, Robust Control and Estimation of Uncertain Dynamic Processes* would like to thank all authors for their high-quality contributions. Moreover, we would like to thank all researchers who supported us by reviewing the individual book chapters and by providing invaluable comments and suggestions for improvements. Finally, we would like to thank Prof. Dr. Claus Hillermeier (University of the Federal Armed Forces, Munich, Germany) and Dr. Jan-Philip Schmidt (Springer–Verlag) for supporting our idea to publish this collection of current research articles in the book series *Mathematical Engineering*.

Contents

Part I
Sliding Mode Control for Continuous and Discrete-Time Systems

In the first part of this book, scientific works are presented which focus on solving control tasks for nonlinear and uncertain dynamic systems by the application as well as the novel development of sliding mode control techniques. In Chap. 1, Saif Siddique Butt, Hao Sun, and Harald Aschemann describe a comparison of backstepping-based sliding mode techniques and adaptive backstepping approaches for the design of robust controllers of a twin rotor helicopter. In contrast to the continuous-time design in the first chapter, Chap. 2 authored by Piotr Leśniewski and Andrzej Bartoszewicz deals with the robust, discrete-time congestion controller design for a single virtual circuit in connection-oriented communication networks. Andreas Rauh and Luise Senkel develop and implement novel interval methods for the robust sliding mode control synthesis of high-temperature fuel cells. Besides the influence of uncertain but bounded parameters and disturbances, the handling of state and input constraints is explicitly addressed in this Chap. 3. Further methodological extensions of interval-based sliding mode controllers are presented in Chap. 4 by Luise Senkel, Andreas Rauh, and Harald Aschemann. It deals with the experimental and numerical validation of a reliable sliding mode control strategy considering uncertainty with the help of interval arithmetic in a real-time capable implementation.

Comparison of Backstepping-Based Sliding Mode and Adaptive Backstepping for a Robust Control of a Twin Rotor Helicopter

Saif Siddique Butt, Hao Sun and Harald Aschemann

Abstract In this contribution, two robust MIMO backstepping control approaches for a twin rotor aerodynamic system (TRAS) test-rig are considered. The TRAS represents a nonlinear system with significant couplings. A nonlinear multibody model of the TRAS with lumped unknown disturbance torques is derived using Lagrange's equations. Herewith, both a backstepping-based sliding mode control and an adaptive backstepping control are designed to track desired trajectories for the azimuth angle and the pitch angle. An explicit expression is derived for the reaching time in the case of the backstepping-based sliding mode control. In order to estimate immeasurable angular velocities and unknown disturbance torques for the backstepping-based sliding mode control, a discrete-time extended Kalman filter (EKF) is employed. For the adaptive backstepping, a robust sliding mode differentiator is used instead to estimate the angular velocities. Moreover, in the adaptive backstepping control approach, the disturbance compensation is realised with the help of additional adaptive control parts driven by the tracking errors of the controlled variables. The overall stability of the proposed controllers in combination with the corresponding estimator is investigated thoroughly by simulations. Furthermore, in order to validate the proposed control schemes, experiments are performed on the dedicated test-rig and a comparison of the two proposed control structures is provided as well.

S.S. Butt · H. Sun (✉) · H. Aschemann
Chair of Mechatronics, University of Rostock, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany
e-mail: hao.sun@uni-rostock.de

S.S. Butt
e-mail: saif.butt@uni-rostock.de

H. Aschemann
e-mail: harald.aschemann@uni-rostock.de

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*,
Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_1

1 Introduction

A twin rotor aerodynamic system (TRAS) with two degrees of freedom (DOF) developed by INTECO [10], Poland, is depicted in Fig. 1. It consists of a beam with two propellers—the main and the tail propellers—at both ends of the beam, driven by DC motors. The beam is pivoted in a cardanic joint and can rotate within confined regions in the horizontal and vertical planes. These limitations arise due to the mechanical construction of the test-rig. Two levers with counter-weights at their ends are fixed to the beam at the pivot. The counter-weights determine the steady-state pitch angle without propeller actuation. Two velocity sensors are coupled with the PWM-driven DC motors for the main and tail rotors. Moreover, two incremental encoders are directly mounted at the pivot point of the beam in order to measure the relative angle of the beam. Based on the mechanical construction, the TRAS test-rig possesses 2-DOF. The first DOF characterises the horizontal rotation of the frame using the azimuth angle, whereas the second one is given by the pitch angle describing the inclination of the frame. In a real helicopter the aerodynamic force is adjusted by changing the angle of attack. The TRAS, however, uses a changing angular velocity for this purpose.

The 2-DOF helicopter system imposes challenging control problems due to its given nonlinearities and significant couplings between both degrees of freedom. To remedy such drawbacks and achieve satisfactory control performance for the accurate tracking of desired trajectories for the azimuth and pitch angles, a control-oriented model of the system is useful. In the past decade, several contributions related to the modelling and experimental identification of similar 2-DOF helicopter set-ups have been published [1, 2, 7, 18, 19]. The methods proposed therein correspond to typical set-ups provided by different manufacturers, various model-based and

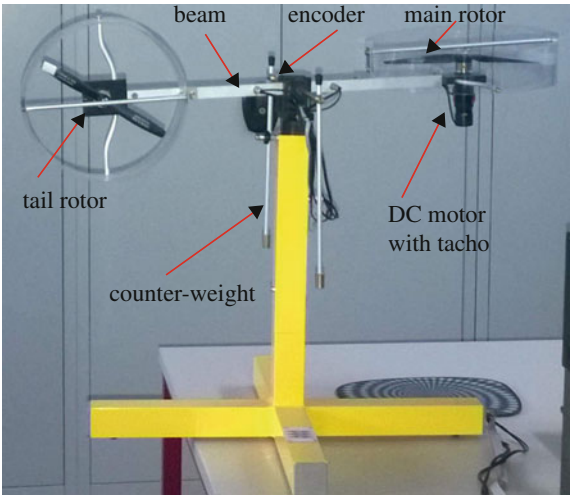


Fig. 1 TRAS test-rig at the Chair of Mechatronics

artificial-intelligence-based methods, e.g. radial basis function, neural networks and genetic algorithms. In [18], a complete mathematical description of the TRAS based on the both the Newton–Euler approach and Lagrange's equations is presented. However, the mechanical construction of the test set-up therein differs from the TRAS test-rig used in this contribution. A complete mathematical description of the TRAS set-up is derived in the form of a detailed ninth-order model in the work of [8]. Regarding the control of the TRAS, the authors performed a simulation study using a nonlinear predictive control for the ninth-order system model. In [16], a feedback linearising control scheme is presented for the pitch motion only. Therein, the yaw position is not considered as a DOF. Lopez et al. [15, 17] proposed an H_∞-controller for the helicopter dynamics. In [4], a nonlinear control-oriented model of the TRAS manufactured by INTECO along with a multi-variable flatness-based control scheme has been proposed for TRAS. However, the influence of the mass moments of inertia of the rotating beam on the kinetic energy of the system has not been addressed. For the estimation of disturbance torques and unmeasured states, the authors employed a discrete-time EKF. In [3, 5], a sliding mode and an integral sliding mode control are considered, respectively, allowing to deal with couplings inherent in the twin rotor dynamics. Following the idea of a cascaded control structure, it is suitable to employ backstepping techniques for the control design. Both the design and the implementation of the separate controllers are simplified in comparison to a central control structure. Disturbances are taken into account by an adaptive backstepping control approach. Although an adaptive control scheme based on backstepping control is already proposed in [11], the authors focus especially on the derivation of adaptive control laws for a set of uncertain parameters within the pitch and the azimuth dynamics. The adaptive backstepping control design guarantees global stabilisation. Nevertheless, due to the large number of adaptive laws, the convergence of the parameter estimates to their true values cannot be guaranteed. This may lead to unrealistic values of the parameters [20]. One possible way to handle this situation is to introduce a lumped disturbance term that represents the overall parameter uncertainty and to design a parameter update law for this lumped disturbance. Therefore, in the given contribution, the parameter uncertainty and the model uncertainty are combined together as a lumped disturbance torque for each axis. As a result, the adaptive laws are needed only for these lumped disturbance torques rather than for the individual parameter uncertainty.

As mentioned earlier, the TRAS system is affected by parameter uncertainty due to the limited knowledge of the true parameter values and unknown disturbances. Hence, a robust way of controlling such a system is the application of variable structure control techniques [23]. For this purpose, a backstepping-based sliding mode control scheme is proposed and investigated as well. Although many significant contributions are already available for the control design of the TRAS, a lot of questions are worth further investigations. This in particular includes the analysis of a damping term in the sliding mode control law concerning the reaching time. One of the main contributions regarding the backstepping-based sliding mode control is the derivation of the corresponding reaching time which—according to the best of the authors knowledge—has not been addressed in previous works.

This chapter is structured as follows: in Sect. 2, the mathematical description of the nonlinear control-oriented model of the TRAS based on Lagrange's equations is presented. Based on the derived system model, two alternative control strategies are described in Sect. 3: a backstepping-based sliding mode control and an adaptive backstepping control. For each control strategy, the asymptotic stability is shown by Lyapunov techniques. Moreover, the overall closed-loop error dynamics for each control scheme emphasise the asymptotic stability. In Sect. 4, a discrete-time EKF is designed to estimate the angular velocities as well as the unknown lumped disturbance torques, which are required for the backstepping-based sliding mode control structure. For the adaptive backstepping control structure, a robust sliding mode differentiator is proposed that provides estimates for the angular velocities. The implementation of the proposed two control strategies together with their corresponding estimators are explained as well in this section. In Sect. 5, relevant simulation results and experimental investigations from the TRAS test-rig are presented and discussed. Finally, conclusions and outlook are provided in Sect. 6.

2 Control-Oriented Model of TRAS

Dynamic system modelling plays a key role in modern control engineering. For a model-based control design, an accurate mathematical description of the system dynamics is essential to improve the overall system performance. For the control-oriented modelling of the TRAS an inertial reference frame is defined and the dynamic model is derived using Lagrange's equations.

The multibody system model for the TRAS consists of a beam, point masses for the two rotors and point masses for the two counterbalances. The mathematical analysis starts with assigning a right-hand coordinate system with the origin O placed at the pivot point of the beam as depicted in Fig. 2. The pitch angle is denoted by ψ and the azimuth angle is given by ϕ. The distance of the main rotor from the origin O is characterised by the length l_m, whereas the distance of the tail rotor from the origin O is given by l_t. The lumped masses corresponding to the main rotor and the tail rotor are denoted by m_m and m_t, respectively. Moreover, l_{cw} represents the relevant length of the two levers with a lumped mass m_{cw} as counter-weight at their end. The propulsive forces acting on the main rotor and the tail rotor are denoted by F_ψ and F_ϕ, respectively. The corresponding position vectors $\mathbf{r}_m$ for the main rotor and $\mathbf{r}_t$ for the tail rotor are given by

$$\mathbf{r}_m = \begin{bmatrix} l_m \cos\psi \cos\phi \\ l_m \cos\psi \sin\phi \\ l_m \sin\psi \end{bmatrix}, \quad \mathbf{r}_t = \begin{bmatrix} -l_t \cos\psi \cos\phi \\ -l_t \cos\psi \sin\phi \\ -l_t \sin\psi \end{bmatrix}. \tag{1}$$

Likewise, the position vectors $\mathbf{r}_{cw1}$ and $\mathbf{r}_{cw2}$ for the counter-weights result in

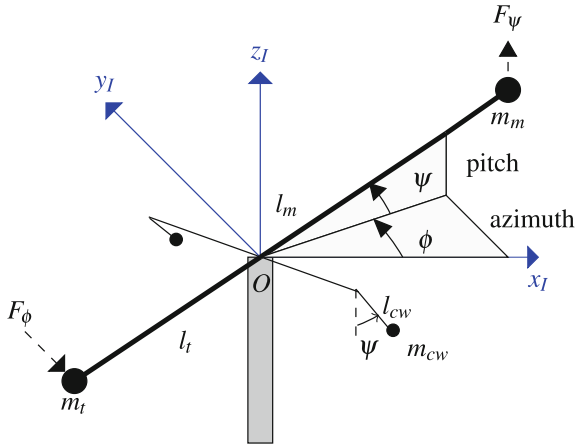


Fig. 2 Free body diagram of TRAS

$$\mathbf{r}_{cw1} = \begin{bmatrix} -l_{cw}\sin\psi\sin\phi \\ -l_{cw}\sin\psi\cos\phi \\ -l_{cw}\cos\psi \end{bmatrix}, \ \mathbf{r}_{cw2} = \begin{bmatrix} l_{cw}\sin\psi\sin\phi \\ l_{cw}\sin\psi\cos\phi \\ -l_{cw}\cos\psi \end{bmatrix}. \tag{2}$$

The overall kinetic energy T of TRAS is determined with the help of velocity vectors as well as the mass moments of inertia of the rotating beam. The kinetic energy of the TRAS in terms of the generalised coordinates is expressed as

$$\begin{aligned} T &= \frac{m_m}{2}\dot{\mathbf{r}}_m^2 + \frac{m_t}{2}\dot{\mathbf{r}}_t^2 + \frac{m_{cw}}{2}(\dot{\mathbf{r}}_{cw1}^2 + \dot{\mathbf{r}}_{cw2}^2) + \frac{1}{2}J_z\dot{\phi}^2 + \frac{1}{2}J_x\dot{\psi}^2 \\ &= \frac{1}{2}\Big((m_m l_m^2 + m_t l_t^2)(\dot{\phi}^2\cos^2\psi + \dot{\psi}^2) + m_{cw}l_{cw}^2(\dot{\phi}^2\sin^2\psi + \dot{\psi}^2) + J_z\dot{\phi}^2 + J_x\dot{\psi}^2\Big), \end{aligned} \tag{3}$$

where the square of the velocity components along the corresponding axes are calculated using the standard vector dot product identity, i.e. $\dot{\mathbf{r}}^2 = \dot{\mathbf{r}}^T \cdot \dot{\mathbf{r}}$. The parameters J_x and J_z denote the moment of inertias of the rotating beam w.r.t. the body-fixed x-axis and z-axis, respectively. Similarly, the overall potential energy U of the point mass system is given as

$$U = g\Big((m_m l_m - m_t l_t)\sin(\psi) + 2m_{cw}l_{cw}(1-\cos(\psi))\Big) + \frac{1}{2}k_\phi\,\phi^2. \tag{4}$$

Here, g is the gravitational constant. The parameter k_ϕ is used to model the restoring energy due to the elasticity in the cable. The Lagrangian of the system is defined as the difference between the kinetic energy and the potential energy, i.e.

$$L = T - U.$$

Finally, Lagrange's equations including non-conservative forces result in [9]

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = \tau_q - \frac{\partial R}{\partial \dot{q}}, \quad q = \{\phi, \psi\}, \tag{5}$$

where τ_ϕ and τ_ψ are torques along the azimuth and the pitch axes, respectively. The Rayleigh dissipation function R is given by [9]

$$R = \frac{1}{2}\, c_\phi\, \dot{\phi}^2 + \frac{1}{2}\, c_\psi\, \dot{\psi}^2. \tag{6}$$

Substituting (3), (4) and (6) into (5), the equations of motion for the system become

$$J_\phi \ddot{\phi} = \tau_\phi + J\dot{\phi}\dot{\psi}\sin(2\psi) - k_\phi \phi - c_\phi \dot{\phi}, \tag{7}$$

$$J_\psi \ddot{\psi} = \tau_\psi - g\Big[(m_m l_m - m_t l_t)\cos\psi + 2m_{cw} l_{cw}\sin\psi\Big] - J\frac{\dot{\phi}^2}{2}\sin(2\psi) - c_\psi \dot{\psi}, \tag{8}$$

with the following definition of the mass moments of inertia

$$J_\phi = J_\phi(\psi) = (m_m l_m^2 + m_t l_t^2)\cos^2\psi + 2m_{cw} l_{cw}^2 \sin^2\psi + J_z, \tag{9}$$

$$J_\psi = m_m l_m^2 + m_t l_t^2 + 2m_{cw} l_{cw}^2 + J_x, \tag{10}$$

$$J = m_m l_m^2 + m_t l_t^2 - 2m_{cw} l_{cw}^2. \tag{11}$$

The dynamics of the beam w.r.t. the pitch angle and azimuth angle is strongly nonlinear and contains couplings. The torques τ_ϕ and τ_ψ are combined in the input vector $\mathbf{u} = [\tau_\phi \;\; \tau_\psi]^T$, with

$$\tau_\phi = F_\phi l_t \cos\psi \quad \text{and} \quad \tau_\psi = F_\psi l_m. \tag{12}$$

Here, F_ψ and F_ϕ denote the propulsive forces provided by the main rotor and the tail rotor, respectively. To handle the parameter uncertainty and unknown disturbances, the model is extended with two lumped disturbance torques $\mathbf{z} = [z_\phi \;\; z_\psi]^T$ acting on the azimuth and pitch axes, respectively. With the state vector $\mathbf{x} = [\phi \;\; \dot{\phi} \;\; \psi \;\; \dot{\psi}]^T$, the extended nonlinear state-space model $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{z})$ becomes

$$\begin{bmatrix} \phi \\ \dot{\phi} \\ \psi \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} \dot{\phi} \\ \frac{1}{J_\phi}\Big(-c_\phi\dot{\phi} - k_\phi\phi + J\dot{\phi}\dot{\psi}\sin(2\psi)\Big) + \frac{l_t\cos(\psi)}{J_\phi}F_\phi + \frac{1}{J_\phi}z_\phi \\ \dot{\psi} \\ \frac{1}{J_\psi}\Big(-c_\psi\dot{\psi} - J\frac{\dot{\phi}^2}{2}\sin(2\psi) - g\big((m_m l_m - m_t l_t)\cos\psi \\ \qquad + 2m_{cw} l_{cw}\sin\psi\big)\Big) + \frac{l_m}{J_\psi}F_\psi + \frac{1}{J_\psi}z_\psi \end{bmatrix}. \tag{13}$$

These disturbance torques, on the one hand, account for parameter uncertainty. On the other hand, the unmodelled effects due to the supply cables and gyroscopic torques as well as the couplings caused by the tail rotor and the main rotor in the case of angular accelerations of the propellers are encompassed by the lumped disturbance torques.

For trajectory tracking, the azimuth angle and the pitch angle are chosen as controlled outputs. The measurement vector $\mathbf{y}_m$ is given by

$$\mathbf{y}_m(t) = \begin{bmatrix} \phi & \psi \end{bmatrix}^T . \tag{14}$$

The nonlinear control-oriented model given in (13) represents a control-affine system. The system under consideration has matched uncertainties, since the unknown disturbance torques appear in the span of the control input. The systematic procedure of backstepping control design, however, allows to achieve global stabilisation under the influence of both matched and mismatched uncertainties. Therefore, in order to track desired trajectories robustly for the azimuth angle as well as the pitch angle, backstepping-based sliding mode control and adaptive backstepping approaches are proposed and investigated thoroughly.

3 Tracking Control Design for the TRAS

The aim of a robust control approach is to accurately track desired trajectories for both the azimuth angle and the pitch angle despite the parameter uncertainty and unknown disturbances. For this purpose, a comparison between two alternative robust nonlinear control strategies based on backstepping techniques—a backstepping-based sliding mode control and an adaptive backstepping control—is carried out. Generally, the backstepping control design is based on a recursive procedure by systematically choosing appropriate control Lyapunov functions, and the corresponding stabilising functions [12]. In the case of backstepping-based sliding mode control, the lumped disturbance torques along with the immeasurable angular velocities are estimated with the help of an EKF. The adaptive backstepping control design, however, includes a dynamic part that involves the derivation of nonlinear adaptive control laws for the estimation of unknown lumped disturbance torques. Thereby, both the robust performance and the trajectory tracking accuracy are improved. For the state estimation within the adaptive backstepping control strategy, robust sliding mode differentiators are employed for the estimation of the angular velocities. The asymptotic stability of the closed-loop systems for both feedback control approaches is proved using Lyapunov methods. Moreover, the stability of the overall control structures—involving the corresponding estimators—is investigated thoroughly by simulations and experiments. The detailed design procedures for both control techniques are presented in the following subsections.

3.1 Backstepping-Based Sliding Mode Control

A backstepping control design is generally carried out in a recursive fashion. The backstepping-based sliding mode control law can be derived in the two following steps:

Step 1:
Given the continuously differentiable reference signals of class C^2 for the azimuth angle ϕ_d and the pitch angle ψ_d, the tracking errors e_ϕ and e_ψ corresponding to the azimuth angle and the pitch angle, respectively, are defined as

$$e_\phi = \phi - \phi_d \text{ and } e_\psi = \psi - \psi_d. \tag{15}$$

Introducing a so-called virtual control input $\boldsymbol{\alpha} = [\alpha_\phi \quad \alpha_\psi]^T$, which has to be chosen properly, the time derivatives of the error dynamics are given by

$$\begin{aligned} \dot{e}_\phi &= \dot{\phi} - \dot{\phi}_d = \alpha_\phi + r_\phi - \dot{\phi}_d, \\ \dot{e}_\psi &= \dot{\psi} - \dot{\psi}_d = \alpha_\psi + r_\psi - \dot{\psi}_d, \end{aligned} \tag{16}$$

with

$$r_\phi = \dot{\phi} - \alpha_\phi \text{ and } r_\psi = \dot{\psi} - \alpha_\psi. \tag{17}$$

The stabilising functions α_ϕ and α_ψ are chosen as

$$\alpha_\phi = -k_\phi e_\phi + \dot{\phi}_d, \quad \alpha_\psi = -k_\psi e_\psi + \dot{\psi}_d, \tag{18}$$

with the strictly positive coefficients k_ϕ and k_ψ. This leads to the error dynamics

$$\dot{e}_\phi = -k_\phi e_\phi + r_\phi, \text{ and } \dot{e}_\psi = -k_\psi e_\psi + r_\psi. \tag{19}$$

Consider a quadratic control Lyapunov function V_1 and its corresponding time derivative,

$$\begin{aligned} V_1 &= \frac{1}{2}e_\phi^2 + \frac{1}{2}e_\psi^2, \\ \dot{V}_1 &= e_\phi \dot{e}_\phi + e_\psi \dot{e}_\psi = -k_\phi e_\phi^2 + e_\phi r_\phi - k_\psi e_\psi^2 + e_\psi r_\psi, \end{aligned} \tag{20}$$

where $e_\phi r_\phi$ and $e_\psi r_\psi$ will be eliminated in the next step. After elimination, the time derivative of the Lyapunov function is negative definite, i.e.

$$\dot{V}_1 = -k_\phi e_\phi^2 - k_\psi e_\psi^2 < 0, \tag{21}$$

hence, the asymptotic stability can be easily established.

Step 2:
Now, the error dynamics r_ϕ and r_ψ are given by

$$\dot{r}_\phi = \ddot{\phi} - \dot{\alpha}_\phi, \quad \text{and} \quad \dot{r}_\psi = \ddot{\psi} - \dot{\alpha}_\psi. \tag{22}$$

From (18) and (19), the following relationships can be obtained, i.e.

$$\dot{\alpha}_\phi = k_\phi^2 e_\phi - k_\phi r_\phi + \ddot{\phi}_d, \quad \text{and} \quad \dot{\alpha}_\psi = k_\psi^2 e_\psi - k_\psi r_\psi + \ddot{\psi}_d, \tag{23}$$

resulting in

$$\begin{aligned} \dot{r}_\phi &= f_2(\mathbf{x}) + \frac{l_t \cos(\psi)}{J_\phi} F_\phi + \frac{1}{J_\phi} z_\phi - k_\phi^2 e_\phi + k_\phi r_\phi - \ddot{\phi}_d, \\ \dot{r}_\psi &= f_4(\mathbf{x}) + \frac{l_m}{J_\psi} F_\psi + \frac{1}{J_\psi} z_\psi - k_\psi^2 e_\psi + k_\psi r_\psi - \ddot{\psi}_d, \end{aligned} \tag{24}$$

where the functions $f_2(\mathbf{x})$ and $f_4(\mathbf{x})$ are defined using the nonlinear control-oriented model (13). They are given by

$$\begin{aligned} f_2(\mathbf{x}) &= \frac{1}{J_\phi}\left(-c_\phi \dot{\phi} - k_\phi \phi + J \dot{\phi}\dot{\psi} \sin(2\psi)\right), \\ f_4(\mathbf{x}) &= \frac{1}{J_\psi}\left(-c_\psi \dot{\psi} - J \frac{\dot{\phi}^2}{2} \sin(2\psi) - g\big((m_m l_m - m_t l_t)\cos\psi + 2 m_{cw} l_{cw} \sin\psi\big)\right). \end{aligned} \tag{25}$$

For the backstepping-based sliding mode control, sliding manifolds need to be defined. Mathematically, the sliding manifolds are expressed as

$$s_\phi(\mathbf{x}) = c_\phi e_\phi + r_\phi\,, \; s_\psi(\mathbf{x}) = c_\psi e_\psi + r_\psi\,. \tag{26}$$

Here, strictly positive coefficients $c_i > 0,\ i \in \{\phi, \psi\}$, are employed. The sliding manifolds $s_\phi(\mathbf{x})$ and $s_\psi(\mathbf{x})$ correspond to the azimuth axis and the pitch axis, respectively. The time derivatives of the sliding manifolds are given by

$$\dot{s}_\phi(\mathbf{x}) = c_\phi \dot{e}_\phi + \dot{r}_\phi\,, \; \dot{s}_\psi(\mathbf{x}) = c_\psi \dot{e}_\psi + \dot{r}_\psi\,. \tag{27}$$

To ensure that the manifolds are reached in a finite period of time and that they are independent of the initial conditions of the system, sufficient reaching conditions

$$s_i(\mathbf{x})\dot{s}_i(\mathbf{x}) < 0, \forall s(\mathbf{x}) \neq 0\,,\ i \in \{\phi, \psi\}\,, \tag{28}$$

have to be fulfilled. A necessary condition for the existence of a sliding mode implies that, $\forall t \geq t_r$ the output trajectory should remain on the sliding surface, i.e., $s_i(\mathbf{x}) =$

$\dot{s}_i(\mathbf{x}) = 0$. Here, t_r denotes the finite reaching time. Hence, to achieve finite-time convergence, the time derivatives of the Lyapunov functions $1/2 \cdot s_i^2(\mathbf{x})$, $i \in \{\phi, \psi\}$, have to fulfil the following reaching conditions

$$s_i(\mathbf{x})\dot{s}_i(\mathbf{x}) \leq s_i(\mathbf{x})(-h_i s_i(\mathbf{x}) - \eta_i \operatorname{sgn}(s_i(\mathbf{x}))) < 0,\ h_i > 0 \text{ and } \eta_i > 0\ ,\ i \in \{\phi, \psi\}\ . \tag{29}$$

The parameters η_i determine the switching height and guarantee that the time-derivative of the Lyapunov functions become negative definite. The introduction of the additional damping terms $-h_i\ s_i(\mathbf{x})$ together with the switching terms ensure a shorter reaching time in comparison to the case where only the switching functions are employed. A detailed mathematical derivation for the reaching condition discussed at the end of this subsection highlights the effectiveness of the damping term.

Define a Lyapunov function V_2 along with its corresponding time differentiation as

$$\begin{aligned} V_2 &= \frac{1}{2}e_\phi^2 + \frac{1}{2}s_\phi^2 + \frac{1}{2}e_\psi^2 + \frac{1}{2}s_\psi^2, \\ \dot{V}_2(\mathbf{x}) &= \underbrace{e_\phi \dot{e}_\phi + s_\phi \dot{s}_\phi}_{\dot{V}_\phi} + \underbrace{e_\psi \dot{e}_\psi + s_\psi \dot{s}_\psi}_{\dot{V}_\psi}. \end{aligned} \tag{30}$$

In order to make the analysis simple, the time derivative of the control Lyapunov function $V_2(\mathbf{x})$ is split into two parts, $\dot{V}_\phi$ and $\dot{V}_\psi$. Considering

$$\dot{V}_\phi = e_\phi \dot{e}_\phi + s_\phi \dot{s}_\phi, \tag{31}$$

$$= e_\phi(-k_\phi e_\phi + r_\phi) + s_\phi\Big(c_\phi(-k_\phi e_\phi + r_\phi) + \dot{r}_\phi\Big), \tag{32}$$

$$= e_\phi(-k_\phi e_\phi + s_\phi - c_\phi e_\phi) + s_\phi\Big(c_\phi(-k_\phi e_\phi + r_\phi) + \dot{r}_\phi\Big), \tag{33}$$

$$= -(k_\phi + c_\phi)e_\phi^2 + s_\phi\Big(e_\phi + c_\phi(-k_\phi e_\phi + r_\phi) + f_2(\mathbf{x}) + \frac{l_t \cos(\psi)}{J_\phi}F_\phi + \frac{1}{J_\phi}z_\phi - k_\phi^2 e_\phi + k_\phi r_\phi - \ddot{\phi}_d\Big). \tag{34}$$

With the following choice of the control input F_ϕ, i.e.

$$F_\phi = \Big(-h_\phi s_\phi - \eta_\phi \operatorname{sgn}(s_\phi) - e_\phi - c_\phi(-k_\phi e_\phi + r_\phi) - f_2(\mathbf{x}) - \frac{1}{J_\phi}z_\phi + k_\phi^2 e_\phi - k_\phi r_\phi + \ddot{\phi}_d\Big)\frac{J_\phi}{l_t \cos(\psi)}\ , \tag{35}$$

the time derivative of the Lyapunov function is negative definite

$$\dot{V}_\phi = -(k_\phi + c_\phi)e_\phi^2 - h_\phi s_\phi^2 - \eta_\phi s_\phi \,\mathrm{sgn}(s_\phi) < 0. \tag{36}$$

It is worth mentioning that the singularity condition due to $\psi = \pm\pi/2$ does not arise in the real set-up which is fortunately outside the range of feasible pitch angles because of the mechanical design of the test-rig.

Likewise, the control input F_ψ for the main rotor can be determined based on the design procedure already described in detail for the control input F_ϕ. Therefore, the following expression is obtained for the F_ψ, i.e.

$$F_\psi = \Bigg(-h_\psi s_\psi - \eta_\psi \,\mathrm{sgn}(s_\psi) - e_\psi - c_\psi(-k_\psi e_\psi + r_\psi) - f_4(\mathbf{x}) - \frac{1}{J_\psi} z_\psi + k_\psi^2 e_\psi - k_\psi r_\psi + \ddot{\psi}_d\Bigg)\frac{J_\psi}{l_m}. \tag{37}$$

With this choice of the control input, the time derivative of the Lyapunov function V_ψ is

$$\dot{V}_\psi = -(k_\psi + c_\psi)e_\psi^2 - h_\psi s_\psi^2 - \eta_\psi s_\psi \,\mathrm{sgn}(s_\psi) < 0. \tag{38}$$

Hence, the time derivative of the control Lyapunov function from (30) can be rewritten as

$$\dot{V}_2(\mathbf{x}) = -(k_\phi + c_\phi)e_\phi^2 - h_\phi s_\phi^2 - \eta_\phi |s_\phi| - (k_\psi + c_\psi)e_\psi^2 - h_\psi s_\psi^2 - \eta_\psi |s_\psi| < 0. \tag{39}$$

Since all the coefficients are positive, the time derivative of the Lyapunov function remains negative definite. As a consequence, the asymptotic stability of the system is guaranteed.

In order to proof a convergence in finite time to the sliding manifold, i.e. $s(\mathbf{x}) = \dot{s}(\mathbf{x}) = 0$, a detailed mathematical analysis is provided in the following.

Proof of Finite Reaching Time

For brevity, only the sliding manifold regarding the azimuth angle is considered. According to Lyapunov's stability theory, the existence as well as the reaching conditions for a sliding mode can be summarised as follows: if there exists a Lyapunov function $V_{s,\phi}$ with a negative-definite time-derivative [12, 20], i.e.

$$V_{s,\phi} = \frac{1}{2} s_\phi^2, \tag{40}$$

$$\dot{V}_{s,\phi} = s_\phi \dot{s}_\phi < 0 \quad \text{for } s_\phi \neq 0, \tag{41}$$

asymptotic stability can be ensured. To achieve finite-time convergence, however, the time derivative of the Lyapunov function has to fulfil the following reaching

condition

$$\dot{V}_{s,\phi} = s_\phi \dot{s}_\phi \leq s_\phi \left(-h_\phi s_\phi - \eta_\phi \operatorname{sgn}(s_\phi) \right) < 0. \tag{42}$$

The parameter $\eta_\phi > 0$ determines the switching height and guarantees that the time derivative of the Lyapunov function becomes negative definite. Moreover, the introduction of the damping constant $h_\phi > 0$ leads to fast finite-time convergence to the sliding manifold as shown in the sequel. To determine the reaching time t_r, inequality (42) can be rewritten in terms of $V_{s,\phi}$ as

$$\dot{V}_{s,\phi} + 2h_\phi V_{s,\phi} \leq -\sqrt{2}\eta_\phi \sqrt{V_{s,\phi}}. \tag{43}$$

This inequality is in the form of a generalised Bernoulli's differential equation and can be easily transformed to a first-order linear non-autonomous differential equation. Multiplying both sides of (43) with $\dfrac{1}{\sqrt{V_{s,\phi}}}$ leads to

$$\frac{1}{\sqrt{V_{s,\phi}}} \dot{V}_{s,\phi} + 2h_\phi \sqrt{V_{s,\phi}} \leq -\sqrt{2}\eta_\phi. \tag{44}$$

Applying the transformation

$$\tilde{V} = \sqrt{V_{s,\phi}}, \tag{45}$$

leads to a first-order differential equation of the form

$$\dot{\tilde{V}} + h_\phi \tilde{V} \leq -\frac{\eta_\phi}{\sqrt{2}}. \tag{46}$$

Keeping in mind that the initial value of $\tilde{V}(t)$ at time $t = 0$ is $\tilde{V}(0)$, the solution of this non-homogeneous linear differential equation results in

$$\tilde{V}(t) \leq \left(\tilde{V}(0) + \frac{\eta_\phi}{\sqrt{2}h_\phi} \right) \exp^{-h_\phi t} - \frac{\eta_\phi}{\sqrt{2}h_\phi}. \tag{47}$$

Transforming back this solution in terms of the original Lyapunov function $V_{s,\phi}(t)$ leads to

$$V_{s,\phi}(t) \leq \left[\left(\sqrt{V_{s,\phi}(0)} + \frac{\eta_\phi}{\sqrt{2}h_\phi} \right) \exp^{-h_\phi t} - \frac{\eta_\phi}{\sqrt{2}h_\phi} \right]^2. \tag{48}$$

Consequently, $V_{s,\phi}(t)$ reaches zero in a finite time t_r bounded by

$$t_r \leq \frac{1}{h_\phi} \ln\left[\frac{\sqrt{2}h_\phi}{\eta_\phi}\left(\sqrt{V_{s,\phi}(0)} + \frac{\eta_\phi}{\sqrt{2}h_\phi}\right)\right]. \tag{49}$$

Here, it becomes obvious that the reaching time is proportional to the natural logarithmic of the square root of the initial value, i.e.$\sqrt{V_{s,\phi}(0)}$. This is in contrast to the case where only the switching term is employed and reaching time is proportional to the square root of the initial value $\sqrt{V_{s,\phi}(0)}$ only [12, 20, 22]. Therefore, condition (43) guarantees a fast finite reaching time. The same procedure can be repeated for the sliding manifold for the pitch axis and a finite reaching time can be guaranteed. It is worth mentioning that the reaching time condition is generally applicable for all sliding mode control design techniques where the finite-time convergence is achieved using a combination of a damping term with a switching term as already introduced in (42).

After reaching the sliding surface $s_i = 0$, $i \in \{\phi, \psi\}$, in finite time $t < t_r$, the closed-loop error dynamics of the azimuth axis and the pitch axes are governed by the following set of state equations

$$\begin{bmatrix} \dot{e}_i \\ \dot{r}_i \end{bmatrix} = \begin{bmatrix} -k_i & 1 \\ -1 + k_i c_i & -c_i \end{bmatrix} \begin{bmatrix} e_i \\ r_i \end{bmatrix}. \tag{50}$$

The characteristic polynomial of the error dynamics results in

$$p_{BS\text{-}SMC}(s) = s^2 + (k_i + c_i)s + 1, \tag{51}$$

which satisfies Hurwitz's stability criterion for strictly positive control gains $k_i > 0$ and $c_i > 0$. The eigenvalues of the closed-loop error dynamics during an ideal sliding mode are, hence, located in the left half s-plane, thus the asymptotic stability of the overall closed-loop system is guaranteed.

The sliding mode control suffers from the chattering phenomenon caused by fast switching actions introduced by the $\mathrm{sgn}(s)$ function and may lead to the excitation of unmodelled high-frequency dynamics. To counteract this effect, smooth switching functions $\tanh(s_i/\epsilon)$ with a strictly positive constant ϵ—representing a boundary layer thickness—are utilised. The chattering reduction depends on the value of ϵ at the cost of robustness. By using a large value for the boundary layer thickness ϵ the reaching time increases. This is due to the fact that the control input within the boundary thickness changes in a smooth way rather in a fast switching way. Therefore, the boundary layer thickness should be carefully selected. A typical value of $\epsilon \ll 1$ is chosen for the boundary layer thickness.

3.2 Adaptive Backstepping Control Design

To compensate the unknown lumped disturbance torques z_ϕ and z_ψ, appropriate adaptive control laws are designed for their estimation. On choosing single integrators as disturbance models, which proved advantageous in many applications, the corresponding dynamics are governed by the following relationship

$$\dot{z}_\phi = 0 \quad \text{and} \quad \dot{z}_\psi = 0.$$

Note that the integrators are driven by the output errors in observer or filter schemes. Using the recursive nature of the backstepping control, the design procedure for the adaptive backstepping control for TRAS is performed in two steps.

Step 1:
For brevity, any derivations in this step are omitted because this step is similar to the first step of the backstepping-based sliding mode control design scheme presented in Sect. 3.1.

Step 2:
To stabilise r_ϕ and r_ψ dynamics as well as to carry out the design of the adaptive laws for the lumped disturbances, the Lyapunov candidate V_2 and the corresponding differentiation $\dot{V}_2$ w.r.t. time are considered

$$V_2 = \frac{1}{2}e_\phi^2 + \frac{1}{2}r_\phi^2 + \frac{1}{2}\tilde{z}_\phi \Gamma_\phi^{-1}\tilde{z}_\phi + \frac{1}{2}e_\psi^2 + \frac{1}{2}r_\psi^2 + \frac{1}{2}\tilde{z}_\psi \Gamma_\psi^{-1}\tilde{z}_\psi, \tag{52}$$

$$\dot{V}_2 = \underbrace{e_\phi \dot{e}_\phi + r_\phi \dot{r}_\phi + \tilde{z}_\phi \Gamma_\phi^{-1}\dot{\tilde{z}}_\phi}_{\dot{V}_\phi} + \underbrace{e_\psi \dot{e}_\psi + r_\psi \dot{r}_\psi + \tilde{z}_\psi \Gamma_\psi^{-1}\dot{\tilde{z}}_\psi}_{\dot{V}_\psi}. \tag{53}$$

The strictly positive parameters Γ_ϕ and Γ_ψ represent the adaptation gains and determine the convergence speed of the estimated values to their true ones. In order to derive the control law in a simpler way, expression (53) is split into two terms denoted by $\dot{V}_\phi$ and $\dot{V}_\psi$. The estimation errors and the corresponding time derivatives of the disturbance torques are given by

$$\begin{aligned} \tilde{z}_\phi &= z_\phi - \hat{z}_\phi, & \dot{\tilde{z}}_\phi &= \dot{z}_\phi - \dot{\hat{z}}_\phi = -\dot{\hat{z}}_\phi, \\ \tilde{z}_\psi &= z_\psi - \hat{z}_\psi, & \dot{\tilde{z}}_\psi &= \dot{z}_\psi - \dot{\hat{z}}_\psi = -\dot{\hat{z}}_\psi. \end{aligned} \tag{54}$$

Regarding the lumped disturbance torque of the azimuth axis, a thorough analysis will be presented in the sequel for the derivation of the control law and the adaptive law. Considering

$$\dot{V}_\phi = e_\phi \dot{e}_\phi + r_\phi \dot{r}_\phi + \tilde{z}_\phi \Gamma_\phi^{-1}\dot{\tilde{z}}_\phi, \tag{55}$$

substituting (13), (24) and (54) into (55) leads to

$$\begin{aligned}\dot{V}_\phi &= e_\phi(-k_\phi e_\phi + r_\phi) + r_\phi\left(f_2(\mathbf{x}) + \frac{l_t\cos(\psi)}{J_\phi}F_\phi + \frac{1}{J_\phi}z_\phi - k_\phi^2 e_\phi + k_\phi r_\phi - \ddot{\phi}_d\right) \\ &\quad - \tilde{z}_\phi \Gamma_\phi^{-1}\dot{\hat{z}}_\phi, \\ &= -k_\phi e_\phi^2 + r_\phi\left(e_\phi + f_2(\mathbf{x}) + \frac{l_t\cos(\psi)}{J_\phi}F_\phi - k_\phi^2 e_\phi + k_\phi r_\phi - \ddot{\phi}_d\right) \\ &\quad + z_\phi\left(\frac{1}{J_\phi}r_\phi - \Gamma_\phi^{-1}\dot{\hat{z}}_\phi\right) + \hat{z}_\phi\Gamma_\phi^{-1}\dot{\hat{z}}_\phi. \end{aligned} \tag{56}$$

Now, the update law for the estimation of the disturbance torque z_ϕ can be stated as

$$\dot{\hat{z}}_\phi = \frac{1}{J_\phi}\Gamma_\phi r_\phi. \tag{57}$$

Substituting back the parameter update law into (56) leads to

$$\dot{V}_\phi = -k_\phi e_\phi^2 + r_\phi \underbrace{\left(e_\phi + f_2(\mathbf{x}) + \frac{l_t\cos(\psi)}{J_\phi}F_\phi - k_\phi^2 e_\phi + k_\phi r_\phi - \ddot{\phi}_d + \frac{1}{J_\phi}\hat{z}_\phi\right)}_{-\tilde{k}_\phi r_\phi} < 0. \tag{58}$$

For asymptotic stability, the Lyapunov function must be negative definite. This condition is ensured by replacing the term within the bracket with $-\tilde{k}_\phi r_\phi$ with the strictly positive control parameter $\tilde{k}_\phi > 0$. Therefore, the control law F_ϕ is given by

$$F_\phi = \left(-(\tilde{k}_\phi + k_\phi)r_\phi - (1 - k_\phi^2)e_\phi - f_2(\mathbf{x}) - \frac{1}{J_\phi}\hat{z}_\phi + \ddot{\phi}_d\right)\frac{J_\phi}{l_t\cos(\psi)}. \tag{59}$$

Similarly, based on the same guidelines provided above for the azimuth axis, the control law for the pitch motion can be derived. Thus, the corresponding disturbance torque update law, the time derivative of the Lyapunov function and the control input are summarised as follows

$$\dot{\hat{z}}_\psi = \frac{1}{J_\psi}\Gamma_\psi r_\psi. \tag{60}$$

$$\dot{V}_\psi = -k_\psi e_\psi^2 + r_\psi \underbrace{\left(e_\psi + f_4(\mathbf{x}) + \frac{l_m}{J_\psi}F_\psi - k_\psi^2 e_\psi + k_\psi r_\psi - \ddot{\psi}_d + \frac{1}{J_\psi}\hat{z}_\psi\right)}_{-\tilde{k}_\psi r_\psi} < 0, \tag{61}$$

$$F_\psi = \left(-(\tilde{k}_\psi + k_\psi)r_\psi - (1 - k_\psi^2)e_\psi - f_4(\mathbf{x}) - \frac{1}{J_\psi}\hat{z}_\psi + \ddot{\psi}_d\right)\frac{J_\psi}{l_m}. \tag{62}$$

Subsequently, the time derivative of the Lyapunov function V_2 is stated as

$$\dot{V}_2 = -k_\phi e_\phi^2 - \tilde{k}_\phi r_\phi^2 - k_\psi e_\psi^2 - \tilde{k}_\psi r_\psi^2 < 0. \tag{63}$$

With the choice of the positive control parameters, the time derivative of the control Lyapunov function is negative definite, and the asymptotic stability of the overall closed-loop system can be guaranteed.

The closed-loop error dynamics with respect to the azimuth axis and the pitch axes are governed by the following set of state equations

$$\begin{bmatrix} \dot{e}_i \\ \dot{r}_i \\ \dot{\tilde{z}}_i \end{bmatrix} = \begin{bmatrix} -k_i & 1 & 0 \\ -1 & -\tilde{k}_i & \frac{1}{J_i} \\ 0 & -\frac{1}{J_i}\Gamma_i & 0 \end{bmatrix} \begin{bmatrix} e_i \\ r_i \\ \tilde{z}_i \end{bmatrix}, \tag{64}$$

with $i \in \{\phi, \psi\}$. The characteristic polynomial for the error dynamics is given by

$$p_{ABS} = s^3 + (k_i + \tilde{k}_i)s^2 + \left(1 + k_i\tilde{k}_i + \frac{1}{J_i^2}\Gamma_i\right)s + \frac{k_i}{J_i^2}\Gamma_i. \tag{65}$$

Since the parameters $k_i, \tilde{k}_i, \Gamma_i$ as well as the moment of inertia J_i are strictly positive, the Hurwitz stability criterion is satisfied. Therefore, the eigenvalues of the closed-loop error dynamics are located in the left half s-plane and the tracking error goes asymptotically to zero. Here, the control gains $k_i, \tilde{k}_i$ and Γ_i determine the convergence rate. Using large control gains, it is possible to increase the decay rate; an asymptotic convergence characteristic, however, is always present. In contrast to this fact, the backstepping-based sliding mode control guarantees a finite-time convergence based on the reaching condition.

4 State and Disturbance Estimation

As the backstepping-based sliding mode control laws (35) and (37) require the knowledge of the immeasurable angular velocities as well as the unknown lumped disturbance torques, a discrete-time EKF is employed for the estimation tasks. In the case of the adaptive backstepping control laws (59) and (62), the angular velocities are estimated with the help of a robust sliding mode differentiator, whereas the disturbance torques are estimated using the corresponding adaptive laws. The discussion in this section focusses on the implementation of both the EKF and the robust sliding mode differentiator.

4.1 Discrete-Time Extended Kalman Filter

For the estimation of the lumped disturbance torques, the integrator disturbance models are introduced according to

$$\dot{z}_\phi = 0, \text{ and } \dot{z}_\psi = 0. \tag{66}$$

Note that these disturbance models are excited in the case of output errors between the measured and the estimated output variables. The state vector of the extended system representation results in

$$\mathbf{x}_e = \left[\phi \ \dot{\phi} \ \psi \ \dot{\psi} \ z_\phi \ z_\psi \right]^T, \tag{67}$$

and the measurement vector becomes

$$\mathbf{y}_m = \mathbf{C}_{m,e}\, \mathbf{x}_e = \begin{bmatrix} 1\,0\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0\,0 \end{bmatrix} \mathbf{x}_e. \tag{68}$$

Given the continuous-time state equation of the extended system

$$\dot{\mathbf{x}}_e = \mathbf{f}(\mathbf{x}_e, \mathbf{u}), \tag{69}$$

an explicit Euler time discretization of (69) and an introduction of additive noise processes lead to the following discrete-time state-space representation used for the EKF design

$$\mathbf{x}_{e,k+1} = \underbrace{\mathbf{x}_{e,k} + T_s \mathbf{f}_k(\mathbf{x}_{e,k}, \mathbf{u}_k)}_{\boldsymbol{\varphi}_k(\mathbf{x}_{e,k}, \mathbf{u}_k)} + \mathbf{w}_k, \tag{70}$$

$$\mathbf{y}_{m,k} = \mathbf{C}_{m,e} \mathbf{x}_{e,k} + \mathbf{v}_k. \tag{71}$$

Here, T_s denotes the sampling time, $\mathbf{x}_{e,k}$ the extended state vector, $\mathbf{u}_k$ the control input vector, and $\mathbf{y}_{m,k}$ the measured output at discrete-time t_k. Furthermore, the process noise and the measurement noise are given by $\mathbf{w}_k$ and $\mathbf{v}_k$, respectively. Both are assumed to be zero-mean Gaussian white noise processes with zero cross-correlation. The vanishing cross-correlation leads to diagonal covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ characterising the process noise $\mathbf{w}_k$ and the measurement noise $\mathbf{v}_k$, respectively. Figure 3 shows that the implementation of the discrete-time EKF can be divided into two stages, namely a prediction stage and an innovation stage [21]. The error covariance matrix is denoted by $\mathbf{P}_k$. The algorithm for the discrete-time EKF can be summarised at each time t_k as follows, cf. [21]:

- State prediction

$$\tilde{\mathbf{x}}_{e,k+1} = \boldsymbol{\varphi}_k(\hat{\mathbf{x}}_{e,k}, \mathbf{u}_k) \tag{72}$$

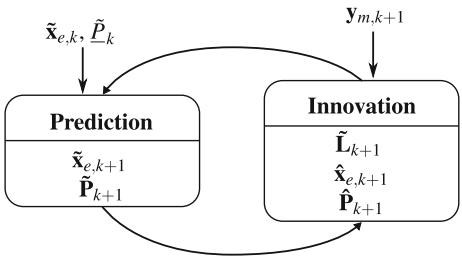


Fig. 3 Implementation of the discrete-time EKF

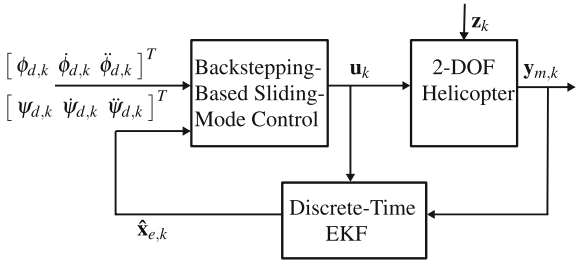


Fig. 4 Implementation scheme of the backstepping-based sliding mode control in combination with the discrete-time EKF

- Prediction of the error covariance matrix $\tilde{\mathbf{P}}_{k+1}$

$$\tilde{\mathbf{P}}_{k+1} = \boldsymbol{\Phi}_k \hat{\mathbf{P}}_k \boldsymbol{\Phi}_k^T + \mathbf{Q}_k \ , \text{ with } \ \boldsymbol{\Phi}_k = \frac{\partial \boldsymbol{\varphi}_k(\mathbf{x}_{e,k}, \mathbf{u}_k)}{\partial \mathbf{x}_{e,k}}\Big|_{\hat{\mathbf{x}}_{e,k}} \tag{73}$$

- Update of the gain matrix $\tilde{\mathbf{L}}_{k+1}$

$$\tilde{\mathbf{L}}_{k+1} = \tilde{\mathbf{P}}_{k+1} \mathbf{C}_{m,e}^T \left(\mathbf{C}_{m,e} \tilde{\mathbf{P}}_{k+1} \mathbf{C}_{m,e}^T + \mathbf{R}_k \right)^{-1} \tag{74}$$

- Update of the state vector $\hat{\mathbf{x}}_{e,k+1}$

$$\hat{\mathbf{x}}_{e,k+1} = \tilde{\mathbf{x}}_{e,k+1} + \tilde{\mathbf{L}}_{k+1} \left(\mathbf{y}_{m,k+1} - \mathbf{C}_{m,e} \tilde{\mathbf{x}}_{e,k+1} \right) \tag{75}$$

- Update of the error covariance matrix for the next sampling interval

$$\hat{\mathbf{P}}_{k+1} = \left(\mathbf{I} - \tilde{\mathbf{L}}_{k+1} \mathbf{C}_{m,e} \right) \tilde{\mathbf{P}}_{k+1} \tag{76}$$

The block diagram of the backstepping-based sliding control along with the EKF in depicted in Fig. 4.

4.2 Robust Sliding Mode Differentiator

The real differentiation using a DT1-system is sensitive to input noise. On the one hand, choosing a low cut-off frequency reduces the negative impact of high-frequency noise. On the other hand, it introduces a significant and undesired time delay. Accordingly, a high cut-off frequency results in small time delay at the cost of high-frequency noise in the output signal. This trade-off between noise and time delay in the estimation of the angular velocities could even lead to instability [13]. For the estimation of the angular velocities as required for implementation of the adaptive backstepping control, Levant's differentiator—based on a robust exact differentiation via sliding mode techniques—is employed [6, 13, 14]. The exact derivatives are calculated by successive implementation of a robust exact first-order differentiator based on a second-order sliding mode control. The finite-time convergence of this robust differentiator is proved in [13]. Provided that γ is the maximum magnitude of the measurement noise, the accuracy of the differentiator is proportional to $\gamma^{1/2}$ for the second time-derivative of the applied signal. The design procedure for the estimation of the exact differentiation of the angular velocity $\dot{\phi}$ is outlined in the following. Consider the azimuth angle ϕ as a basis signal with the third time-derivative having a known Lipschitz constant L. The robust sliding mode differentiator aiming at the estimation of $\dot{\phi}(t)$, $\ddot{\phi}(t)$ and $\dddot{\phi}(t)$ is such a way that it is exact in the absence of measurement noise. A second-order differentiator for the input ϕ with $|\dddot{\phi}(t_0)| \leq L$ according to [14] is given by

$$\begin{aligned} \dot{z}_0 &= \nu_0, \quad \nu_0 = -3L^{1/3}|z_0 - \phi|^{2/3}\,\mathrm{sgn}(z_0 - \phi) + z_1, \\ \dot{z}_1 &= \nu_1, \quad \nu_1 = -1.5L^{1/2}|z_1 - \nu_0|^{1/2}\,\mathrm{sgn}(z_1 - \nu_0) + z_2, \\ \dot{z}_2 &= -1.1L\,\mathrm{sgn}(z_2 - \nu_1), \end{aligned} \tag{77}$$

with $[z_0\ z_1\ z_2]^T = [\hat{\phi}\ \dot{\hat{\phi}}\ \ddot{\hat{\phi}}]^T$. Similarly, the angular velocity for the pitch angle can be easily estimated using the robust sliding mode differentiation.

Figure 5 shows the implementation scheme of the controller in combination with the robust sliding mode differentiator.

The closed-loop stability of the overall control structure consisting of the nonlinear control techniques along with the estimators has been investigated thoroughly in simulations.

5 Simulation and Experimental Results

In this section, the proposed backstepping-based sliding mode control (BS-SMC) and the adaptive backstepping (ABS) control laws in combination with the EKF and the robust state differentiators, respectively, are investigated by both simulations and experimental evaluations. To guarantee realistic simulation results, the system model is extended with measurement noise concerning the incremental encoders for

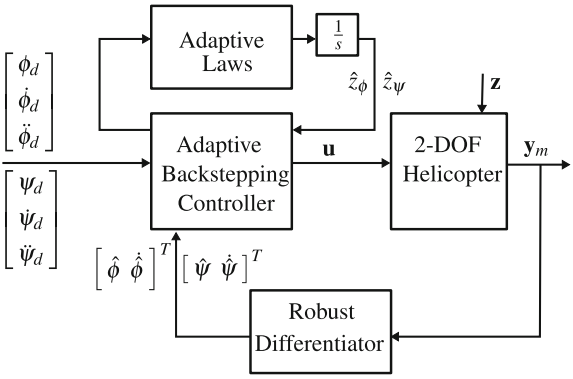


Fig. 5 Implementation scheme of the adaptive backstepping control along with the robust sliding mode differentiator

the azimuth and pitch angles. The mechanical construction of the TRAS permits the azimuth angle and the pitch angle to lie in the following regions

$$-3.2\,\text{rad} \leq \phi \leq 3.2\,\text{rad} \quad \text{and} \quad -1\,\text{rad} \leq \psi \leq 1\,\text{rad}.$$

The counter-weights have been adjusted in such a way that the initial inclination of the pitch angle vanishes. Moreover, the initial azimuth angle is set to 0 rad. The nominal parameter values for the mass moments of inertia, the restoring spring coefficients, and the viscous damping are identified experimentally by means of a nonlinear least-squares minimization based on the available measurement data. The control inputs and the measurements are updated with a sample time of $T_s = 5\,\text{ms}$. This small sample time makes it possible to realise a quasi-continuous implementation of the backstepping-based sliding mode control in combination with the discrete-time EKF as depicted in Fig. 4. In previous studies like [3, 4, 11], either relatively slow trajectories or step changes are considered for the azimuth angle as well as the pitch angle. In this contribution, however, fast but feasible trajectories are employed. The desired—three times continuously differentiable desired trajectories of class C^3 for the azimuth and the pitch angles are depicted in Fig. 6. In order to gain insight of the contour of the two rotor aerodynamic system in a real 3D space, the time evolution of the desired trajectories is illustrated in Fig. 7. It can be seen clearly that the chosen combination of the desired trajectories leads to quite general operating scenarios for the azimuth angle and the pitch angle. Herein, the starting point of the trajectory is marked with a black point, whereas the end point is indicated by a grey point.

Since the backstepping-based sliding mode control technique requires a discrete-time EKF for the estimation of immeasurable states and unknown disturbance torques, the covariance matrix $\mathbf{Q}_k$, $\mathbf{P}_0$ and $\mathbf{R}_k$ needs to be properly initialized. The covariance matrix $\mathbf{Q}_k$ and the initial error covariance matrix $\mathbf{P}_0$ are chosen heuristically to achieve fast error convergence rate as well as to minimise the effect of process noise on the estimates. Furthermore, the covariance matrix $\mathbf{R}_k$ corresponding to the sensor noise is parametrised based on the standard deviation of the incremental encoders. It is worth mentioning that the covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ remain

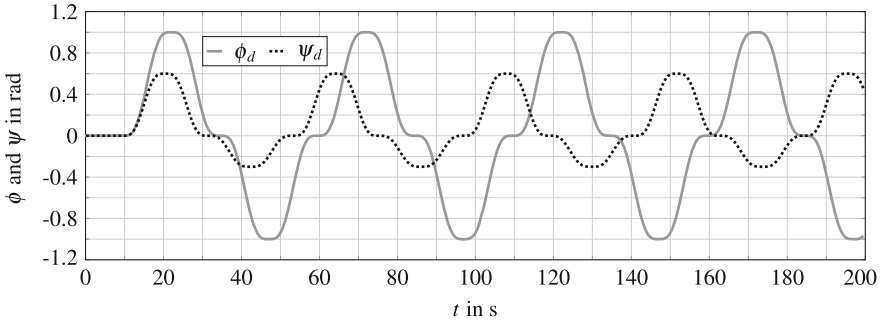


Fig. 6 Desired trajectories for the azimuth angle and the pitch angle

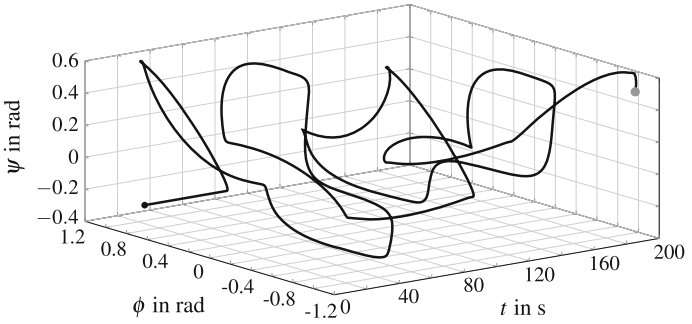


Fig. 7 Representation of the desired trajectory for the pitch angle and the azimuth angle w.r.t. the time t

constant irrespective of the time argument k. After thorough simulation studies, the control algorithm can be directly implemented on the TRAS test-rig. Nevertheless, the design parameters of the model-based controller have been slightly adjusted due to the influence of the disturbance torques at the test-rig in order to further improve the tracking performance.

To make a fair comparison between the aforementioned robust nonlinear control strategies, two performance metrics are introduced—the mean absolute tracking error e_{MAE} as well as the root-mean-square (RMS) error e_{RMS}, which are defined as

$$e_{MAE} = \frac{1}{N}\sum_{k=1}^{N} |e(k)|, \quad \text{and} \quad e_{RMS} = \sqrt{\frac{1}{N}\sum_{k=1}^{N} e^2(k)},$$

where N is the length of the measurement vector.

5.1 Simulation Results

The system model (13) shows that the disturbances $z_q,\ q \in \{\phi, \psi\}$ are acting on the system. Hence, in the simulations analysis, the disturbance torque z_q is modelled as follows

$$z_q = z_{q,nom} - c_{f,q}\,\mathrm{sgn}(\dot{q}_d). \tag{78}$$

Here, $z_{q,nom}$ represents a constant disturbance torque acting on the system during the whole length of the simulation. Moreover, the Coulomb friction torque represents a time-varying disturbance acting on the system.

The desired and simulated trajectories for the azimuth angle as well as the pitch angle along with the corresponding tracking errors are shown in Fig. 8. The desired and simulated trajectories for the azimuth angle and the pitch angle are in good agreement during both the transient phase and the steady state for both nonlinear control approaches, cf. Fig. 8a, b. The magnitude of the tracking errors during the transient phase for the backstepping-based sliding mode control is smaller in comparison to the adaptive backstepping control, see Fig. 8c, d. Moreover, an excellent tracking behaviour is obtained during the steady-state phase, where the steady-state errors for both nonlinear robust control techniques converge to zero. The control inputs for the

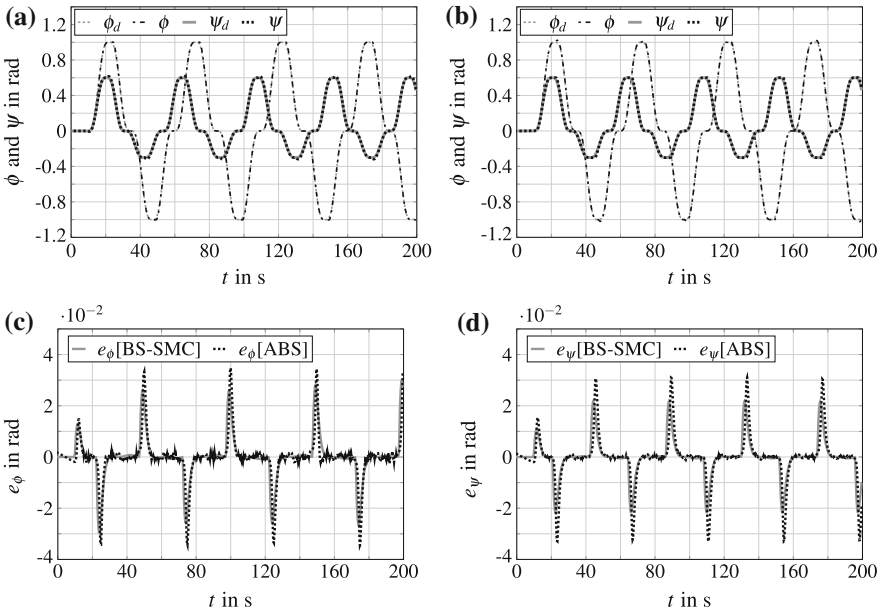


Fig. 8 Trajectory tracking for the azimuth angle and the pitch angle with the corresponding tracking errors. **a** Trajectory tracking using BS-SMC. **b** Trajectory tracking using ABS. **c** Tracking error for the azimuth angle. **d** Tracking error for the pitch angle

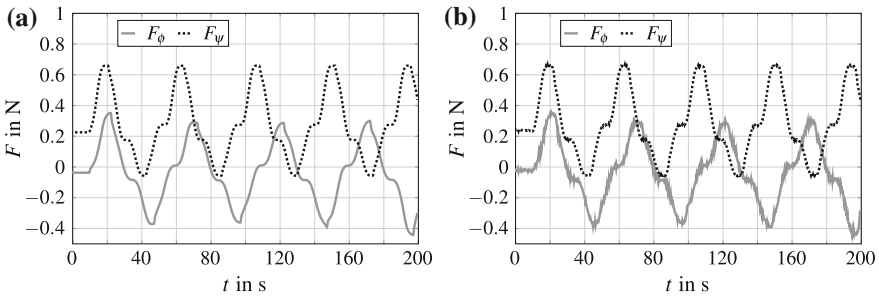


Fig. 9 Control inputs for the main rotor and the tail rotor. **a** Control inputs using BS-SMC. **b** Control inputs using ABS

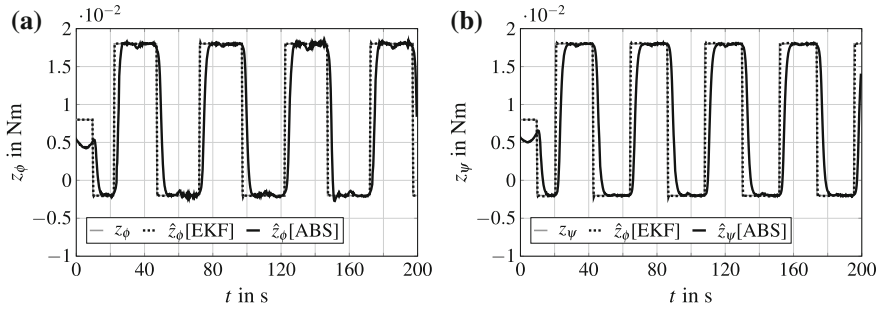


Fig. 10 Estimated lumped disturbance torques acting along the azimuth axis and the pitch axis. **a** Estimated disturbance torque along the azimuth axis. **b** Estimated disturbance torque along the pitch axis

main rotor and the tail rotor provided by the backstepping-based sliding mode and the adaptive backstepping strategies are depicted in Fig. 9.

Figure 10 illustrates the estimated disturbances for both the azimuth axis and the pitch axis. It can be seen clearly that the convergence speed of the estimated disturbances in the case of EKF is faster in comparison to the adaptive backstepping control law. Nevertheless, increasing the adaptation gain parameters makes it possible to achieve a faster convergence of the estimated disturbance to their true values. However, this introduces oscillations around the steady-state values of the desired disturbances, which in turns generates oscillatory control actions that could even lead to instability. The oscillations in the control input for the adaptive backstepping method can be seen easily in the corresponding figure, cf. Fig. 9b. Hence, an upper limit on the choice of adaptation gain parameters must not be exceeded during the experiments.

For a quantitative comparison between the control strategies, the performance metrics described earlier are considered. The numerical evaluation according to mean absolute error e_{MAE} and root-mean-square error e_{RMS} in case of the simulation studies are listed in Table 1.

Table 1 Evaluation criteria for the robust nonlinear control laws

Tracking error	MAE in rad	RMSE in rad
e_ϕ (BS-SMC)	0.0033	0.0077
e_ϕ (ABS)	0.0034	0.0078
e_ψ (BS-SMC)	0.0031	0.0066
e_ψ (ABS)	0.0038	0.0083

Regarding the tracking errors, it becomes obvious that the backstepping-based sliding mode control outperforms the adaptive backstepping control. In order to validate the simulation results, a thorough experimental study on a dedicated test-rig at the Chair of Mechatronics is conducted as well.

5.2 Experimental Results

The control laws are implemented on the TRAS set-up with a sampling time of $T_s = 5\,\text{ms}$. The covariance matrices $\mathbf{Q}_k$, $\mathbf{P}_0$ and $\mathbf{R}_k$ are initialized with the same values as in the simulation study. Moreover, an identical parameter value L is used for the robust state differentiator in both the simulation and the experiment. The trajectory tracking for the azimuth angle along with the tracking errors using both robust nonlinear control techniques are shown in Fig. 11.

The desired and measured trajectories are in good agreement during both the transient and steady-state phases, see Fig. 11a, b. Furthermore, it can be clearly seen that a superior error convergence is obtained for the azimuth angle in the case of the backstepping-based sliding mode controller.

The tracking errors shown in Fig. 11c, d illustrate the excellent performance of the backstepping-based sliding mode control. Herein, the magnitude of the transient tracking error and the steady-state error are smaller for the backstepping-based sliding mode control as compared to the adaptive backstepping approach.

The corresponding control inputs for the main rotor and the tail rotor for both robust control strategies are depicted in Fig. 12.

Figure 13 shows that the estimated disturbances both for the azimuth axis and the pitch axis. The discrete-time EKF provides slightly smoother estimates for the disturbance torques in comparison with the ones of the adaptive backstepping control law.

The estimates for the angular velocities—using the EKF in combination with the backstepping-based sliding mode control and the robust sliding mode differentiator (RSMD) in combination with the adaptive backstepping control—are depicted in Fig. 14. The estimates are of comparable magnitude and quality.

Likewise to the simulation studies, the experimental results are also assessed based on the same evaluation criteria, namely the mean absolute error e_{MAE} and the root-mean-square error e_{RMS}. A quantitative comparison of the nonlinear control techniques is stated in Table 2.

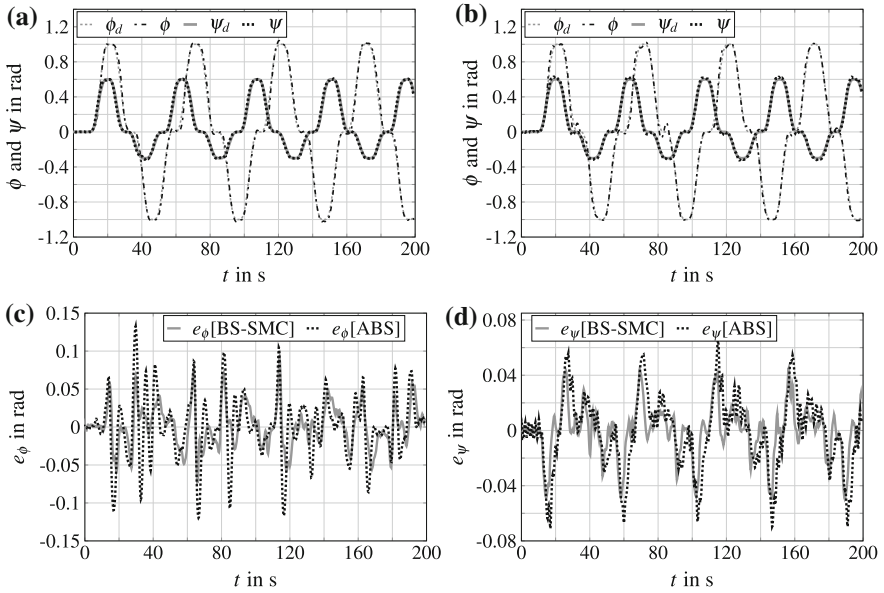


Fig. 11 Trajectory tracking for the azimuth angle and the pitch angle with the corresponding tracking errors. **a** Trajectory tracking using BS-SMC. **b** Trajectory tracking using ABS. **c** Tracking error for the azimuth angle. **d** Tracking error for the pitch angle

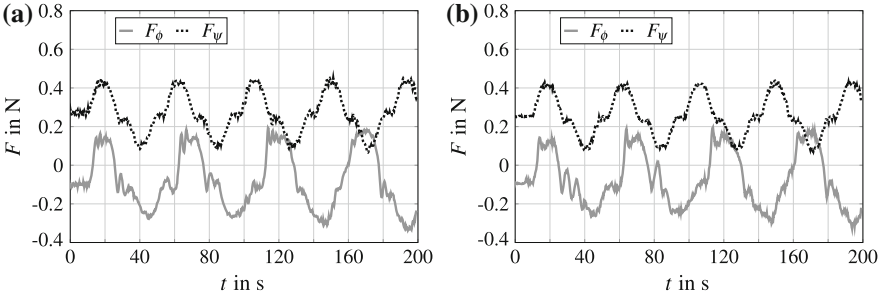


Fig. 12 Control inputs for the main rotor and the tail rotor. **a** Control inputs using BS-SMC. **b** Control inputs using ABS

Obviously, the backstepping-based sliding mode control leads to an improved performance in terms of the tracking error as compared to the adaptive backstepping control. Thus, the experimental results comply with the simulation results and validate both control approaches.

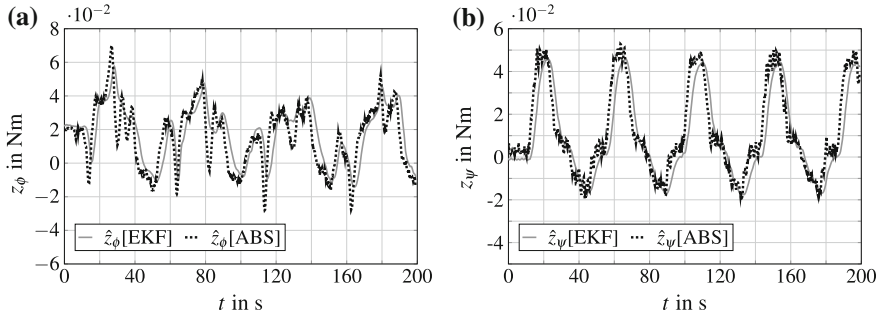


Fig. 13 Estimated lumped disturbance torques acting along the azimuth axis and the pitch axis. **a** Estimated torque along the azimuth axis. **b** Estimated torque along the pitch axis

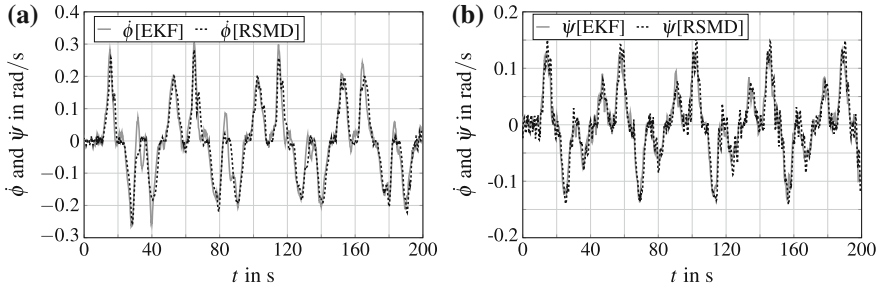


Fig. 14 Estimated angular velocities using EKF and robust sliding mode differentiator (RSMD). **a** Angular velocity along the azimuth axis. **b** Angular velocity along the pitch axis

Table 2 Evaluation criteria for the robust nonlinear control laws

Tracking error	MAE in rad	RMSE in rad
e_ϕ (BS-SMC)	0.0187	0.0258
e_ϕ (ABS)	0.0260	0.0381
e_ψ (BS-SMC)	0.0113	0.0173
e_ψ (ABS)	0.0175	0.0245

6 Conclusions and Future Work

In this contribution, a nonlinear control-oriented model is proposed for the tracking control design of a two rotor aerodynamic system (TRAS). The control-oriented modelling is based on Lagrange's equations. Two alternative nonlinear robust control approaches are investigated: a backstepping-based sliding mode control and an adaptive backstepping control. To deal with immeasurable state variables and unknown disturbances, a discrete-time EKF is employed in combination with the backstepping-based sliding mode control, whereas, the adaptive backstepping

control uses a robust sliding mode differentiator to estimate the angular velocities. In the latter case, adaptive laws are derived to estimate the lumped disturbance torques acting on the system. An important conclusion of this chapter is the derivation of the reaching condition for the backstepping-based sliding mode control. For both control approaches, the asymptotic stability of the overall closed-loop error dynamics is proven by Hurwitz's stability criterion. Moreover, both control structures are thoroughly investigated by means of simulations regarding the achievable tracking accuracy and the estimation quality. A successful experimental validation points out that both proposed approaches allow for an excellent tracking of the desired trajectories. Nevertheless, the backstepping-based sliding mode control in combination with a discrete-time EKF slightly outperforms the adaptive backstepping control during both the transient and the steady-state phases. The nonlinear control methods proposed in this contribution can be easily applied to a general class of control-affine system under the influence of the lumped matched disturbance.

In future work, the robust backstepping-based sliding mode control will be compared with higher order sliding mode controllers which can elegantly cope with the chattering phenomena. Furthermore, a comparison with a robust model predictive control strategy will be performed as well. From the control point of view, the proposed control strategies will be investigated for the nonlinear systems under the influence of both matched and mismatched lumped disturbances.

References

1. Ahmad S, Chipperfield A, Tokhi M (2000) Dynamic modeling and optimal control of a twin rotor MIMO system. In: Proceedings of the IEEE conference on national aerospace and electronics (NAECON), pp 391–398
2. Ahmad S, Shaheed M, Chipperfield A, Tokhi M (2000) Nonlinear modelling of a twin rotor MIMO system using radial basis function networks. In: Proceedings of the IEEE conference on national aerospace and electronics (NAECON), pp 313–320
3. Ahmed Q, Bhatti A, Iqbal S (2009) Nonlinear robust decoupling control design for twin rotor system. In: Proceedings of the 7th Asian Control Conference (ASCC), pp 937–942
4. Butt SS, Prabel R, Aschemann H (2014) Multi-variable flatness-based control of a two degrees of freedom helicopter. In: Proceedings of the international conference on control, decision and information technologies (CoDIT), pp 321–326
5. Butt SS, Prabel R, Aschemann H (2015) Multi-variable sliding mode control of a two degrees of freedom helicopter. In: Proceedings of the Vienna conference on mathematical modelling, (MATHMOD), pp 802–807
6. Chawda V, Celik O, O'Malley M (2011) Application of Levant's differentiator for velocity estimation and increased Z-width in haptic iinterfaces. In: Proceedings of the IEEE world haptics conference (WHC), pp 403–408
7. Darus I, Aldebrez F, Tokhi M (2004) Parametric modelling of a twin rotor system using genetic algorithms. In: Proceedings of the first international symposium on control, communications and signal processing, pp 115–118
8. Dutka A, Ordys A, Grimble M (2003) Non-linear predictive control of 2 DOF helicopter model. In: Proceedings of the 42nd IEEE conference on decision and control (CDC), vol 4, pp 3954–3959
9. Goldstein H, Poole P, Safko J (2002) Classical mechanics. Addison Wesley, San Francisco

10. INTECO (2013) Two rotor aero-dynamical system user's manual
11. Jung-Hua Y, Wen-Chun H (2009) Adaptive backstepping control for electrically driven unmanned helicopter. Control Eng Pract 17(8):903–913
12. Khalil HK (1996) Nonlinear systems. Prentice Hall, Upper Saddle River
13. Levant A (1998) Robust exact differentiation via sliding mode technique. Automatica 34(3):379–384
14. Levant A (2003) Higher-order sliding modes, differentiation and output-feedback control. Int J Control 76(9):924–941
15. Lopez-Martinez M, Ortega M, Rubio F (2003) An H_∞ controller for a double rotor system. In: IEEE conference on emerging technologies and factory automation (ETFA), vol 1, pp 253–259
16. Lopez-Martinez M, Diaz JM, Ortega M, Rubio F (2004) Control of a laboratory helicopter using switched 2-step feedback linearization. In: Proceedings of the American control conference (ACC), vol. 5, pp 4330–4335
17. Lopez-Martinez M, Vivas C, Ortega M (2005) A multivariable nonlinear H_∞ controller for a laboratory helicopter. In: Proceedings of the 44th IEEE conference on decision and control, European control conference (CDC-ECC), pp 4065–4070
18. Rahideh A, Shaheed MH (2007) Mathematical dynamic modelling of a twin-rotor multiple input-multiple output system. Proc Inst Mech Eng, Part I: J Syst Control Eng 221(1):89–101
19. Shaheed M (2004) Performance analysis of 4 types of conjugate gradient algorithms in the nonlinear dynamic modelling of a TRMS using feedforward neural networks. In: Proceedings of the IEEE international conference on systems, man and cybernetics, vol 6, pp 5985–5990
20. Slotine J, Li W (1991) Applied nonlinear control. Prentice-Hall Inc, Upper Saddle River
21. Stengel R (1994) Optimal control and estimation. Dover Publications Inc, New York
22. Utkin VI, Guldner J, Shi J (1999) Sliding mode control in electromechanical systems. The Taylor & Francis systems and control book series. Taylor & Francis, London
23. Young K, Utkin V, Ozguner U (1999) A control engineer's guide to sliding mode control. IEEE Trans Control Syst Technol 7(3):328–342

Robust Congestion Controller for a Single Virtual Circuit in Connection-Oriented Communication Networks

Piotr Leśniewski and Andrzej Bartoszewicz

Abstract In this contribution, we consider the problem of data flow control for a single virtual connection in communication networks. The connection is described by the maximum link capacity, the non-negligible propagation delay and an unknown, time-varying data loss rate. We propose a discrete-time sliding mode controller, which generates non-negative and upper bounded transmission rates. In addition, it ensures that the queue length in the bottleneck link buffer is always limited. Moreover, with a sufficiently large memory buffer in the bottleneck node, it guarantees full utilization of the available bandwidth. The controller uses a dead-beat sliding hyperplane in order to ensure fast response to unknown changes of the link capacity and to an unpredictable data loss rate. However, if the straightforward dead-beat paradigm was used, unacceptably large transmission rates would be generated. Therefore, we use the reaching law approach in this chapter to decrease excessive magnitudes of the control signal at the start of the control process.

1 Introduction

In connection-oriented communication networks, data units are almost never sent directly from their source to their destination. Instead, they must pass through a series of intermediate nodes. When one of those nodes cannot pass on all of the received data, due to the limited bandwidth of its outgoing link, a part of the incoming data is stored in the memory buffer of this node, awaiting for later transmission. Such an event is called a congestion. In order to minimize queueing delays and to maximize throughput, congestion controllers must be applied [4–6, 15, 17, 19, 24, 26]. The difficulty in the design of congestion control algorithms is caused by rapid changes of the available bandwidth, unpredictable packet losses, and long propagation delays.

P. Leśniewski (✉) · A. Bartoszewicz
Institute of Automatic Control, Technical University of Łódź,
18/22 Bohdana Stefanowskiego St., 90-924 Łódź, Poland
e-mail: piotr.lesniewski2@gmail.com

A. Bartoszewicz
e-mail: andrzej.bartoszewicz@p.lodz.pl

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*,
Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_2

When the congestion of a specific link is observed, an information must be sent to all sources that are transmitting data using this link. However, the delivery of this information is not instantaneous, but occurs after the feedback propagation delays. Then data sources adjust their transmission rates in order to reduce the congestion. The adjusted rates begin to affect the congested link after the propagation delay of the sent data. Thus, in modern communication networks that are characterized by time delays and large bandwidth the need for efficient data flow control algorithms cannot be neglected.

Sliding mode control is a widely recognized methodology [8, 10, 25, 27, 29], suitable for a large group of nonlinear, time-varying and uncertain systems. The main advantages of sliding mode control are its high computational performance and robustness [9]. This technique was first designed for continuous-time systems. However, as an overwhelming majority of control algorithms are nowadays implemented in digital hardware, discrete-time sliding mode control [1–3, 7, 11, 14, 18, 20, 23, 28, 31] is an interesting and up-to-date research field. The main concept of sliding mode control is to force the representative point (state) of the system from its initial position towards the sliding hyperplane. The period when the state approaches the hyperplane is called the reaching phase. The controller should then enforce a sliding mode, during which the representative point moves on ("slides" along) the hyperplane or in its vicinity.

The first step in designing a sliding mode control is selecting the parameters of the hyperplane so as to obtain the desired performance of the closed-loop system. This can be done in various ways such as dead-beat design, quadratic optimization or the pole placement method. Then, a controller that guarantees the convergence to the hyperplane and the stability of the sliding mode is designed. There are two approaches to solving this problem. The first one begins with proposing a control law, and then proving, that the properties mentioned above are guaranteed when it is applied. However, in this contribution we will use the other approach, which is based on the reaching law. Using this approach, one first defines the desired evolution of the sliding variable. Then, the sliding mode control that enforces this evolution is derived. This methodology was presented for continuous-time systems in [12], and then extended to discrete-time systems in [13]. The approach presented in [13] has then become quite popular among sliding mode control researchers [14, 16, 21, 22, 30].

In this work, we design a discrete-time reaching law-based sliding mode congestion controller for connection-oriented networks. During the design process, we take into account not only unpredictable bandwidth changes and inevitable propagation delays, but also time-varying transmission losses.

The remainder of this work is organized as follows. In Sect. 2 we present the model of the virtual circuit. Then, we design the reaching law-based sliding mode control in Sect. 3. In Sect. 4 we demonstrate and prove analytically the important properties of the presented controller. Section 5 contains the results of computer simulations, that verify the performance of the controller. Finally, Sect. 6 comprises the conclusions of this work.

2 Network Model

Let us analyze a virtual circuit of a connection-oriented network. The circuit consists of a data source, some intermediate nodes and a destination. Figure 1 depicts the scheme of the model. There is a single bottleneck in the considered network, and a congestion controller is placed at the bottleneck node. This controller generates a signal (denoted by u), that determines the transmission rate of the source. This signal arrives at the source after the backward delay T_{B}, and upon receiving it, the source transmits the specified amount of data. The data are passed from node to node, until, after the forward delay T_{F} they reach the bottleneck node. Inevitably, during transmission, some data packets are lost, and so only αu data arrive at the congested node, where

$$0 < \alpha_{\min} \leq \alpha \leq \alpha_{\max} \leq 1. \tag{1}$$

The round trip time (RTT), which is the delay between generating the control signal and the requested data arrival at the bottleneck node, can be expressed as the sum of the backward and forward propagation delays

$$RTT = T_{\mathrm{B}} + T_{\mathrm{F}}. \tag{2}$$

We denote the discretization period by T, the bottleneck queue length at time kT by $y(kT)$, and its demand value by y_{d}. There are no data in the buffer before the start of the control process, i.e. $y(kT < 0) = 0$. Furthermore, it is assumed that the round trip time is a multiple of the discretization period, i.e. $RTT = mT$, where m is a positive integer. Since the first data will arrive at the queue after RTT, then $y(kT \leq RTT) = 0$.

The available bandwidth of the bottleneck link is modelled as an a priori unknown non-negative function of time $d(kT)$. Only the maximum value of this function denoted by $d_{\max}$ is known in advance. As sometimes the amount of data stored in the bottleneck node can be insufficient to fully utilize the bandwidth, an additional function $h(kT)$ representing the amount of data actually leaving the buffer at time kT is introduced. Therefore,

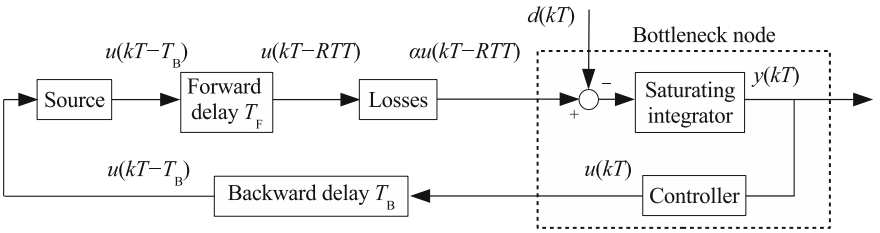


Fig. 1 Network model

$$0 \leq h(kT) \leq d(kT) \leq d_{\max} \tag{3}$$

for any $k \geq 0$.

We can represent the queue length as the difference between incoming and outgoing amounts of data, i.e.

$$y(kT) = \alpha \sum_{j=0}^{k-1} u(jT - RTT) - \sum_{j=0}^{k-1} h(jT) = \alpha \sum_{j=0}^{k-m-1} u(jT) - \sum_{j=0}^{k-1} h(jT). \tag{4}$$

The system can also be expressed using the standard state-space notation

$$\begin{aligned} \mathbf{x}[(k+1)T] &= \mathbf{A}\mathbf{x}(kT) + \boldsymbol{\delta}\mathbf{A}\mathbf{x}(kT) + \mathbf{b}u(kT) + \mathbf{o}h(kT) \\ y(kT) &= \mathbf{r}^T\mathbf{x}(kT), \end{aligned} \tag{5}$$

where $\mathbf{x}(kT) = [x_1(kT) \;\; x_2(kT) \ldots x_n(kT)]^T$ is the state vector, $y(kT) = x_1(kT)$ represents the queue length, and the remaining state variables are the delayed values of the control signal, i.e.

$$x_i(kT) = u[(k - n + i - 1)T] \tag{6}$$

for $i = 2, \ldots, n$, where $n = m + 1$. In (5) $\mathbf{A}$ is a $n \times n$ state matrix and $\boldsymbol{\delta}\mathbf{A}$ is a $n \times n$ model uncertainty matrix

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{\max} & 0 & & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & & 0 \end{bmatrix}, \quad \boldsymbol{\delta}\mathbf{A} = \begin{bmatrix} 0 & \delta\alpha & 0 & & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & & 0 \end{bmatrix}, \tag{7}$$

where $\delta\alpha \in [\alpha_{\min} - \alpha_{\max}, 0]$, and $\mathbf{b}$, $\mathbf{o}$, and $\mathbf{r}$ are the following $n \times 1$ vectors

$$\mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{o} = \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}. \tag{8}$$

The desired state of the system is denoted by $\mathbf{x}_{\mathrm{d}} = [y_{\mathrm{d}} \;\; 0 \;\; \cdots \;\; 0]^T$, where $y_{\mathrm{d}} = \text{const}$. The model may represent a single virtual circuit in Asynchronous Transfer Mode (ATM) or MultiProtocol Label Switching networks.

3 Non-Switching Reaching Law-Based SM Controller

We now derive a non-switching reaching law-based sliding mode controller, and apply it to the system described in the previous section. We start by choosing the sliding variable as

$$s(kT) = \mathbf{c}^T \mathbf{e}(kT), \tag{9}$$

where

$$\mathbf{e}(kT) = \mathbf{x}_\mathrm{d} - \mathbf{x}(kT) \tag{10}$$

is the closed-loop control error. With this choice of the switching variable s the equation $s(kT) = 0$ determines the sliding hyperplane. The vector $\mathbf{c}$ is chosen so that $\mathbf{c}^T\mathbf{b} \neq 0$ and the closed-loop system exhibits the desired performance. We select the vector $\mathbf{c}$ so that the closed-loop system exhibits dead-beat dynamics to obtain finite-time error convergence to zero. We begin by calculating the value of the sliding variable in the next time instant. As we are interested in the poles of the closed-loop system, we omit the disturbance and the modelling uncertainty terms. Using Eqs. (5) and (9) we arrive at

$$s[(k+1)T] = \mathbf{c}^T \left[\mathbf{x}_\mathrm{d} - \mathbf{A}\mathbf{x}(kT) - \mathbf{b}u(kT)\right]. \tag{11}$$

We find that the control signal that satisfies $s[(k+1)T] = 0$ is given by

$$u(kT) = \left(\mathbf{c}^T\mathbf{b}\right)^{-1} \mathbf{c}^T \left[\mathbf{x}_\mathrm{d} - \mathbf{A}\mathbf{x}(kT)\right]. \tag{12}$$

Let us point out, that control signal (12) is calculated only to find the form of vector $\mathbf{c}$ that ensures dead-beat dynamics in sliding mode. This signal is not used in the final reaching law-based sliding mode controller. By substituting the control signal (12) into (5) we obtain the closed-loop system matrix as

$$\mathbf{A}_\mathrm{c} = \left[\mathbf{I}_\mathrm{n} - \mathbf{b}(\mathbf{c}^T\mathbf{b})^{-1}\mathbf{c}^T\right]\mathbf{A}. \tag{13}$$

The matrix $\mathbf{A}_\mathrm{c}$ has the following characteristic polynomial

$$\det\left(z\mathbf{I}_\mathrm{n} - \mathbf{A}_\mathrm{c}\right) = z^n + \frac{c_{n-1} - c_n}{c_n} z^{n-1} + \cdots + \frac{c_2 - c_3}{c_n} z^2 + \frac{\alpha_{\max} c_1 - c_2}{c_n} z. \tag{14}$$

A discrete-time system is asymptotically stable if and only if all of its eigenvalues are located inside the unit circle. Moreover, to obtain dead-beat performance all of the eigenvalues must be located in the origin of the z-plane. Therefore, the characteristic polynomial must have the following form

$$\det\left(z\mathbf{I}_{\mathrm{n}} - \mathbf{A}_{\mathrm{c}}\right) = kz^{n}, \tag{15}$$

where k is an arbitrary constant. We find that (14) simplifies to (15) when the following vector $\mathbf{c}$ is chosen

$$\mathbf{c} = \begin{bmatrix} \dfrac{1}{\alpha_{\max}} & 1 & 1 & \cdots & 1 \end{bmatrix}^{T}. \tag{16}$$

The aim of the controller is to decrease the absolute value of the switching variable until it reaches a band around $s(kT) = 0$, which is further called the quasi-sliding band. After entering this band, the switching variable should never leave it again. Contrary to some previous works [13], in the chosen definition of the quasi-sliding mode, crossing the hyperplane is allowed but not necessary. As it will be demonstrated later, this modification allows us to eliminate the undesirable phenomenon of chattering.

After determining the appropriate sliding hyperplane coefficients we now consider the following reaching law, which describes the desired evolution of the sliding variable

$$s[(k+1)T] = \{1 - q[s(kT)]\}s(kT) - \tilde{F}(kT) - \tilde{S}(kT) + F_1, \tag{17}$$

where

$$\tilde{S}(kT) = \mathbf{c}^{T}\boldsymbol{\delta}\mathbf{A}\mathbf{x}(kT) = \frac{\delta\alpha x_2(kT)}{\alpha_{\max}} \tag{18}$$

is the impact of the model uncertainty (the unknown and time-varying transmission losses). Function

$$\tilde{F}(kT) = \mathbf{c}^{T}\mathbf{o}h(kT) = -\frac{h(kT)}{\alpha_{\max}} \tag{19}$$

represents the influence of the disturbance (in our case the outgoing flow of data) on the sliding variable. The term

$$F_1 = -\frac{d_{\max}}{2\alpha_{\max}} \tag{20}$$

is employed to compensate for the mean value of $\tilde{F}(kT)$. The term $S_1 = \frac{S_U + S_L}{2}$ (where S_U and S_L are the upper and lower bounds of $\tilde{S}(kT)$), used in some earlier works [13] to compensate for the mean value of the model uncertainty, could lead to generating negative control signals. As in the considered system this is not feasible, we discard this term. The variable convergence rate $q[s(kT)]$ is given by

$$q[s(kT)] = \frac{s_0}{s_0 + |s(kT)|}, \tag{21}$$

where

$$s_0 > \frac{d_{\max}(2\alpha_{\max} - \alpha_{\min})}{2\alpha_{\max}\alpha_{\min}} \tag{22}$$

is a design parameter, that must satisfy the above inequality in order to ensure the convergence to the vicinity of the sliding hyperplane. This is another modification of the reaching law developed by Gao et al. [13] as in their work the parameter q was constant. As we can observe from (21), the parameter $q[s(kT)]$ decreases with the increase of $|s(kT)|$. Therefore, the value of the control signal needed to ensure the desired sliding variable evolution (17) will not change so dramatically for different absolute values of the sliding variable. This allows us to find a better compromise between acceptable control signal values when $|s(kT)|$ is large and fast convergence in the vicinity of the sliding hyperplane.

We now calculate the control signal that makes the sliding variable actually evolve according to (17). We will start by using (5) to rewrite (9) as

$$s[(k+1)T] = \mathbf{c}^T \mathbf{x}_{\mathrm{d}} - \mathbf{c}^T \left[\mathbf{A}\mathbf{x}(kT) + \boldsymbol{\delta}\mathbf{A}\mathbf{x}(kT) + \mathbf{b}u(kT) + \mathbf{o}h(kT)\right]. \tag{23}$$

By comparing the right-hand sides of (17) and (23) we obtain

$$u(kT) = (\mathbf{c}^T\mathbf{b})^{-1}\left\{q[s(kT)]s(kT) + \frac{d_{\max}}{2\alpha_{\max}} - \mathbf{c}^T(\mathbf{A} - \mathbf{I}_{\mathrm{n}})\mathbf{x}(kT)\right\}. \tag{24}$$

We notice, that by choosing $\mathbf{c}$ according to (16), we have obtained

$$\mathbf{c}^T(\mathbf{A} - \mathbf{I}_{\mathrm{n}}) = [0 \quad \ldots \quad 0]. \tag{25}$$

Using (8), (16), (21) and (25) we can rewrite (24) as

$$u(kT) = \frac{s_0 s(kT)}{s_0 + |s(kT)|} + \frac{d_{\max}}{2\alpha_{\max}}. \tag{26}$$

In this way, we have completed the design of the reaching law-based sliding mode congestion controller. It will be shown in the next section that control strategy (26) as opposed to (12) ensures upper bounded transmission data rates, which is a highly desirable property in the application considered in this chapter.

4 Properties of the System

In this section we demonstrate that the application of the proposed controller guarantees several important properties of the considered system.

Theorem 1 *The sliding variable satisfies*

$$s(kT) \geq -\frac{d_{\max} s_0}{2\alpha_{\max} s_0 - d_{\max}} \tag{27}$$

for all $k \geq 0$.

Proof The sliding variable in the initial time instant is equal to $y_{\mathrm{d}}/\alpha_{\max}$, and thus satisfies (27). Therefore, in order to prove the theorem, it is sufficient to demonstrate that if (27) holds for every $k \leq l$, then it also holds for $k = l + 1$.

We observe from (26) that the control signal value always increases with the increase of $s(kT)$. Therefore, substituting the right-hand side of (27) into (26), we observe, that $u(kT) \geq 0$ for all $k \leq l$. This in turn implies that $\tilde{S}(kT) \leq 0$ for all $k \leq l$. Moreover, we observe from (19) that $\tilde{F}(kT) \leq 0$ for all $k \geq 0$.

Substituting the right-hand side of (27) and the worst-case scenario values $\tilde{F}(kT) = 0$, $\tilde{S}(kT) = 0$ into (17), we notice that indeed if (27) holds for all $k \leq l$ then it must also hold for $k = l + 1$. Using this observation and the principle of mathematical induction we conclude, that (27) will hold for all $k \geq 0$.

Since (27) holds for all $k \geq 0$ then, following the reasoning from Theorem 1, the control signal is always non-negative. In the next theorem we show that this control signal is also upper bounded by an a priori known value. As the amount of data sent by the source during a single discretization period is limited by the capacity of the outgoing link (and obviously cannot be negative) both of these properties are crucial for applying the flow controller in a real network.

Theorem 2 *The control signal satisfies the inequality*

$$u(kT) \leq \frac{y_d s_0}{s_0 \alpha_{\max} + y_d} + \frac{d_{\max}}{2\alpha_{\max}} \tag{28}$$

for any $k \geq 0$.

Proof The outgoing amount of data $h(kT)$ is defined in such a way that $x_1(kT) \geq 0$ for any $k \geq 0$. Furthermore, since the control signal is always non-negative, then we notice from (6) that $x_i(kT) \geq 0$ for $i = 2, \ldots, n$ and all $k \geq 0$. Using these observations together with (9) and (16), we see that the sliding variable cannot exceed its initial value, i.e.

$$s(kT) \leq \frac{y_{\mathrm{d}}}{\alpha_{\max}} \tag{29}$$

for all $k \geq 0$. As already stated the control signal always increases with the increase of $s(kT)$. Therefore, by substituting the right-hand side of inequality (29) into (26) we obtain the greatest possible value of $u(kT)$ and find that (28) indeed holds.

Theorem 3 *Once the inequality*

$$s(kT) \leq \frac{s_0\left[(\alpha_{\max}-\alpha_{\min})\left(\frac{y_d s_0}{s_0\alpha_{\max}+y_d}+\frac{d_{\max}}{2\alpha_{\max}}\right)+\frac{d_{\max}}{2}\right]}{\alpha_{\max}s_0-\left[(\alpha_{\max}-\alpha_{\min})\left(\frac{y_d s_0}{s_0\alpha_{\max}+y_d}+\frac{d_{\max}}{2\alpha_{\max}}\right)+\frac{d_{\max}}{2}\right]} \tag{30}$$

is satisfied, it remains true for the remainder of the control process.

Proof As $x_2(kT)$ is a delayed value of the control signal, using (28) with (18) we get

$$\tilde{S}(kT) \geq \frac{\alpha_{\min}-\alpha_{\max}}{\alpha_{\max}}\left(\frac{y_{\mathrm{d}} s_0}{s_0\alpha_{\max}+y_{\mathrm{d}}}+\frac{d_{\max}}{2\alpha_{\max}}\right) \tag{31}$$

for all $k \geq 0$. Furthermore, we notice from (19) that

$$\tilde{F}(kT) \geq -\frac{d_{\max}}{\alpha_{\max}} \tag{32}$$

also for all $k \geq 0$.

Substituting these lower bounds of $\tilde{S}(kT)$ and $\tilde{F}(kT)$ and the right-hand side of inequality (30) into (26), we conclude that indeed once (30) holds, it will remain true for the remainder of the control process.

Remark 1 Taking into account the results of Theorems 1 and 3, we observe that the sliding variable will converge to the quasi-sliding mode band defined by

$$\begin{aligned}\frac{-d_{\max}s_0}{2\alpha_{\max}s_0-d_{\max}} &\leq s(kT)\\ &\leq \frac{s_0\left[(\alpha_{\max}-\alpha_{\min})\left(\frac{y_{\mathrm{d}} s_0}{s_0\alpha_{\max}+y_{\mathrm{d}}}+\frac{d_{\max}}{2\alpha_{\max}}\right)+\frac{d_{\max}}{2}\right]}{\alpha_{\max}s_0-\left[(\alpha_{\max}-\alpha_{\min})\left(\frac{y_{\mathrm{d}} s_0}{s_0\alpha_{\max}+y_{\mathrm{d}}}+\frac{d_{\max}}{2\alpha_{\max}}\right)+\frac{d_{\max}}{2}\right]},\end{aligned} \tag{33}$$

and after entering it, it will never leave it again.

Buffer overflows occur when the incoming data and the current amount of stored data exceed the buffer capacity of the congested node. As this event leads to data losses and causes the need for retransmission, it is highly undesirable in communication networks. In the next theorem we determine the size, which the queue length at the bottleneck node will never exceed. If the buffer capacity is equal to or greater than this value, then there is no risk of buffer overflows.

Theorem 4 *The queue length never exceeds the following value*

$$y(kT) \leq y_d + \frac{d_{\max}s_0}{2s_0-d_{\max}/\alpha_{\max}}. \tag{34}$$

Proof Using (9) we can rewrite (27) as

$$y(kT) \leq y_{\mathrm{d}} + \frac{d_{\max} s_0}{2s_0 - d_{\max}/\alpha_{\max}} - \alpha_{\max} \sum_{i=2}^{n} x_i(kT). \tag{35}$$

All of the state variables, except for the first one are the delayed values of the control signal. As already demonstrated, the control signal is always non-negative, therefore (35) implies (34).

In order for the flow control algorithm to be efficient, it should result in the greatest throughput that is possible in the network. In Theorem 5, we calculate the smallest value of the demand queue length that guarantees, that the queue length remains always strictly positive, after the first data reach it. This property corresponds to 100 % utilization of the available bandwidth.

Theorem 5 *If the demand queue length satisfies*

$$y_d > \frac{\alpha_{\max} s_0 d_{\max}(2\alpha_{\max} - \alpha_{\min})}{2\alpha_{\max}\alpha_{\min} s_0 - d_{\max}(2\alpha_{\max} - \alpha_{\min})} + \frac{\alpha_{\max} d_{\max}(n-1)}{\alpha_{\min}}, \tag{36}$$

then the queue length will be strictly positive for any $k \geq n$.

Proof To keep the notation clear, we introduce

$$\theta = \frac{2\alpha_{\max} - \alpha_{\min}}{2\alpha_{\max}\alpha_{\min}}. \tag{37}$$

In the proof we will consider two possible ranges of the sliding variable separately

Case 1 If

$$s[(k-n+1)T] > \frac{s_0 d_{\max}\theta}{s_0 - d_{\max}\theta}, \tag{38}$$

then the queue length for any $k \geq n-1$ can be calculated, using (4) as

$$\begin{aligned} y[(k+1)T] &= y(kT) + \alpha u[(k-n+1)T] - h(kT) \\ &\geq y(kT) + \alpha_{\min}\left[\frac{s_0 s[(k-n+1)T]}{s_0 + |s[(k-n+1)T]|} + \frac{d_{\max}}{2\alpha_{\max}}\right] - d_{\max}. \end{aligned} \tag{39}$$

The expression in square brackets in (39) always increases with the increase of $s[(k-n+1)T]$. Therefore, substituting (38) we get

$$\begin{aligned} y[(k+1)T] > (s_0 - d_{\max}\theta)\frac{d_{\max}\theta}{s_0 - d_{\max}\theta} + \frac{d_{\max}}{2\alpha_{\max}} \\ + y(kT) - d_{\max} = y(kT) + d_{\max} - d_{\max} \geq 0. \end{aligned} \tag{40}$$

Case 2 On the other hand, if

$$s[(k-n+1)T] \leq \frac{s_0 d_{\max}\theta}{s_0 - d_{\max}\theta}, \tag{41}$$

then the queue length for $k \geq n-1$ can be expressed as

$$\begin{aligned} y[(k+1)T] &= y[(k-n+1)T] + \alpha \sum_{i=2}^{n} x_i[(k-n+1)T] - \sum_{i=k-n}^{k} h(iT) \\ &\geq y[(k-n+1)T] + \alpha_{\min} \sum_{i=2}^{n} x_i[(k-n+1)T] - (n-1)d_{\max}. \end{aligned} \tag{42}$$

Furthermore, we obtain from (9) and (16) that

$$s(kT) = \frac{y_{\mathrm{d}}}{\alpha_{\max}} - \frac{y(kT)}{\alpha_{\max}} - \sum_{i=2}^{n} x_i(kT). \tag{43}$$

Multiplying both sides of (43) by $\alpha_{\min}$ and rearranging the terms one obtains

$$\frac{\alpha_{\min} y(kT)}{\alpha_{\max}} + \alpha_{\min} \sum_{i=2}^{n} x_i(kT) = \frac{\alpha_{\min} y_{\mathrm{d}}}{\alpha_{\max}} - \alpha_{\min} s(kT). \tag{44}$$

As $\alpha_{\min}/\alpha_{\max} \leq 1$, and the queue length cannot be negative, we can observe, that

$$y(kT) + \alpha_{\min} \sum_{i=2}^{n} x_i(kT) \geq \frac{\alpha_{\min} y_{\mathrm{d}}}{\alpha_{\max}} - \alpha_{\min} s(kT). \tag{45}$$

Substituting (45) and (41) into (42), we get

$$y[(k+1)T] \geq \frac{\alpha_{\min} y_{\mathrm{d}}}{\alpha_{\max}} - \alpha_{\min} \frac{s_0 d_{\max}\theta}{s_0 - d_{\max}\theta}. \tag{46}$$

We find, that if (36) holds, then the right-hand side of the above inequality is always strictly positive. This conclusion, together with the result of Case 1 finalizes the proof.

5 Simulation Results

To verify the properties of the proposed control strategy, computer simulations were performed. The round trip time RTT =12 ms and the discretization period $T = 1$ ms. Therefore, $m = 12$ and $n = 13$. The packet loss ratio is bounded by $\alpha_{\min} = 0.82$ and

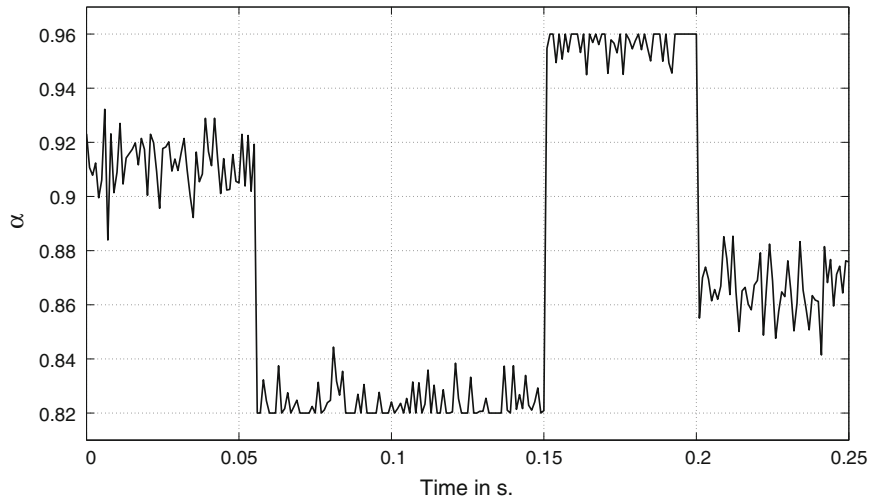


Fig. 2 Data loss rate

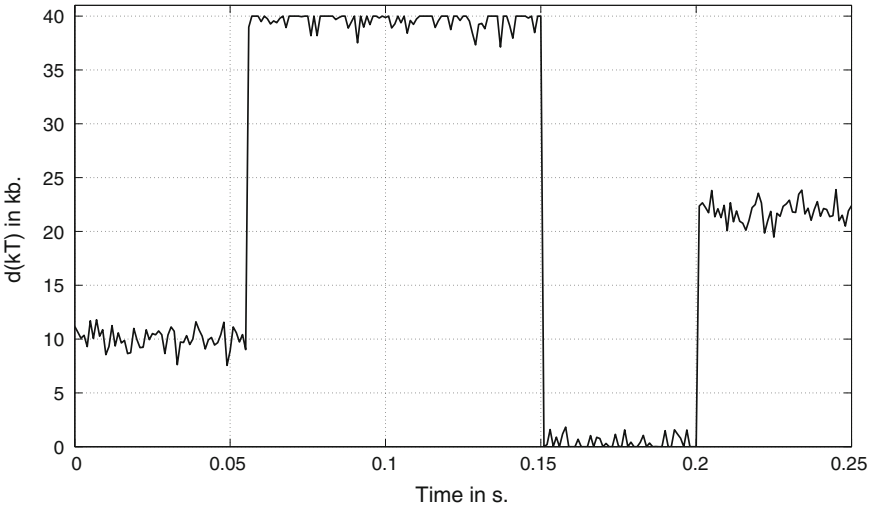


Fig. 3 Available bandwidth

$\alpha_{\max} = 0.96$, and the actual transient of the loss ratio is depicted in Fig. 2. The parameter $d_{\max} = 40$ kb and the available bandwidth chosen for the simulation is shown in Fig. 3. Both of the unpredictable functions exhibit rapid changes between small and large values which correspond to the most difficult conditions that could arise in the system. Note that we have taken into account both of the worst-case scenario combinations, namely

- $d(kT)$ is close to its maximum value, while α is close to minimum between 0.05 and 0.15 s
- $d(kT)$ is close to its minimum value, while α is close to maximum between 0.15 and 0.2 s.

Let us point out that irrespective of the chosen control algorithm, the upper bound of the transmission rate of the data source must satisfy

$$u_{\max} \geq \frac{d_{\max}}{\alpha_{\min}} \tag{47}$$

in order to fully utilize the available bandwidth of the bottleneck node. This can be observed as follows: imagine a situation, in which the available bandwidth is equal to $d_{\max}$, and the transmission loss parameter to $\alpha_{\min}$. Generating a constant control signal, that would be smaller then the right-hand side of (47) would result in less data arriving at the buffer, then leaving it in every discretization period. In this way, the buffer would become empty after some time, and the available bandwidth would not be fully utilized.

The choice of controller parameters y_{d} and s_0 will now be considered. The generated control signal will never exceed the right-hand side of inequality (28). As this signal is upper bounded by $u_{\max}$ we obtain the following condition on the controller parameters

$$\frac{y_{\mathrm{d}} s_0}{s_0 \alpha_{\max} + y_{\mathrm{d}}} \leq u_{\max} - \frac{d_{\max}}{2\alpha_{\max}}. \tag{48}$$

The expression $s_0\alpha_{\max} + y_{\mathrm{d}}$ is always positive, therefore it follows from (48) that

$$y_{\mathrm{d}} \left(s_0 - u_{\max} + \frac{d_{\max}}{2\alpha_{\max}} \right) \leq \left(u_{\max} - \frac{d_{\max}}{2\alpha_{\max}} \right) s_0 \alpha_{\max}. \tag{49}$$

Depending on the choice of s_0, the term $s_0 - u_{\max} + \frac{d_{\max}}{2\alpha_{\max}}$ can be positive, negative or equal to zero. We will now analyze these three cases separately.

Case 1 If

$$s_0 > u_{\max} + \frac{d_{\max}}{2\alpha_{\max}}, \tag{50}$$

then it follows from (49) that parameter y_{d} must satisfy

$$y_{\mathrm{d}} \leq \frac{\left(u_{\max} - \frac{d_{\max}}{2\alpha_{\max}} \right) s_0 \alpha_{\max}}{s_0 - u_{\max} + \frac{d_{\max}}{2\alpha_{\max}}} \tag{51}$$

so that the control signal will never exceed $u_{\max}$.

Case 2 If

$$s_0 < u_{\max} + \frac{d_{\max}}{2\alpha_{\max}}, \tag{52}$$

then parameter y_{d} must satisfy

$$y_{\mathrm{d}} \geq \frac{\left(u_{\max} - \frac{d_{\max}}{2\alpha_{\max}}\right) s_0 \alpha_{\max}}{s_0 - u_{\max} + \frac{d_{\max}}{2\alpha_{\max}}} \tag{53}$$

so that the maximum value of the control signal will not exceed $u_{\max}$. It is easy to observe using (47) and (52), that the right-hand side of inequality (53) is strictly negative. Therefore, any positive value of the demand queue length y_{d} satisfies this inequality.

Case 3 We now consider

$$s_0 = u_{\max} + \frac{d_{\max}}{2\alpha_{\max}}. \tag{54}$$

Using (47) we notice that the right-hand side of (49) is strictly positive. Therefore, for s_0 given by (54) any value of y_d ensures that the control signal will not exceed $u_{\max}$.

To sum up the three cases: for all $s_0 \leq u_{\max} + \frac{d_{\max}}{2\alpha_{\max}}$ any positive value of y_d ensures that the control signal will not exceed $u_{\max}$. On the other hand, if $s_0 > u_{\max} + \frac{d_{\max}}{2\alpha_{\max}}$, then the demand queue length y_{d} must be selected according to (51).

It is assumed that the source can send a maximum of 54 kb of data in a single discretization period. Moreover, inequality (36) must be satisfied to achieve full bandwidth utilization. The choice of parameters s_0 and y_{d} is depicted in Figs. 4 and 5. The dashed black line in Fig. 4 corresponds to $s_0 = u_{\max} + \frac{d_{\max}}{2\alpha_{\max}}$. In both figures the solid black line reflects the smallest y_{d} that satisfies condition (36), while the dotted black line represents the greatest value of y_{d} that for a given s_0 ensures, that the control signal will not exceed $u_{\max} = 54$ kb. Therefore, to achieve maximum throughput and at the same time not exceed the data source capabilities, one should select a combination of parameters that is below the dotted black line and above the solid black one. The set of admissible parameter combinations is marked in grey. The last question is which particular point from this set should be chosen. It is clear from inequality (34) that the maximum bottleneck buffer queue length decreases with the increase of s_0 and decrease of y_{d}. Therefore, to reduce memory requirements, one should select the combination of parameters that is close to the intersection of the solid and dotted lines. Motivated by this reasoning, we have selected $s_0 = 34.72$ kb and $y_{\mathrm{d}} = 710$ kb.

The results of the simulation are shown in Figs. 6, 7 and 8. Figure 6 shows the control signal. As predicted by Theorems 1 and 2, it never exceeds 54 kb and is always non-negative. Furthermore, when comparing Figs. 3 and 6 we observe, that although the control signal tracks the value of the available bandwidth, the rapid changes have

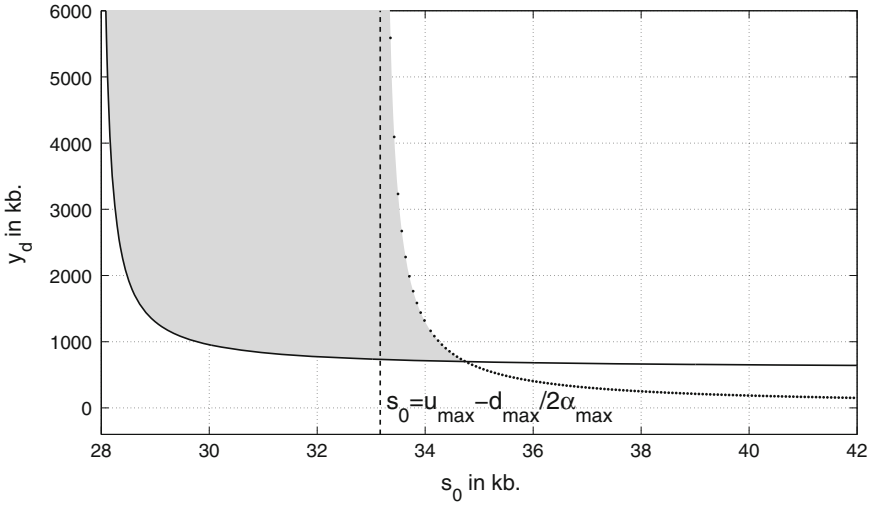


Fig. 4 The admissible set of parameter combinations for the first simulation scenario

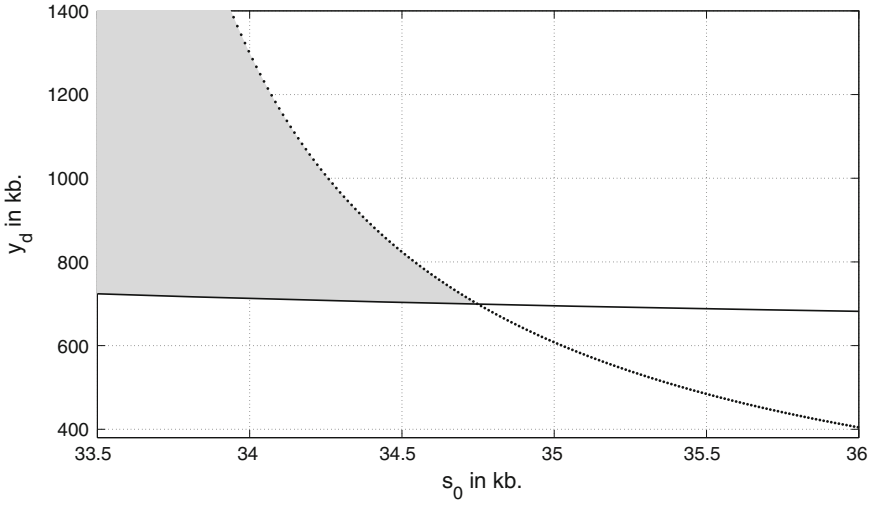


Fig. 5 The enlarged neighbourhood of the selected parameter values

been smoothed out, which is advantageous for the general transmission efficiency in the network. The bottleneck queue length is shown in Fig. 7. It never exceeds the value of 760 kb predicted by Theorem 4 and as shown in Theorem 5, after the first data reach it, it never drops to zero. Therefore, the risk of buffer overflow is eliminated and full utilization of the available bandwidth is guaranteed. Figure 8 depicts the evolution of the sliding variable. As predicted by Theorem 1 it is always greater

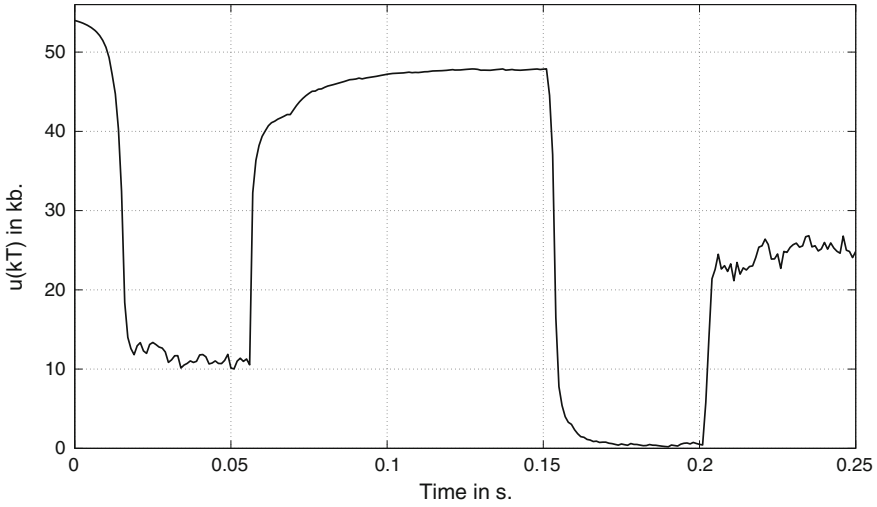


Fig. 6 Control signal

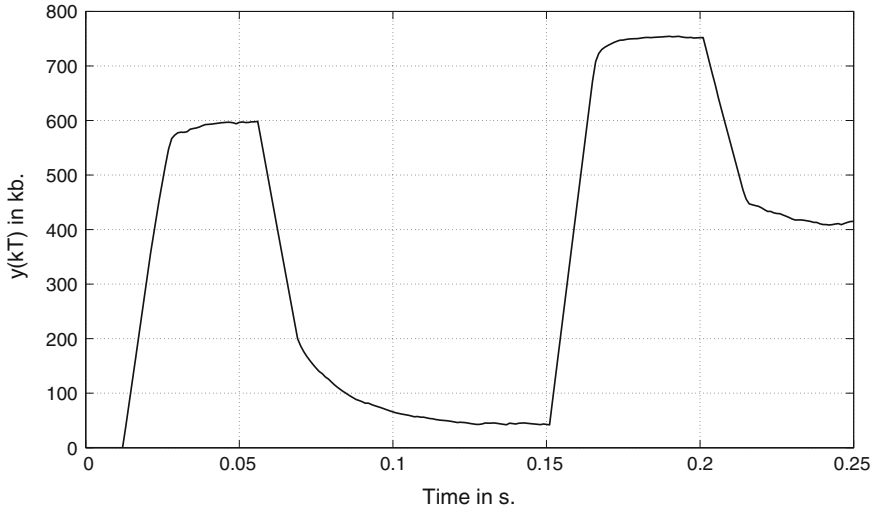


Fig. 7 Bottleneck node queue length

than -52.08 kb, and as demonstrated in Theorem 3 once it drops below 370 kb it never exceeds this value again.

In the second simulation scenario the available memory capacity at the bottleneck node is insufficient to enable full bandwidth utilization. The buffer capacity is $y_{\max} = 620$ kb. Therefore, in order to prevent buffer overflows, we have to select such values

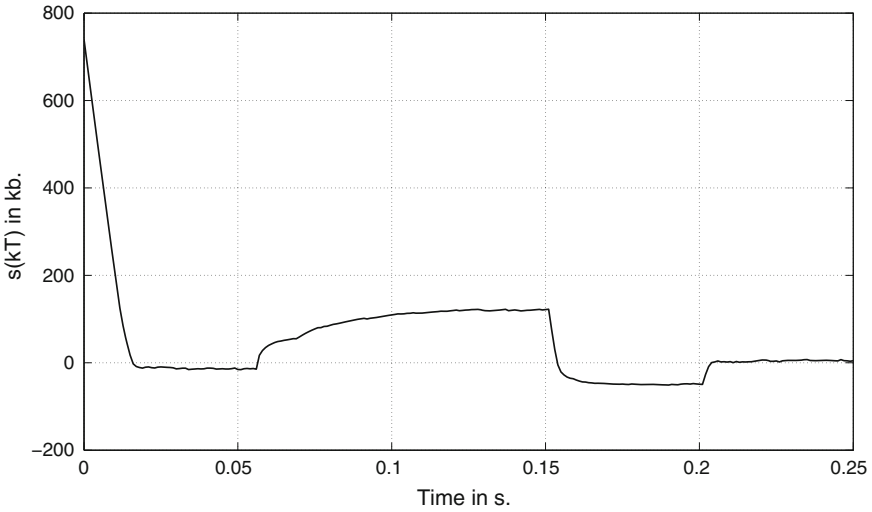


Fig. 8 Sliding variable evolution

of y_d and s_0 that the right-hand side of inequality (34) does not exceed y_{max}. This corresponds to satisfying the following condition

$$y_d \leq y_{max} - \frac{d_{max} s_0}{2s_0 - d_{max}/\alpha_{max}} \tag{55}$$

The choice of parameters for this simulation scenario is presented in Fig. 9. As in the previous case, the dotted black line represents the greatest value of y_d that prevents from exceeding the admissible control signal range, and the solid black line corresponds to the smallest values of y_d that satisfy condition (36). Moreover, the dashed black line corresponds to points in which the right-hand side of (55) is equal to the left-hand side of this inequality. Therefore, one should select parameter values that are below this line to eliminate the risk of buffer overflows. As we can observe from Fig. 9 it is impossible to select a combination of parameters which satisfies all the three requirements at the same time, namely there are no points that are at the same time below both the dotted black line and the dashed black line and above the solid black line. Therefore, in order to prevent buffer overflows, we select the intersection of the dashed line with the dotted black line. In this way we fully utilize the available control signal range without exceeding it. The risk of buffer overflows is also eliminated. However, as this intersection point lies below the solid black line, we do not obtain full utilization of the available bandwidth. This will be apparent in the simulation results.

The results of simulations are shown in Figs. 10, 11, 12 and 13. Figure 10 depicts the control signal, which again is always non-negative and never exceeds the available value of 54 kb. When compared to the previous scenario, we observe, that the changes

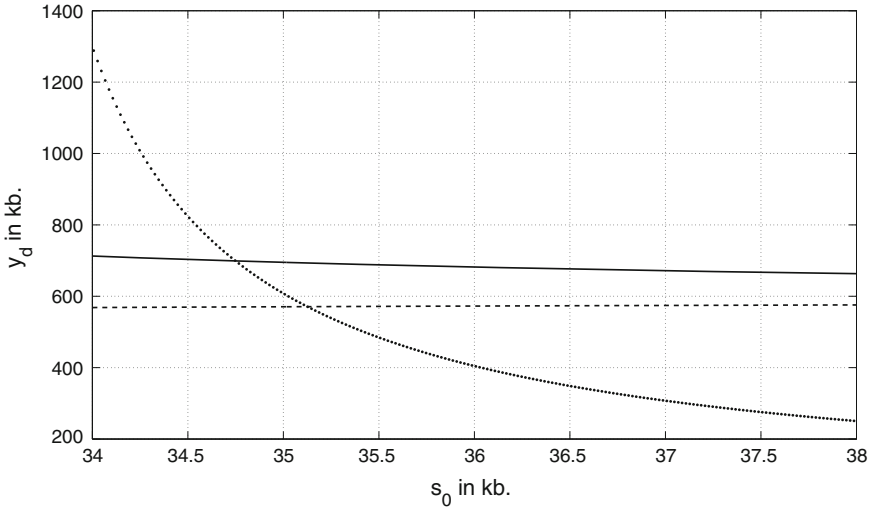


Fig. 9 The choice of parameters for the reduced bottleneck node memory capacity

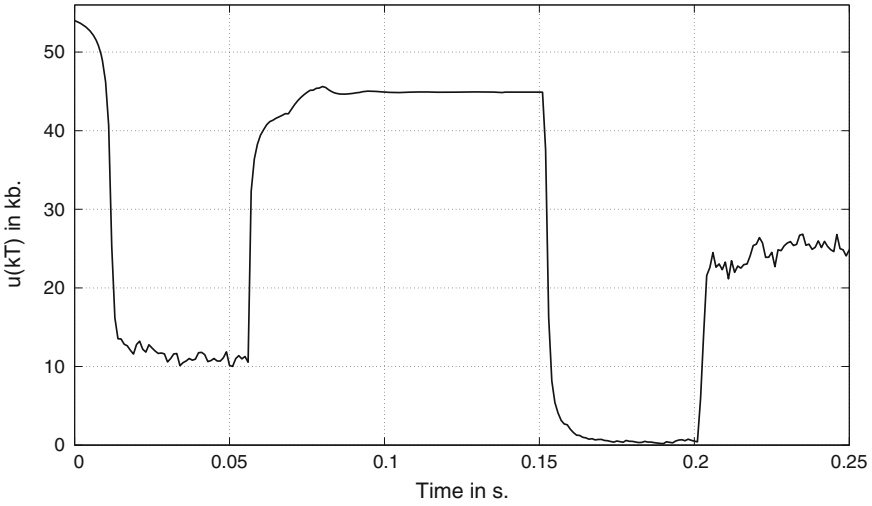


Fig. 10 Control signal for the reduced bottleneck node memory capacity

of the control signal are slightly more rapid. This is due to the fact, that the smaller available memory capacity can be used to smooth the changes in bandwidth to a lesser extent. The bottleneck node queue length is shown in Fig. 11. It never exceeds the value of 620 kb predicted by Theorem 4, but in contrast to the previous case, because of the insufficient memory capacity, at some times the queue is empty. When this occurs, the available bandwidth is not fully utilized, and we depict the unused part of the bandwidth in Fig. 12. The unused bandwidth at the start of the control process

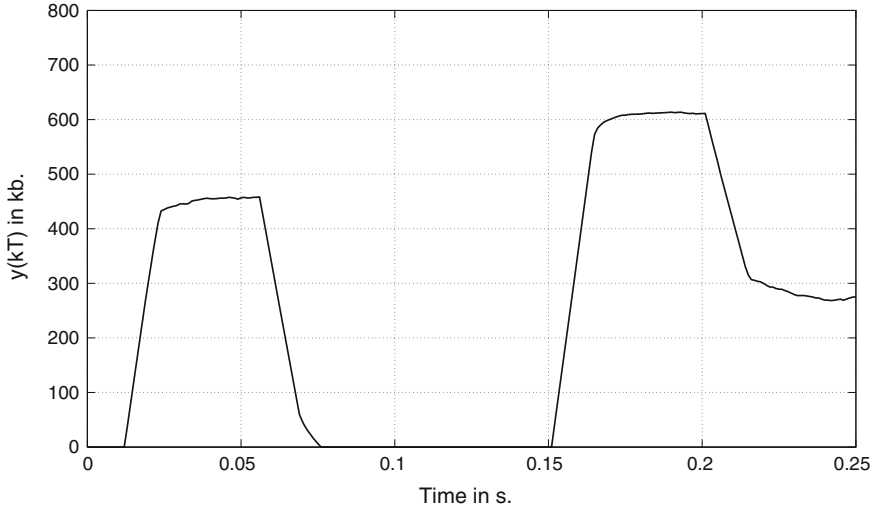


Fig. 11 Bottleneck node queue length for the reduced bottleneck node memory capacity

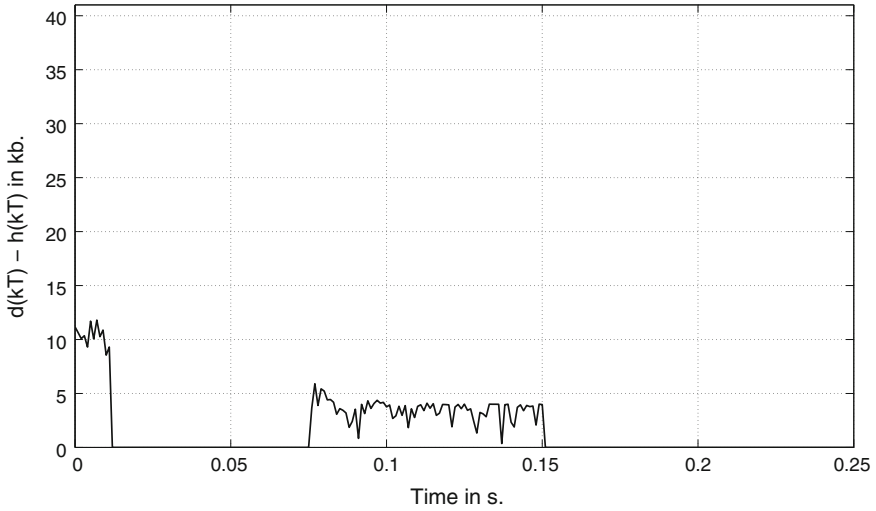


Fig. 12 Unused bandwidth for the reduced bottleneck node memory capacity

occurs before the first data arrive at the bottleneck node, and therefore cannot be eliminated by any control strategy. As we can notice from Fig. 12 the bandwidth after the first round trip time passes is still utilized fairly well, despite the limited buffer size. Figure 13 shows the evolution of the sliding variable. Again, as predicted in Sect. 4, it never drops below -51.14 kb and once it gets smaller than 328 kb it never exceeds this value again.

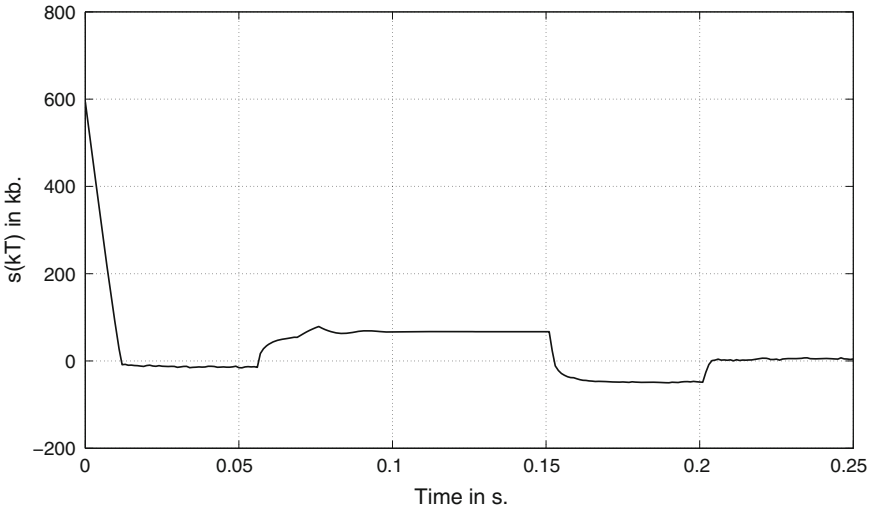


Fig. 13 Sliding variable evolution for the reduced bottleneck node memory capacity

6 Conclusions

In this chapter we have presented a robust sliding mode control strategy for congestion control of a single virtual circuit in connection-oriented communication networks. We have applied the reaching law methodology and the dead-beat control technique in designing the control law. The presented controller enforces a chattering free quasi-sliding mode. Furthermore, it eliminates the risk of data losses, and can ensure full utilization of the bottleneck link bandwidth, in spite of rapid, unpredictable changes of the transmission loss rate and the available bandwidth. In our future work we will extend the presented approach to the case of multisource data transmission networks with varying time delays.

Acknowledgments This work has been performed in the framework of a project "Optimal sliding mode control of time delay systems" financed by the National Science Centre of Poland decision number DEC 2011/01/B/ST7/02582. Kind support provided by the Foundation for Polish Science under Mistrz grant is also acknowledged.

References

1. Abidi K, Xu JX, Yu X (2007) On the discrete-time integral sliding-mode control. IEEE Trans Autom Control 52(4):709–715
2. Bandyopadhyay B, Fulwani D (2009) High-performance tracking controller for discrete plant using nonlinear sliding surface. IEEE Trans Ind Electron 56(9):3628–3637

3. Bartolini G, Ferrara A, Utkin V (1995) Adaptive sliding mode control in discrete-time systems. Automatica 31:769–773
4. Bartoszewicz A (2006) Nonlinear flow control strategies for connection oriented communication networks. Proc IEE Part D: Control Theory Appl 153(1):21–28
5. Bartoszewicz A, Leśniewski P (2014) An optimal sliding mode congestion controller for connection-oriented communiction networks with lossy links. Int J Appl Math Comput Sci 24(1):87–97
6. Bartoszewicz A, Leśniewski P (2014) Reaching law-based sliding mode congestion control for communication networks. IET Control Theory Appl 8(17):1914–1920
7. Corradini ML, Orlando G (1998) Variable structure control of discretized continuous-time systems. IEEE Trans Autom Control 43(9):1329–1334
8. DeCarlo RS, Żak S, Mathews G (1988) Variable structure control of nonlinear multivariable systems: a tutorial. Proc of the IEEE 76:212–232
9. Draženović B (1969) The invariance conditions in variable structure systems. Automatica 5:287–295
10. Edwards C, Spurgeon S (1998) Sliding mode control: theory and applications. Taylor & Francis, London
11. Galias Z, Yu X (2008) Analysis of zero-order holder discretization of two-dimensional sliding mode control systems. IEEE Trans Circuits Syst II 55(12):1269–1273
12. Gao W, Hung J (1993) Variable structure control of nonlinear systems: a new approach. IEEE Trans Ind Electron 40(1):45–55
13. Gao W, Wang Y, Homaifa A (1995) Discrete-time variable structure control systems. IEEE Trans Ind Electron 42(2):117–122
14. Golo G, Milosavljević C (2000) Robust discrete-time chattering free sliding mode control. Syst Control Lett 41(1):19–28
15. Jagannathan S, Talluri J (2002) Predictive congestion control of ATM networks: multiple sources/single buffer scenario. Automatica 38(5):815–820
16. Janardhanan S, Bandyopadhyay B (2007) Multirate output feedback based robust quasi-sliding mode control of discrete-time systems. IEEE Trans Autom Control 52(3):499–503
17. Jing Y, Yu N, Kong Z, Dimirovski G (2008) Active queue management algorithm based on fuzzy sliding model controller. In: Proceedings of the 17th IFAC world congress, pp 6148–6153
18. Kurode S, Bandyopadhyay B, Gandhi P (2011) Discrete sliding mode control for a class of underactuated systems. In: Proceedings of 37th annual conference on IEEE industrial electronics society, pp 3936–3941
19. Laberteaux KP, Rohrs CE, Antsaklis PJ (2002) A practical controller for explicit rate congestion control. IEEE Trans Autom Control 47(6):960–978
20. Lai NO, Edwards C, Spurgeon S (2007) On output tracking using dynamic output feedback discrete-time sliding-mode controllers. IEEE Trans Autom Control 52(10):1975–1981
21. Mehta A, Bandyopadhyay B (2010) The design and implementation of output feedback based frequency shaped sliding mode controller for the smart structure. In: Proceedings of IEEE international symposium on industrial electronics, pp 353–358
22. Mija S, Susy T (2010) Reaching law based sliding mode control for discrete MIMO systems. In: Proceedings of IEEE international conference on control, automation, robotics and vision, pp 1291–1296
23. Milosavljević C (1985) General conditions for the existence of a quasisliding mode on the switching hyperplane in discrete variable structure systems. Autom Remote Control 46(3):307–314
24. Quet PF, Ataslar B, Iftar A, Ozbay H, Kalyanaraman S, Kang T (2002) Rate-based flow controllers for communication networks in the presence of uncertain time-varying multiple time-delays. Automatica 38(6):917–928
25. Shtessel Y, Edwards C, Fridman L, Levant A (2014) Sliding mode control and observation. Springer, New York
26. Tanenbaum AS, Wetherall DJ (2010) Computer networks. Prentice Hall, Boston

27. Utkin V (1977) Variable structure systems with sliding modes. IEEE Trans Autom Control 22(2):212–222
28. Utkin V, Drakunov SV (1989) On discrete-time sliding mode control. In: Proceedings of IFAC conference on nonlinear control, pp 484–489
29. Utkin V, Guldner J, Shi J (1999) Sliding mode control in electromechanical systems. Taylor & Francis, London
30. Xia Y, Liu GP, Shi P, Chen J, Rees D, Liang J (2007) Sliding mode control of uncertain linear discrete time systems with input delay. IET Control Theory Appl 1(4):1169–1175
31. Yu X, Wang B, Li X (2012) Computer-controlled variable structure systems: the state of the art. IEEE Trans Ind Inf 8(2):197–205

Interval Methods for Robust Sliding Mode Control Synthesis of High-Temperature Fuel Cells with State and Input Constraints

Andreas Rauh and Luise Senkel

Abstract Fuel cell systems provide a way to produce electric energy in future decentralized power supply grids. In the case of using high-temperature fuel cells, it becomes possible to exploit not only the provided electric power but also the process heat in order to maximize the overall system efficiency. However, the efficiency maximization goes along with a high flexibility with respect to temporal variations of the electric power that is demanded by corresponding consumers. Such power variations impose restrictions on intelligent fuel cell control systems. Such control strategies do not only have to make sure that the supplied fuel gas (typically hydrogen and mixtures with methane or carbon monoxide) is stoichiometrically balanced with the demanded electric power. It is also inevitable to control the fuel cell itself from a thermodynamic point of view. This control has to make sure that sufficiently smooth temperature trajectories can be tracked during the heating phase of the system and that a priori unknown but bounded disturbances are robustly compensated at high-temperature operating points. For this purpose, interval-based sliding mode control procedures can be implemented. This contribution gives an overview of how interval methods can be combined with the fundamental sliding mode methodology in a variable-structure control synthesis. The efficiency of the presented methods is highlighted for the control of solid oxide fuel cells in various simulations.

1 Introduction

The control of nonlinear dynamic systems is an important topic for many practical applications. Especially, in cases in which dynamic system models are significantly influenced by uncertain parameters and bounded (additive) uncertainty, it is challenging to determine feedback control procedures that reliably stabilize the

A. Rauh (✉) · L. Senkel
Chair of Mechatronics, University of Rostock,
Justus-von-Liebig-Weg 6, 18059 Rostock, Germany
e-mail: andreas.rauh@uni-rostock.de

L. Senkel
e-mail: luise.senkel@uni-rostock.de

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*,
Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_3

system dynamics and—at the same time—guarantee that specific state constraints are not violated. As mentioned in the abstract of this contribution, the control of high-temperature solid oxide fuel cell systems (SOFC systems) belongs to this class of applications.

As shown in previous work, it is possible to use interval arithmetic techniques for the implementation of feedback controllers that can be applied in real time [19–27, 29]. On the one hand, these controllers have to fulfill the requirement of robustness against bounded uncertainty and disturbances. On the other hand, the asymptotic stability of the dynamics of the overall closed-loop control structure has to be shown in a guaranteed way. The basic idea of the chosen control structure is motivated by the principle of sliding mode control [31, 32, 36, 37].

In classical sliding mode techniques, so-called equivalent control strategies are determined for the exact tracking of sufficiently smooth reference trajectories. These equivalent controllers represent the control for states located exactly on the sliding surface, which serves as a specification of the desired closed-loop system dynamics. State values which are not (exactly) located on this sliding surface are forced to converge toward it using a variable-structure control approach. The amplitude of this variable-structure control component is usually selected as a constant in such a way that the influence of uncertain parameters is overcompensated. However, the choice of such constant variable-structure gains may have the drawback of unnecessarily large chattering phenomena. Such chattering should be reduced as far as possible for practical applications to avoid the associated non-advantageous actuator wear and energy consumption [19].

In contrast to classical sliding mode approaches, the fundamental idea of interval-based sliding mode control is the online adaptation of the variable-structure control component with respect to both the current uncertain system state and uncertain parameters, cf. [19, 23, 29]. This computation can be performed in real time using software libraries for basic interval functionalities. For this purpose, the calculation of the control signal is implemented in such a way that asymptotic stability of the closed-loop control system can be shown using suitable candidates for Lyapunov functions. Such candidates were so far investigated for first-order sliding mode control without and with one-sided barrier functions. These barriers serve as a guaranteed means to avoid the violation of hard upper bounds for selected state variables. For example, such upper bounds may represent the maximum admissible temperature of SOFC stacks. One of the important generalizations presented in this contribution is the extension of these techniques to an interval-based sliding mode control of second order as well as a generalization to two-sided barriers.

Section 2 gives an overview of fundamental first- and second-order sliding mode control approaches, their generalization to interval-based implementations, and illustrating simulation examples that highlight the properties and advantages of the chosen options. Thereafter, Sect. 3 describes a brief summary of the control-oriented modeling of the thermal behavior of SOFCs. The control design for these systems is described in Sect. 4 by the presented interval-based variable-structure methods. Representative simulation results are summarized in Sect. 5. Finally, conclusions and an outlook on future work are given in Sect. 6.

2 Sliding Mode Control with Guaranteed State Constraints

In this section, a summary is given for fundamental sliding mode control approaches that can be generalized in two different ways. First, they are generalized in such a way that violations of hard constraints on state variables and deviations of the states from the sliding surface are penalized in a reliable way. This penalization makes use of strict inequalities that are represented using a barrier Lyapunov function approach [9, 17, 35]. Second, techniques are introduced that allow for a verified treatment of uncertainty in the state equations by means of interval arithmetic [11, 14, 16]. These techniques aim at the online computation of control laws in such a manner that chattering due to unnecessarily large switching amplitudes is reduced as much as possible. The usage of these approaches is described for an illustrative example that is similar to the dynamics of the considered fuel cell system after a suitable coordinate transformation.

2.1 Fundamental Sliding Mode Control Laws

Both the treatment of strict inequality constraints and bounded interval uncertainty can be combined with first- and second-order sliding mode techniques. The fundamental stages are the definition of appropriate sliding surfaces and the guaranteed proof of asymptotic stability using suitable candidates for Lyapunov functions.

2.1.1 First-Order Sliding Mode Control

As an illustrative example, the nth-order linear system model

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_{n-1}(t) \\ \dot{x}_n(t) \end{bmatrix} = \begin{bmatrix} x_2(t) \\ x_3(t) \\ \vdots \\ x_n(t) \\ u(t) \end{bmatrix} \tag{1}$$

with the state vector $\mathbf{x}(t) \in \mathbb{R}^n$ and the scalar control input $u(t) \in \mathbb{R}$ is considered. The system output is represented by the first state variable according to

$$y(t) = x_1(t). \tag{2}$$

Obviously, the dynamic system (1) with the output variable (2) has the relative degree n [12, 15]. This is confirmed by the fact that the nth time derivative $x_1^{(n)}(t) = u(t)$ of the system output is the lowest-order derivative that explicitly depends on the control input $u(t)$. Therefore, the output $y(t)$ corresponds to a (trivial) flat system

output [5] with which the complete system dynamics and suitable feedforward and feedback control approaches can be parameterized for sufficiently smooth desired trajectories $x_{1,\mathrm{d}}(t)$.

Using this desired output, the corresponding tracking error ($r = 0$) and its rth time derivative are given by

$$\tilde{\xi}_1^{(r)}(t) = x_1^{(r)}(t) - x_{1,\mathrm{d}}^{(r)}(t) \tag{3}$$

with $r \in \{0, 1, \ldots, n\}$.

Using the definition (3) of the tracking error, the sliding surface

$$s := s(t) = \sum_{r=0}^{n-1} \alpha_r \tilde{\xi}_1^{(r)}(t) \tag{4}$$

with the normalized coefficient $\alpha_{n-1} = 1$ can be defined. To guarantee asymptotic stability of the system dynamics on this sliding surface, the parameters α_r have to fulfill the necessary and sufficient stability conditions for a Hurwitz polynomial [6] of linear dynamic systems of the order $n - 1$.

First-order sliding mode control approaches can be derived with the help of the quadratic radially unbounded candidate for a Lyapunov function

$$V^{\langle \mathrm{I} \rangle} = \frac{1}{2} s^2 > 0 \quad \text{for} \quad s \neq 0. \tag{5}$$

(Global) Asymptotic stability of the dynamic system corresponds to the (global) negative definiteness of the corresponding time derivative

$$\dot{V}^{\langle \mathrm{I} \rangle} = s \cdot \dot{s} = \left(\sum_{r=0}^{n-1} \alpha_r \tilde{\xi}_1^{(r)}(t) \right) \cdot \left(\sum_{r=0}^{n-1} \alpha_r \tilde{\xi}_1^{(r+1)}(t) \right) < 0 \quad \text{for} \quad s \neq 0. \tag{6}$$

During the derivation of the variable-structure sliding mode control approach [21, 27], the right-hand side of the inequality (6) is replaced by the more conservative formulation

$$\left(\sum_{r=0}^{n-1} \alpha_r \tilde{\xi}_1^{(r)}(t) \right) \cdot \left(\sum_{r=0}^{n-1} \alpha_r \tilde{\xi}_1^{(r+1)}(t) \right) < -\eta \cdot |s| = -\eta \cdot \left(\sum_{r=0}^{n-1} \alpha_r \tilde{\xi}_1^{(r)}(t) \right) \cdot \operatorname{sign}(s) \tag{7}$$

which guarantees global asymptotic stability for arbitrary parameters $\eta > 0$. Note that the actual choice of η significantly influences the dynamics and the maximum absolute values of the control signal in the so-called reaching phase in which $s \neq 0$ holds. As soon as the sliding surface $s = 0$ has been reached in a finite time, the control amplitudes depend on the actual choice of the reference trajectory $x_{1,\mathrm{d}}(t)$ and on the coefficients α_r. The latter values have the major influence on the control

amplitudes as soon as non-modeled errors and disturbances influence the system dynamics and if the error signals $\tilde{\xi}_1^{(r)}(t)$ are corrupted by non-negligible measurement noise or state reconstruction errors.

The derivation of the control law is completed by enforcing that the second factor in (7) becomes proportional to the sign of the actual value of s according to

$$\left(\sum_{r=0}^{n-1} \alpha_r \tilde{\xi}_1^{(r)}(t)\right) \cdot \underbrace{\left(\sum_{r=0}^{n-2} \alpha_r \tilde{\xi}_1^{(r+1)}(t) + u(t) - x_{1,\mathrm{d}}^{(n)}(t) + \eta \cdot \operatorname{sign}(s)\right)}_{-\beta \cdot \operatorname{sign}(s)} < 0 \tag{8}$$

with $\beta > 0$. The definition of $\tilde{\eta} := \eta + \beta > 0$ leads to the final control signal

$$u(t) = u^{\langle \mathrm{I} \rangle}(t) = x_{1,\mathrm{d}}^{(n)}(t) - \sum_{r=0}^{n-2} \alpha_r \tilde{\xi}_1^{(r+1)}(t) - \tilde{\eta} \cdot \operatorname{sign}(s)\,. \tag{9}$$

In principle, the robustness of the closed-loop control system can be improved by adding the integral of the tracking error with a suitable gain value to the definition of the sliding surface. Such additional measures are investigated in the following subsection for the derivation of second-order sliding mode controllers.

2.1.2 Second-Order Sliding Mode Control

A second-order sliding mode is defined in the sense that not only $s = s(t) = 0$ but also $\dot{s} = \dot{s}(t) = 0$ are ensured by the designed feedback controller [1, 4, 7]. This can be achieved by additionally low-pass filtering (first-order lag dynamics) the left-hand side of

$$\gamma_1 \dot{s} + \gamma_0 s = \sum_{r=0}^{n-1} \alpha_r \tilde{\xi}_1^{(r)}(t). \tag{10}$$

For the sake of asymptotic stability, the coefficients γ_0 and γ_1 need to be strictly positive, while the coefficients on the right-hand side of (10) are again chosen as parameters of a Hurwitz polynomial of the order $n - 1$. As before, this sliding surface has a PD (proportional, differentiating) characteristic.

To enhance steady-state accuracy, the sliding surface in (10) is extended by an additional time integral of the tracking error with

$$\gamma_1 \dot{s} + \gamma_0 s = \alpha_{-1} \int_0^t \tilde{\xi}_1(\tau)\mathrm{d}\tau + \sum_{r=0}^{n-1} \alpha_r \tilde{\xi}_1^{(r)}(t). \tag{11}$$

For a short-hand notation, this extension of the sliding surface (11) is abbreviated by

$$\gamma_1 \dot{s} + \gamma_0 s = \sum_{r=-1}^{n-1} \alpha_r \tilde{\xi}_1^{(r)}(t) \quad \text{with} \quad \tilde{\xi}_1^{(-1)}(t) := \int_0^t \tilde{\xi}_1(\tau) \mathrm{d}\tau. \tag{12}$$

The second time derivative of the PID-type sliding variable s, required subsequently for the control design, is given by the differentiation of (11), (12) and subsequently solving it for $\ddot{s}$ with

$$\ddot{s} = -\frac{\gamma_0}{\gamma_1}\dot{s} + \frac{1}{\gamma_1} \sum_{r=-1}^{n-1} \alpha_r \tilde{\xi}_1^{(r+1)}(t) = -\frac{\gamma_0}{\gamma_1}\dot{s} + \frac{1}{\gamma_1} \sum_{r=0}^{n} \alpha_{r-1} \tilde{\xi}_1^{(r)}(t). \tag{13}$$

Here, the special case $\alpha_{-1} \equiv 0$ corresponds to the case of a sliding surface of PD type. In analogy to the previous subsection, an appropriate Lyapunov function candidate needs to be defined to parameterize a variable-structure controller that guarantees asymptotic stability of the closed-loop system dynamics. Because $\dot{s} = 0$ has to be ensured in addition to $s = 0$, the definition

$$V^{\langle \mathrm{II} \rangle} = \frac{1}{2} \cdot \left(s^2 + \lambda \dot{s}^2 \right) \quad \text{with the scaling factor} \quad \lambda > 0 \tag{14}$$

is employed. Its time derivative results in

$$\begin{aligned} \dot{V}^{\langle \mathrm{II} \rangle} &= s \cdot \dot{s} + \lambda \cdot \dot{s} \cdot \ddot{s} \\ &= s \cdot \dot{s} + \dot{s} \cdot \left(-\frac{\lambda \gamma_0}{\gamma_1} \dot{s} + \frac{\lambda}{\gamma_1} \sum_{r=0}^{n} \alpha_{r-1} \tilde{\xi}_1^{(r)}(t) \right) < 0, \end{aligned} \tag{15}$$

where the special parameterization $\lambda = \gamma_1 > 0$ can be used without loss of generality. This is due to the fact that scaling of $\dot{V}^{\langle \mathrm{II} \rangle}$ in (15) can be performed by a suitable choice of γ_0 and α_r.

Using $\lambda = \gamma_1$ allows for simplifying the expression (15) under consideration of the control-dependent term $\tilde{\xi}^{(n)}(t) = \dot{x}_n(t) - x_{1,\mathrm{d}}^{(n)}(t) = u(t) - x_{1,\mathrm{d}}^{(n)}(t)$ according to

$$\dot{V}^{\langle \mathrm{II} \rangle} = s \cdot \dot{s} + \dot{s} \cdot \left(-\gamma_0 \dot{s} + \sum_{r=0}^{n-1} \alpha_{r-1} \tilde{\xi}_1^{(r)}(t) + \alpha_{n-1} \cdot \left(u(t) - x_{1,\mathrm{d}}^{(n)}(t) \right) \right) < 0. \tag{16}$$

As before, a conservative stabilization of the closed-loop system is desired that allows for a finite-time convergence toward $s = 0$. This can be achieved by setting

$$\dot{V}^{\langle \mathrm{II} \rangle} < -\eta_1 \cdot |\dot{s}| - \eta_2 \cdot |s| \cdot |\dot{s}| = -\dot{s} \cdot \operatorname{sign}(\dot{s}) \cdot (\eta_1 + \eta_2 \cdot |s|) \tag{17}$$

which finally leads to the nonlinear feedback controller [4, generalized form of Eqs. (22), (23)]

$$u(t) = u^{\langle \mathrm{II} \rangle}(t) = x_{1,\mathrm{d}}^{(n)}(t) + \frac{1}{\alpha_{n-1}} \cdot \left(\gamma_0 \dot{s} - s - \sum_{r=0}^{n-1} \alpha_{r-1} \tilde{\xi}_1^{(r)}(t) - \operatorname{sign}(\dot{s}) \cdot (\tilde{\eta}_1 + \tilde{\eta}_2 \cdot |s|) \right) \tag{18}$$

with $\tilde{\eta}_i \geq \eta_i > 0$ for both $i \in \{1, 2\}$.

2.2 Extension by One-Sided Barrier Lyapunov Functions

Both the control laws $u^{\langle \mathrm{I} \rangle}(t)$ and $u^{\langle \mathrm{II} \rangle}(t)$ can be extended by a one-sided barrier Lyapunov function approach in such a way that the generally time-varying strict state (respectively output) constraint

$$x_1(t) < \bar{x}_{1,\max}(t) := x_{1,\mathrm{d}}(t) + \Delta x_{1,\max}(t) \tag{19}$$

with $\Delta x_{1,\max}(t) > 0$ is guaranteed not to be violated for each point of time $t > 0$. Note that the initial conditions for the state vector $\mathbf{x}(t)$ at the point of time $t = 0$ have to be compatible with this constraint. Moreover, it is necessary that the sliding surface $s = 0$ for $x_1(t) = x_{1,\mathrm{d}}(t)$ lies within the admissible operating range that is defined by (19).

Then, the extended Lyapunov function ansatz

$$V^{\langle j,\mathrm{A} \rangle} = V^{\langle j \rangle} + V^{\langle \mathrm{A} \rangle} > 0 \quad \text{for} \quad s \neq 0 \tag{20}$$

with

$$V^{\langle \mathrm{A} \rangle} = \rho_{\mathrm{V}} \cdot \ln \left(\frac{\sigma_{\mathrm{V}} \cdot \bar{x}_{1,\max}(t)}{\bar{x}_{1,\max}(t) - x_1(t)} \right) \quad \text{and} \quad x_1(t) < \bar{x}_{1,\max}(t) \tag{21}$$

is introduced for both alternatives $j \in \{\mathrm{I}, \mathrm{II}\}$. In (21), the parameter $\rho_{\mathrm{V}} > 0$ needs to be chosen in such a way that the singularity $\bar{x}_{1,\max}(t) - x_1(t) = 0$ represents a repelling potential, where control constraints are not violated for usual operating conditions, and that the term $V^{\langle j \rangle}$ has dominating influence in the neighborhood of $s = 0$. In addition, the parameter $\sigma_{\mathrm{V}} > 0$ can be utilized to adapt the steepness of the barrier function near its singularity.

The time derivative of (20) can be computed as[1]

$$\dot{V}^{\langle j,\mathrm{A} \rangle} = \dot{V}^{\langle j \rangle} + \dot{V}^{\langle \mathrm{A} \rangle} < 0 \quad \text{with}$$

$$\dot{V}^{\langle \mathrm{A} \rangle} := \left(\frac{\partial V^{\langle \mathrm{A} \rangle}}{\partial \mathbf{x}} \right)^T \cdot \dot{\mathbf{x}}(t) = \frac{\rho_{\mathrm{V}}}{\bar{x}_{1,\max}(t)} \cdot \left(\frac{-x_1(t) \cdot \dot{\bar{x}}_{1,\max}(t) + \dot{x}_1(t) \cdot \bar{x}_{1,\max}(t)}{\bar{x}_{1,\max}(t) - x_1(t)} \right). \tag{22}$$

[1] Note that the expression $\dot{V}^{\langle \mathrm{A} \rangle}$ does not explicitly depend on the system input u in any of the applications considered in this chapter.

In analogy to the fundamental first-order sliding mode control law $u^{\langle \mathrm{I} \rangle}(t)$ derived from (8), the inequality

$$s \cdot \underbrace{\left(\sum_{r=0}^{n-2} \alpha_r \tilde{\xi}_1^{(r+1)}(t) + u(t) - x_{1,\mathrm{d}}^{(n)}(t) + \eta \cdot \operatorname{sign}(s) + \frac{1}{s} \cdot \dot{V}^{\langle \mathrm{A} \rangle} \right)}_{-\beta \cdot \operatorname{sign}(s)} < 0 \tag{23}$$

has to be fulfilled to prevent overshooting the state constraint (19) and to stabilize the error dynamics in a reliable way.

Under consideration of the term $u^{\langle \mathrm{I} \rangle}(t)$ defined in (9), the modified control law

$$u(t) = u^{\langle \mathrm{I,A} \rangle}(t) = u^{\langle \mathrm{I} \rangle}(t) - \frac{s}{s^2 + \tilde{\varepsilon}} \cdot \dot{V}^{\langle \mathrm{A} \rangle} \tag{24}$$

is obtained, in which the rational term $\frac{1}{s}$ in (23) has been approximated by the expression $\frac{s}{s^2+\tilde{\varepsilon}}$ with the small positive constant $\tilde{\varepsilon} > 0$. The approximation of this rational term ensures that the control law $u^{\langle \mathrm{I,A} \rangle}(t)$ is regular on the sliding surface $s = 0$ and that the barrier Lyapunov function becomes inactive as soon as the control goal has been reached. This is especially true in the case that interval uncertainty has a non-negligible influence on the system dynamics. This uncertainty leads to the fact that the sign of s can usually no longer be determined unambiguously in the close vicinity of $s = 0$. Hence, a good approximation of the rational term $\frac{1}{s}$ is only necessary for $|s| \gg 0$, where $\frac{1}{s} \approx \frac{s}{s^2+\tilde{\varepsilon}}$ holds.

For the special case of a time-independent state constraint $\bar{x}_{1,\max} = \text{const}$ with $\dot{\bar{x}}_{1,\max} = 0$, the control law (24) simplifies to

$$u(t) = u^{\langle \mathrm{I,A} \rangle}(t) = u^{\langle \mathrm{I} \rangle}(t) - \frac{s}{s^2 + \tilde{\varepsilon}} \cdot \rho_{\mathrm{V}} \cdot \left(\frac{x_2(t)}{\bar{x}_{1,\max} - x_1(t)} \right). \tag{25}$$

In a similar way, the second-order sliding mode control procedure can be extended by the barrier function (21). Following the same steps as in Eqs. (22)–(25) yields the control law

$$u(t) = u^{\langle \mathrm{II,A} \rangle}(t) = u^{\langle \mathrm{II} \rangle}(t) - \frac{1}{\alpha_{n-1}} \cdot \frac{\dot{s}}{\dot{s}^2 + \tilde{\varepsilon}} \cdot \dot{V}^{\langle \mathrm{A} \rangle}, \tag{26}$$

that can again be simplified as in (22) to obtain the control signal

$$u(t) = u^{\langle \mathrm{II,A} \rangle}(t) = u^{\langle \mathrm{II} \rangle}(t) - \frac{1}{\alpha_{n-1}} \cdot \frac{\dot{s}}{\dot{s}^2 + \tilde{\varepsilon}} \cdot \rho_{\mathrm{V}} \cdot \left(\frac{x_2(t)}{\bar{x}_{1,\max} - x_1(t)} \right) \tag{27}$$

for constant state constraints with $\dot{\bar{x}}_{1,\max} = 0$.

2.3 Extension by Two-Sided Barrier Lyapunov Functions

As for the case of one-sided state constraints, also two-sided barrier Lyapunov functions can be combined with the fundamental first- and second-order sliding mode controllers. For this purpose, the Lyapunov functions $V^{\langle j \rangle}$, $j \in \{\mathrm{I}, \mathrm{II}\}$, are extended by an additive term $V^{\langle \mathrm{B} \rangle}$ according to

$$V^{\langle j,\mathrm{B} \rangle} = V^{\langle j \rangle} + V^{\langle \mathrm{B} \rangle} > 0 \quad \text{for} \quad s \neq 0. \tag{28}$$

In (28), the additive term $V^{\langle \mathrm{B} \rangle}$ can either be chosen to avoid state deviations $\left| x_1(t) - x_{1,\mathrm{d}}(t) \right| \geq \bar{\chi}$ or to avoid large tracking errors with $|s(t)| \geq \bar{\chi}$. In this contribution, only the second option is considered, where $\bar{\chi}$ is assumed to be constant. Note that all corresponding equations can be generalized in a straightforward manner to the first option as well and to the case of time-dependent bounds $\bar{\chi}(t)$.

Penalizing errors with respect to the absolute value of s leads to the definition

$$V^{\langle \mathrm{B} \rangle} = \rho_{\mathrm{V}} \cdot \ln\left(\frac{\bar{\chi}^{2l}}{\bar{\chi}^{2l} - s^{2l}} \right) \quad \text{with} \quad l \in \mathbb{N} \tag{29}$$

and the even powers $2l$, enforcing symmetric bounds for the sliding variable s. Increasing values for l typically lead to the fact that resulting state trajectories come closer to the edges of the admissible operating range.

The time derivative of (29) is then given by

$$\dot{V}^{\langle \mathrm{B} \rangle} = \left(\frac{\partial V^{\langle \mathrm{B} \rangle}}{\partial \mathbf{x}} \right)^T \cdot \dot{\mathbf{x}} = \rho_{\mathrm{V}} \cdot \frac{2l \cdot s^{2l-1} \dot{s}}{\bar{\chi}^{2l} - s^{2l}}. \tag{30}$$

According to the previous subsections, the requirement $\dot{V}^{\langle j,\mathrm{B} \rangle} < 0$ for $s \neq 0$ (and $\dot{s} \neq 0$, resp.) leads to the control laws

$$u^{\langle \mathrm{I},\mathrm{B} \rangle}(t) = u^{\langle \mathrm{I} \rangle}(t) - \frac{1}{s} \cdot \dot{V}^{\langle \mathrm{B} \rangle} = u^{\langle \mathrm{I} \rangle}(t) - \rho_{\mathrm{V}} \cdot \frac{2l \cdot s^{2l-2} \dot{s}}{\bar{\chi}^{2l} - s^{2l}} \tag{31}$$

in the case of the first-order sliding mode or to

$$u^{\langle \mathrm{II},\mathrm{B} \rangle}(t) = u^{\langle \mathrm{II} \rangle}(t) - \frac{1}{\alpha_{n-1}} \cdot \frac{1}{\dot{s}} \cdot \dot{V}^{\langle \mathrm{B} \rangle} = u^{\langle \mathrm{II} \rangle}(t) - \frac{1}{\alpha_{n-1}} \cdot \rho_{\mathrm{V}} \cdot \frac{2l \cdot s^{2l-1}}{\bar{\chi}^{2l} - s^{2l}} \tag{32}$$

for the second-order sliding mode. To preserve the additive superposition of a control $u^{\langle \mathrm{I} \rangle}(t)$ with a correction term resulting from the barrier function as in (24) and (25), the term $\dot{s}$ is typically estimated by a suitable low-pass filtered differentiation or by means of an observer for the implementation of (31).

2.4 Interval Extensions of Sliding Mode Control Strategies

To guarantee asymptotic stability despite bounded uncertainty in system parameters $\mathbf{p}$, it is possible to apply interval techniques in real time for the implementation of the before-mentioned sliding mode control approaches [21]. The fundamental prerequisite for the applicability of interval techniques is that all parameters (and a priori unknown disturbances as well as measurement and state reconstruction errors) are bounded by closed interval vectors $[\mathbf{p}]$ that are defined component-wise according to $\mathbf{p} \in [\mathbf{p}] = \left[\underline{\mathbf{p}};\ \overline{\mathbf{p}}\right]$ with $\underline{p}_i \leq p_i \leq \overline{p}_i, i \in \{1, \ldots, n_\mathrm{p}\}$. Furthermore, it is assumed that the dynamic systems are given as nth-order sets of ordinary differential equations (ODEs)

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_{n-1}(t) \\ \dot{x}_n(t) \end{bmatrix} = \begin{bmatrix} x_2(t) \\ x_3(t) \\ \vdots \\ x_n(t) \\ a\left(\mathbf{x}(t), \mathbf{p}\right) + b\left(\mathbf{x}(t), \mathbf{p}\right) \cdot u(t) \end{bmatrix} \tag{33}$$

in nonlinear controller canonical form. These ODEs are a natural generalization of the pure integrator chain in Eq. (1). Note that a suitable nonlinear coordinate transformation of the fuel cell model in Sect. 4 leads exactly to this type of system structure.

For the control synthesis it is assumed as before that the system output is given by the first state variable according to

$$y(t) = x_1(t) \tag{34}$$

and that all state variables are known at each point of time in terms of guaranteed interval bounds $\mathbf{x}(t) \in [\mathbf{x}](t) = \left[\underline{\mathbf{x}}(t);\ \overline{\mathbf{x}}(t)\right]$.

For the sake of controllability (and, therefore, also for the existence of the following generalized control laws), it has to be guaranteed that

$$0 \notin b\left([\mathbf{x}](t), [\mathbf{p}]\right) := \left\{ b\left(\mathbf{x}(t), \mathbf{p}\right) | b\left(\mathbf{x}(t), \mathbf{p}\right) \text{ for all } \mathbf{x}(t) \in [\mathbf{x}](t), \mathbf{p} \in [\mathbf{p}] \right\} \tag{35}$$

holds.

These assumptions lead to the possibility to define the output tracking error and its rth derivative by the interval expression

$$\tilde{\xi}_1^{(r)}(t) \in \left[\tilde{\xi}_1^{(r)}\right](t) = \left[x_1^{(r)}\right](t) - x_{1,\mathrm{d}}^{(r)}(t) \tag{36}$$

for each $r \in \{0, 1, \ldots, n\}$. Furthermore, these tracking errors can be used to generalize the first-order sliding mode control laws according to

$$\left[u^{\langle \mathrm{I} \rangle}\right](t) = \frac{-a\left([\mathbf{x}](t), [\mathbf{p}]\right) + x_{1,\mathrm{d}}^{(n)}(t) - \sum_{r=0}^{n-2} \alpha_r \cdot \left[\tilde{\xi}_1^{(r+1)}\right](t) - \tilde{\eta} \cdot \operatorname{sign}([s])}{b\left([\mathbf{x}](t), [\mathbf{p}]\right)}, \tag{37}$$

$$\left[u^{\langle \mathrm{I,A} \rangle}\right](t) = \left[u^{\langle \mathrm{I} \rangle}\right](t) - \frac{1}{b\left([\mathbf{x}](t), [\mathbf{p}]\right)} \cdot \frac{[s]}{[s]^2 + \tilde{\varepsilon}} \cdot \left[\dot{V}^{\langle \mathrm{A} \rangle}\right](t), \tag{38}$$

and

$$\left[u^{\langle \mathrm{I,B} \rangle}\right](t) = \left[u^{\langle \mathrm{I} \rangle}\right](t) - \frac{1}{b\left([\mathbf{x}](t), [\mathbf{p}]\right)} \cdot \rho_{\mathrm{V}} \cdot \frac{2l\,[s]^{2l-2}\,[\dot{s}]}{\bar{\chi}^{2l} - [s]^{2l}}. \tag{39}$$

Similarly, the interval-based generalization for the second-order sliding mode approach is given either by

$$\left[u^{\langle \mathrm{II} \rangle}\right](t) = \frac{-a\left([\mathbf{x}](t), [\mathbf{p}]\right) + x_{1,\mathrm{d}}^{(n)}(t) + \frac{1}{\alpha_{n-1}} \cdot \left[\tilde{v}^{\langle \mathrm{II} \rangle}\right](t)}{b\left([\mathbf{x}](t), [\mathbf{p}]\right)} \tag{40}$$

with

$$\left[\tilde{v}^{\langle \mathrm{II} \rangle}\right](t) := \left(\gamma_0 \cdot [\dot{s}] - [s] - \sum_{r=0}^{n-1} \alpha_{r-1} \cdot \left[\tilde{\xi}_1^{(r)}\right](t) - \operatorname{sign}([\dot{s}]) \cdot (\tilde{\eta}_1 + \tilde{\eta}_2 \cdot |[s]|)\right), \tag{41}$$

$$\left[u^{\langle \mathrm{II,A} \rangle}\right](t) = \left[u^{\langle \mathrm{II} \rangle}\right](t) - \frac{1}{b\left([\mathbf{x}](t), [\mathbf{p}]\right)} \cdot \frac{1}{\alpha_{n-1}} \cdot \frac{[\dot{s}]}{[\dot{s}]^2 + \tilde{\varepsilon}} \cdot \left[\dot{V}^{\langle \mathrm{A} \rangle}\right](t), \tag{42}$$

or by

$$\left[u^{\langle \mathrm{II,B} \rangle}\right](t) = \left[u^{\langle \mathrm{II} \rangle}\right](t) - \frac{1}{b\left([\mathbf{x}](t), [\mathbf{p}]\right)} \cdot \frac{1}{\alpha_{n-1}} \cdot \rho_{\mathrm{V}} \cdot \frac{2l\,[s]^{2l-1}}{\bar{\chi}^{2l} - [s]^{2l}}. \tag{43}$$

The choice between these different options is made as before in dependence of the type of barrier function to be taken into account by the robust control synthesis. Furthermore, the expressions $a\left([\mathbf{x}](t), [\mathbf{p}]\right)$, $b\left([\mathbf{x}](t), [\mathbf{p}]\right)$, $\left[\dot{V}^{\langle \mathrm{A} \rangle}\right](t)$, $[s] := [s](t)$, and $[\dot{s}] := [\dot{s}](t)$ denote the interval-dependent evaluations of the corresponding entries of the state equations, the time derivatives of the barrier function, the sliding surface, and its time derivative, respectively.

For the actual control implementation in a real-time environment, the interval expressions mentioned above are evaluated by means of the C++ toolbox C-XSC [13]. To guarantee asymptotic stability for all possible operating conditions, the corresponding interval variables have to be chosen in such a way that they include the state and parameter uncertainties in a rigorous way. Assuming a quasi-continuous implementation, in which the effect of time discretization errors is negligibly small, the final control signal $u(t)$ needs to be chosen from the previous intervals in such

a way that it guarantees asymptotic stability regardless of the sign of $b(\mathbf{x}(t), \mathbf{p})$, $0 \notin b([\mathbf{x}](t), [\mathbf{p}])$.

For this reason, the set of possible control variables, guaranteeing a minimum signal amplitude, consists of the infima and the suprema of the before-mentioned interval-valued control strategies. These values are given as $\underline{u} := \inf\{[u]\}$ and $\overline{u} := \sup\{[u]\}$, respectively, where $[u]$ is either of the control laws (37)–(40), (42), or (43). To account for roundoff and representation errors, the infima and suprema are inflated by a small positive value $\varepsilon > 0$ to obtain the final set of control candidates

$$\mathcal{U} := \left\{\underline{u} - \varepsilon, \underline{u} + \varepsilon, \overline{u} - \varepsilon, \overline{u} + \varepsilon\right\}. \tag{44}$$

From this set, the control (with minimum absolute value) is chosen, which guarantees to satisfy the inequality $\dot{V} < 0$ (or its generalization for the barrier Lyapunov function approach) despite the considered interval uncertainty.

2.5 Illustrative Simulation Examples

In this section, an illustrative benchmark example is used to visualize the effectiveness of the before-mentioned fundamental sliding mode control approaches and their interval-based extensions. The considered system model with $n = 3$ is given as

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} x_2(t) \\ x_3(t) \\ p_1 x_1 + p_2 x_2 + p_3 x_3 + p_4 u(t) \end{bmatrix} \tag{45}$$

with

(a) the nominal parameters $p_1 = p_2 = p_3 = 0$ and $p_4 = 1$ as well as
(b) the uncertain parameters $p_i \in [-0.1;\ 0.1]$, $i \in \{1, 2, 3\}$, and $p_4 = 1$.

This system model corresponds to the dynamics of a point mass (position $x_1(t)$, velocity $x_2(t)$) and a normalized input force $x_3(t)$, where the underlying actuator dynamics with the control input $u(t)$ are characterized by both the nominal and uncertain parameters p_j, $j \in \{1, \ldots, 4\}$.

In all simulations, the desired reference trajectory is given by

$$x_{1,\mathrm{d}}(t) = 1 - e^{-t}\ ,\quad t \geq 0, \tag{46}$$

with the initial system states $\mathbf{x}(0) = \mathbf{0}$. Note that these initial states do not satisfy the sliding condition $s = 0$ at the initial point of time $t = 0$. Hence, even without uncertainty, there exists a transition between the reaching and sliding phase.

A summary of the selected system parameters as well as of the parameterization of the Lyapunov function candidates and the corresponding barrier functions is given for all considered simulation scenarios in Table 1. For the sake of simplicity, it is assumed

Table 1 Parameterization of sliding mode control approaches for the considered benchmark application

Scenario	Definition of Lyapunov function V	Barrier	System parameters	Measurement tolerance for x_1	Parameters of s	Variable-structure gains
Case 1	$V^{\langle \mathrm{I} \rangle}$	–	$p_1 = p_2 = p_3 = 0$, $p_4 = 1$	–	$\alpha_0 = 1, \alpha_1 = 0.9$	$\tilde{\eta} = 20$
Case 2	$V^{\langle \mathrm{I,A} \rangle}$	$\rho_\mathrm{V} = 0.5, \sigma_\mathrm{V} = 1$, $\Delta x_{1,\max} = 0.01$	$p_1 = p_2 = p_3 = 0$, $p_4 = 1$	–	$\alpha_0 = 1, \alpha_1 = 0.9$	$\tilde{\eta} = 20$
Case 3	$V^{\langle \mathrm{I,A} \rangle}$	$\rho_\mathrm{V} = 0.75, \sigma_\mathrm{V} = 1$, $\Delta x_{1,\max} = 0.01$	$p_i \in [-0.1;\ 0.1]$, $i = \{1, 2, 3\}$, $p_4 = 1$	$0.0025 \cdot [-1;\ 1]$	$\alpha_0 = 15, \alpha_1 = 0.9$	$\tilde{\eta} = 20$
Case 4	$V^{\langle \mathrm{II} \rangle}$	–	$p_1 = p_2 = p_3 = 0$, $p_4 = 1$	–	$\alpha_{-1} = 0.01, \alpha_0 = 1$, $\alpha_1 = 0.9, \alpha_2 = 1$, $\gamma_0 = 1, \gamma_1 = 10$	$\tilde{\eta}_1 = \tilde{\eta}_2 = 20$
Case 5	$V^{\langle \mathrm{II,A} \rangle}$	$\rho_\mathrm{V} = 0.02, \sigma_\mathrm{V} = 1$, $\Delta x_{1,\max} = 0.01$	$p_1 = p_2 = p_3 = 0$, $p_4 = 1$	–	$\alpha_{-1} = 0.01, \alpha_0 = 1$, $\alpha_1 = 0.9, \alpha_2 = 1$, $\gamma_0 = 1, \gamma_1 = 10$	$\tilde{\eta}_1 = \tilde{\eta}_2 = 20$
Case 6	$V^{\langle \mathrm{II,B} \rangle}$	$\rho_\mathrm{V} = 0.5$, $\bar{\chi} = 5{\cdot}10^{-6}, l = 5$	$p_1 = p_2 = p_3 = 0$, $p_4 = 1$	–	$\alpha_{-1} = 0.1, \alpha_0 = 1$, $\alpha_1 = 0.9, \alpha_2 = 1$, $\gamma_0 = 1, \gamma_1 = 10$	$\tilde{\eta}_1 = \tilde{\eta}_2 = 20$
Case 7	$V^{\langle \mathrm{II} \rangle}$	–	$p_1 = p_2 = p_3 = 0$, $p_4 = 1$	–	$\alpha_{-1} = 0.01, \alpha_0 = 10$, $\alpha_1 = 0.9, \alpha_2 = 0.1$, $\gamma_0 = 1, \gamma_1 = 10$	$\tilde{\eta}_1 = \tilde{\eta}_2 = 20$
Case 8	$V^{\langle \mathrm{II} \rangle}$	–	$p_1 = p_2 = p_3 = 0$, $p_4 = 1$	–	$\alpha_{-1} = 0.01, \alpha_0 = 1$, $\alpha_1 = 0.9, \alpha_2 = 1$, $\gamma_0 = 1, \gamma_1 = 10$	$\tilde{\eta}_1 = \tilde{\eta}_2 = 0$
Case 9	$V^{\langle \mathrm{II} \rangle}$	–	$p_1 = p_2 = p_3 = 0$, $p_4 = 1$	–	$\alpha_{-1} = 0.01, \alpha_0 = 10$, $\alpha_1 = 0.9, \alpha_2 = 0.1$, $\gamma_0 = 1, \gamma_1 = 10$	$\tilde{\eta}_1 = \tilde{\eta}_2 = 0$

in the non-interval implementations (Cases 1 and 2 for the first-order sliding mode) that all states $\mathbf{x}(t)$ can be measured accurately for the quasi-continuous sliding mode implementation. This simplifying assumption is removed in the simulation Case 3 and in the further course of this contribution for the application of sliding mode techniques to the temperature control of a high-temperature SOFC system.

For both the first- and second-order sliding mode controllers $u^{\langle \mathrm{I} \rangle}(t)$ and $u^{\langle \mathrm{II} \rangle}(t)$, the parameters of the sliding surface are chosen purposefully in such a way that the roots of the associated characteristic polynomial, corresponding to

$$\sum_{r=0}^{n-1} \alpha_r \zeta^r = \alpha_0 + \alpha_1 \zeta + \zeta^2 = 0 \tag{47}$$

in the first-order case, are conjugate complex. In such a way, the straightforward sliding mode implementation without state barriers leads to an oscillatory behavior in the reaching phase with overshooting the reference trajectory $x_{1,\mathrm{d}}(t)$, see the Cases 1 and 4 as well as Figs. 1 and 2.

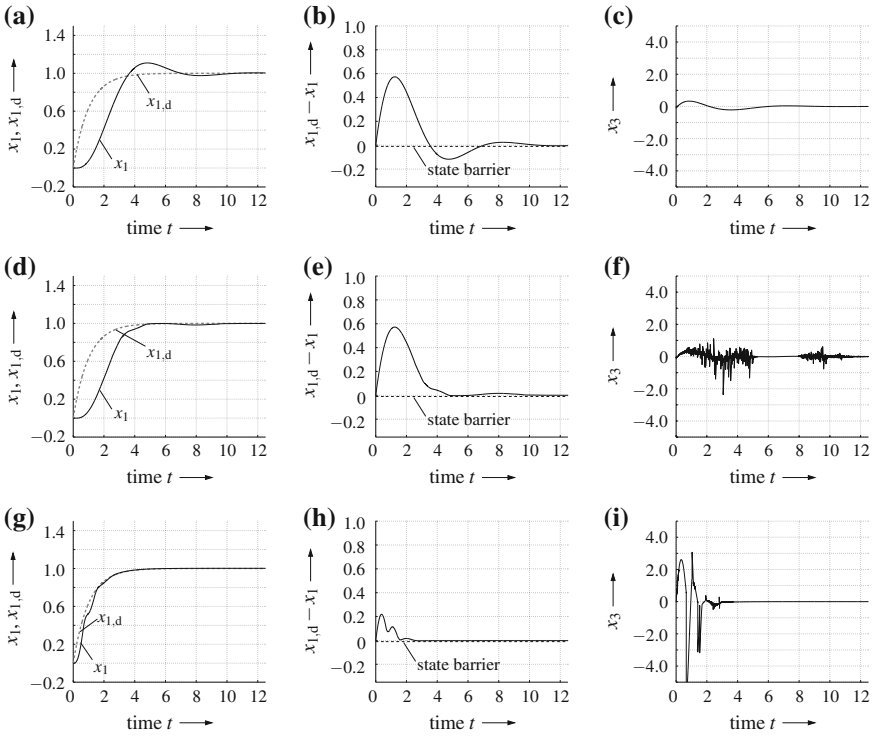


Fig. 1 Simulation results for the benchmark application with first-order sliding mode techniques. **a** System output (Case 1). **b** Tracking error (Case 1). **c** State x_3 (Case 1). **d** System output (Case 2). **e** Tracking error (Case 2). **f** State x_3 (Case 2). **g** System output (Case 3). **h** Tracking error (Case 3). **i** State x_3 (Case 3)

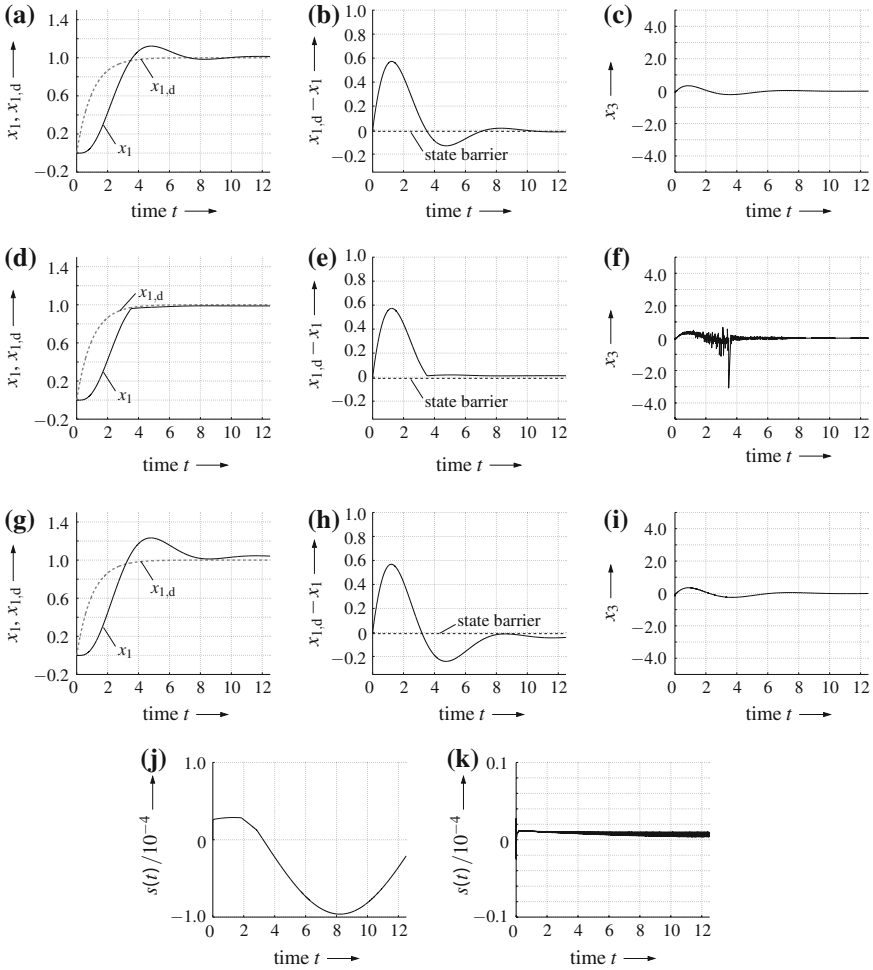


Fig. 2 Simulation results for the benchmark application with second-order sliding mode (part 1). **a** System output (Case 4). **b** Tracking error (Case 4). **c** State x_3 (Case 4). **d** System output (Case 5). **e** Tracking error (Case 5). **f** State x_3 (Case 5). **g** System output (Case 6). **h** Tracking error (Case 6). **i** State x_3 (Case 6). **j** Variation of $s(t)$ (Case 4). **k** Variation of $s(t)$ (Case 6)

Introducing a strict time-varying barrier for the state $x_1(t)$ according to (19) in the Cases 2 and 5 helps to reliably avoid the overshoot. This holds for both the first- and second-order sliding mode controllers with $u(t) = u^{\langle \mathrm{I,A} \rangle}(t)$ and $u(t) = u^{\langle \mathrm{II,A} \rangle}(t)$.

In addition to using the state barrier according to (19), the two-sided constraint (29) is employed in the Case 6 with $u(t) = u^{\langle \mathrm{II,B} \rangle}(t)$. Generally, the results for the first- and second-order cases show the same behavior. Subsequently, only the second-order result is depicted because it highlights the advantage of the extension $V^{\langle \mathrm{B} \rangle}$ more clearly for the considered application scenario: Although the two-sided barrier

extension $V^{\langle \mathrm{B} \rangle}$ according to (28) does not allow for limiting the range of state variables explicitly, it allows for limiting the deviations of states from the sliding surface s in a reliable way. This limits the variable-structure control amplitudes for the Case 6, where fixed gains $\tilde{\eta}_1$ and $\tilde{\eta}_2$ are coupled with the absolute value of s in a multiplicative way. It has to be pointed out that the maximum admissible deviation $\bar{\chi}$ as well as the necessary order l in the barrier term $V^{\langle \mathrm{B} \rangle}$ have to be chosen carefully to make sure that the admissible operating range is not violated. This may happen for specifying too tight operating ranges that inevitably lead to large control amplitudes. Such large amplitudes may, on the one hand, not be compatible with actuator constraints in a real-life application. On the other hand, they may also lead to violations of the admissible operating range if too large integration step sizes (resp. sampling times) are specified for the numerical evaluation of the system ODEs in a quasi-continuous control implementation. Especially due to the latter issue, future work will deal with the direct consideration of time discretization phenomena within the sliding mode design for continuous-time dynamic systems. Note that violations of barrier terms—due to time discretization effects in a regularized control implementation that ensures finite control values in the second-order case (the same holds for the one-sided barrier functions)—lead to nonlinear integrator wind-up phenomena in the computation of $s(t)$ and $\tilde{\xi}^{(-1)}(t)$. These wind-up effects have to be avoided by suitable parameterizations since their presence inevitably deteriorates the control quality.

The robustness of the interval-based control extensions presented in Sect. 2.4 is confirmed in the extended simulations in Case 3. It can be seen that the application of the fundamental interval-based control approach (37) guarantees asymptotic stability of the closed-loop dynamics after a careful setting of the controller's parameters. Here, uncertainties in the system parameters p_j as well as additive bounded errors in the measured state $x_1(t)$ were considered.

Note that the choice of $\Delta x_{1,\max}$ must be made in such a way that the sliding surface is reachable despite the above-mentioned measurement errors for $x_1(t)$. Due to the fact that the state barrier $\bar{x}_{1,\max}(t)$ is only considered explicitly in the Cases 2, 3, and 5, the violation of this constraint is obvious in all remaining scenarios.

Simulation results for $V^{\langle \mathrm{II,A} \rangle}$ are not presented due to an identical behavior as in Case 3. Note that the use of the second-order sliding mode is not advantageous for the case of interval uncertainty with large diameters that are directly included in the expression (11). Then, a definite statement about the signs of s and $\dot{s}$ may no longer be possible. Although the second-order sliding mode control approach leads to more smooth control signals than the first-order one for the non-interval case, this advantageous filtering property is lost as soon as ambiguities in the signs of s and $\dot{s}$ arise. Therefore and due to the fact that the uncertain parameters in the following sections have a large influence, interval-based implementations currently focus on first-order sliding mode techniques. Extensions to higher-order cases and toward an improved systematic parameterization of interval-based control approaches are subjects of ongoing work.

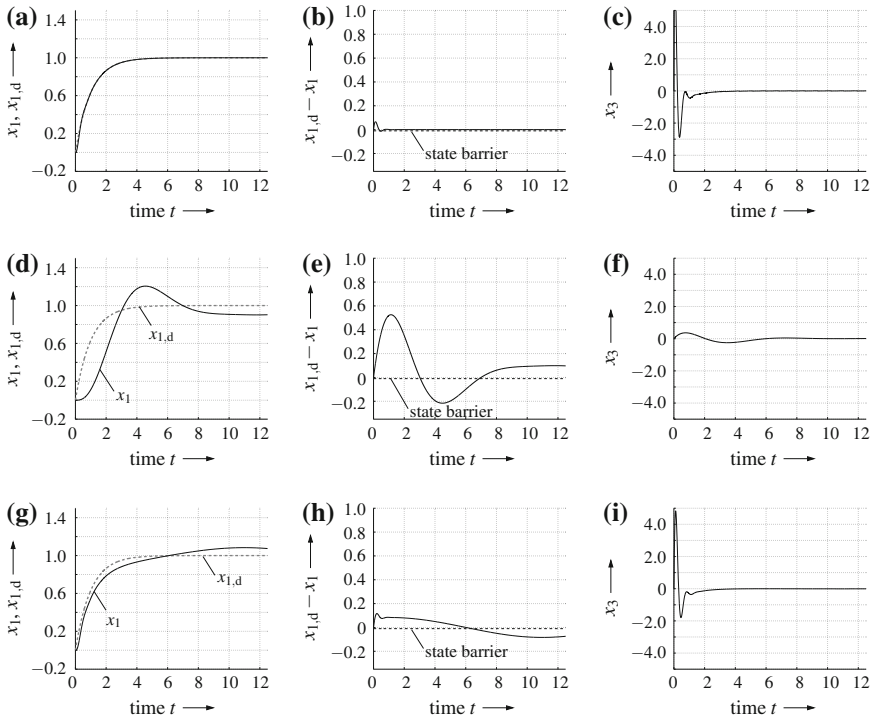


Fig. 3 Simulation results for the benchmark application with second-order sliding mode (part 2). **a** System output (Case 7). **b** Tracking error (Case 7). **c** State x_3 (Case 7). **d** System output (Case 8). **e** Tracking error (Case 8). **f** State x_3 (Case 8). **g** System output (Case 9). **h** Tracking error (Case 9). **i** State x_3 (Case 9)

A detailed investigation of the influence of the parameters of the sliding surface with and without variable-structure control parts can be found in Fig. 3 (Cases 7–9). In general, increased gain values α_0 lead to smaller tracking errors. However, accurate trajectory tracking is only possible when the variable-structure part is active. The same also holds in an analogous manner for the first-order sliding mode.

3 Control-Oriented SOFC Modeling

As described, for example, in [3, 10, 18, 33], SOFCs are characterized by a nonlinear dynamic behavior if large domains are considered for the temperature operating point as well as for the electric load of the system. Focusing on the temperature distribution in the interior of high-temperature fuel cell stack modules, the corresponding dynamics can be described by sets of partial differential equations. The goal of control design for SOFC systems is to prevent local over-temperatures in the

interior of the fuel cell stack module by a suitable actuation of the gas preheaters. This helps to reduce mechanical strain of the SOFC components due to different thermal expansion properties and thus avoids accelerated aging or in the worst case the destruction of the SOFC. The control task consists in varying the enthalpy flow provided to the fuel cell stack by changes in the temperature of the preheated gases and by variations of the gas mass flows. Typically, only the enthalpy flow of the cathode gas is used for this purpose, while the anode gas mass flow is employed to specify the electric power that can be produced by the system.

However, system models given by partial differential equations are often too complex to design controllers and state estimators that can be implemented in real time. Therefore, a control-oriented modeling procedure is used in the following to approximate the system dynamics by a finite-dimensional set of ODEs. For such systems of ODEs, the design of feedforward as well as feedback controllers can be performed by state-of-the-art approaches. Classically, this is done in the frame of SOFC systems by (gain-scheduled) PI (proportional, integral) controllers as well as by linear model-predictive control techniques [10, 34]. Since the applicability of these techniques requires that the operating temperature of the SOFC stack does not deviate too far from the point at which the nonlinear system model is linearized for design purposes, they may not be well suited if larger operating domains are considered in a flexible future power supply grid. Moreover, the use of classical linear control approaches requires an accurate knowledge of the parameters of the describing sets of ODEs. However, parameters such as heat conductivities of the fuel cell material and specific heat capacities of the fuel gases are uncertain and cannot be identified experimentally with absolute accuracy. Hence, robust control procedures have to handle such uncertainties in a reliable way. For this reason, interval-based sliding mode procedures [21, 27] are extended in this contribution to implement robust control strategies under state, input, and input rate constraints.

The prerequisite for this type of control design is the derivation of control-oriented system models. These models are derived from a spatial semi-discretization of the SOFC stack which consists of a finite number of planar fuel cells in electric series connection. The fuel cell stack is constructed in such a way that the electric current through the individual cells is orthogonal to the gas mass flow. The control-oriented model, described in detail in [27], assumes that all dynamic variables are spatially homogeneous over finitely large domains. This homogeneity assumption holds for the stack module temperatures, the electric currents as well as the corresponding internal gas mass flows. In such a way, thermodynamic quantities such as heat conductivities and specific heat capacities represent effective quantities holding in an integral balance for each of the finite volume elements.

After setting up a parameterizable set of ODEs for the thermal behavior of the SOFC stack, the parameters were identified experimentally in previous work either by local or global optimization procedures. Note that the semi-discretization procedure is based on integral heat flow and energy balances for each of the finite volume elements $\mathcal{I}$ in the stack that is depicted in Fig. 4.

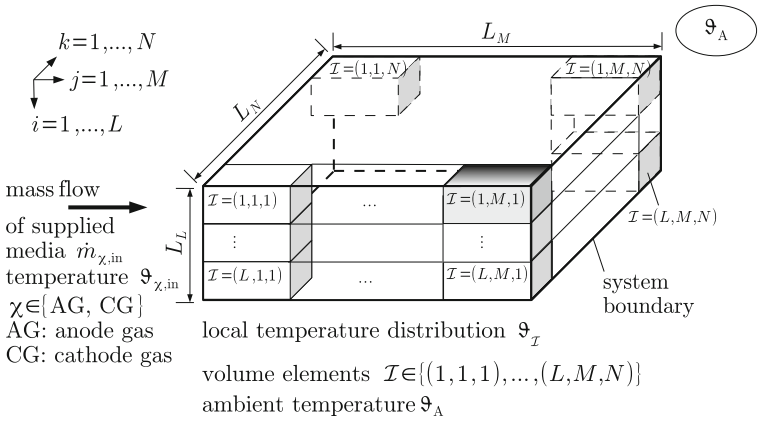


Fig. 4 Spatial semi-discretization of the SOFC stack

These integral heat flow balances lead to the ODEs

$$\dot{\vartheta}_\mathcal{I}(t) = \frac{\dot{Q}^\mathcal{I}_\mathrm{HT}(t) + \sum_\mathrm{G} \dot{Q}^\mathcal{I}_{\mathrm{G},\mathcal{I}_j^-}(t) + \dot{Q}^\mathcal{I}_\mathrm{EL}(t) + \dot{Q}^\mathcal{I}_\mathrm{R}(t)}{c_\mathcal{I} m_\mathcal{I}} \tag{48}$$

for the temperature $\vartheta_\mathcal{I}(t)$ in each volume element $\mathcal{I}$, where the index $\mathrm{G} \in \{\mathrm{AG}, \mathrm{CG}\}$ denotes the anode gas (AG) and the cathode gas (CG). For the test rig, available at the Chair of Mechatronics at the University of Rostock, the AG consists of a mixture of hydrogen (H_2), nitrogen (N_2), and water vapor (H_2O) which are jointly heated in the electric AG preheater. The CG is further given as a preheated flow of air. Additional parameters in (48) are the heat capacity $c_\mathcal{I}$ of the volume element $\mathcal{I}$ and its local mass parameter $m_\mathcal{I}$.

The heat flow term

$$\begin{aligned}\dot{Q}^\mathcal{I}_\mathrm{HT}(t) = {} & \dot{Q}^\mathcal{I}_{\mathrm{HT},\mathcal{I}_i^-}(t) + \dot{Q}^\mathcal{I}_{\mathrm{HT},\mathcal{I}_i^+}(t) + \dot{Q}^\mathcal{I}_{\mathrm{HT},\mathcal{I}_j^-}(t) \\ & + \dot{Q}^\mathcal{I}_{\mathrm{HT},\mathcal{I}_j^+}(t) + \dot{Q}^\mathcal{I}_{\mathrm{HT},\mathcal{I}_k^-}(t) + \dot{Q}^\mathcal{I}_{\mathrm{HT},\mathcal{I}_k^+}(t)\end{aligned} \tag{49}$$

in (48) consists of heat transport between directly neighboring volume elements as well as heat transfer to the ambiance. In (49), the heat flows

$$\dot{Q}^\mathcal{I}_{\mathrm{HT},\mathcal{J}}(t) = \beta^\mathcal{I}_\mathcal{J} \cdot (\vartheta_\mathcal{J}(t) - \vartheta_\mathcal{I}(t)) \tag{50}$$

are assumed to be directed from the volume elements $\mathcal{J} \in \{\mathcal{I}_i^-, \mathcal{I}_i^+, \mathcal{I}_j^-, \mathcal{I}_j^+, \mathcal{I}_k^-, \mathcal{I}_k^+\}$ into the element $\mathcal{I}$. Here, the coefficient $\beta^\mathcal{I}_\mathcal{J}$ is either the effective parameter for heat conduction in the interior of the fuel cell or the convective heat transfer coefficient (containing radiation effects in a locally linearized form) between the ambient medium and the elements at the stack boundary.

In detail, the index set $\mathcal{J}$ consists of the following entries: $\mathcal{I}_i^- := (i-1, j, k)$, $\mathcal{I}_i^+ := (i+1, j, k)$, $\mathcal{I}_j^- := (i, j-1, k)$, $\mathcal{I}_j^+ := (i, j+1, k)$, $\mathcal{I}_k^- := (i, j, k-1)$, and $\mathcal{I}_k^+ := (i, j, k+1)$. The temperatures $\vartheta_{\mathcal{I}_i^-}(t)$ and $\vartheta_{\mathcal{I}_i^+}(t)$ denote stack temperatures for $i \geq 2$ and $i \leq L-1$. The same holds for $\vartheta_{\mathcal{I}_j^-}(t)$ and $\vartheta_{\mathcal{I}_j^+}(t)$ with $j \geq 2$ and $j \leq M-1$ and for $\vartheta_{\mathcal{I}_k^-}(t)$ and $\vartheta_{\mathcal{I}_k^+}(t)$ with $k \geq 2$ and $k \leq N-1$. In all other cases, the values $\vartheta_{\mathcal{J}}(t)$ are set to the ambient temperature $\vartheta_{\mathrm{A}}(t) = \text{const}$.

In addition to these internal effects, the total enthalpy flow

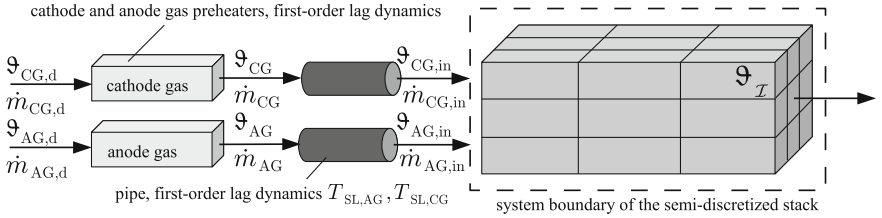
$$\sum_{\mathrm{G}} \dot{Q}^{\mathcal{I}}_{\mathrm{G},\mathcal{I}_j^-}(t)\,, \quad \mathrm{G} \in \{\mathrm{AG}, \mathrm{CG}\}, \tag{51}$$

of AG and CG is included in the ODEs, where the mass flow $\dot{m}_{\mathrm{CG}}$ and its desired temperature $\vartheta_{\mathrm{CG,d}}$ are used to design a guaranteed stabilizing control strategy. Ohmic heat production $\dot{Q}^{\mathcal{I}}_{\mathrm{EL}}(t)$ and heat flows $\dot{Q}^{\mathcal{I}}_{\mathrm{R}}(t)$ due to an exothermic reaction between AG and CG conclude the energy balance. Detailed models for the local variations of the reacting gas mass flows and their temperature-dependent parameterizations as well as explicit expressions for the reaction enthalpies are given in [27].

The finite volume model from (48)–(51) is coupled with the dynamics of the AG and CG preheaters according to Fig. 5. As shown in [27, 28], it is essential to account for the preheater dynamics during the control design for non-stationary operating conditions to avoid unnecessary chattering of the system inputs.

According to [28], each preheater is described by two sets of first-order ODEs ($\mathrm{G} \in \{\mathrm{AG}, \mathrm{CG}\}$, $\chi \in \{\mathrm{H_2}, \mathrm{N_2}, \mathrm{H_2O}, \mathrm{CG}\}$)

$$T_{\mathrm{G}} \cdot \dot{v}_{\chi}(t) + v_{\chi}(t) + \tilde{d}_{\chi}(t) = v_{\chi,\mathrm{d}}(t) = \vartheta_{\chi,\mathrm{d}}(t) \cdot \dot{m}_{\chi,\mathrm{d}}(t) \tag{52}$$



$\vartheta_{\mathrm{AG,d}}, \vartheta_{\mathrm{CG,d}}$	- desired values for temperatures of anode and cathode gas
$\vartheta_{\mathrm{AG,in}}, \vartheta_{\mathrm{CG,in}}$	- temperature of anode and cathode gas in the inlet gas manifold
$\vartheta_{\mathrm{AG}}, \vartheta_{\mathrm{CG}}$	- temperatures of anode and cathode gas at the preheater outlet
$\dot{m}_{\mathrm{AG}}, \dot{m}_{\mathrm{CG}}$	- measured mass flows of anode and cathode gas at the preheater outlet
$\dot{m}_{\mathrm{AG,d}}, \dot{m}_{\mathrm{CG,d}}$	- desired mass flows of anode and cathode gas
$T_{\mathrm{AG}}, T_{\mathrm{CG}}$	- time constants of the first-order lag dynamics of the preheaters

Fig. 5 Semi-discretization of the fuel cell stack module with gas preheaters

and

$$T_{\mathrm{SL,G}} \cdot \frac{\dot{v}_{\chi,\mathrm{in}}(t)}{L \cdot N} + \frac{v_{\chi,\mathrm{in}}(t)}{L \cdot N} = \frac{v_{\chi}(t)}{L \cdot N} = \frac{\vartheta_{\chi}(t) \cdot \dot{m}_{\chi}(t)}{L \cdot N} \tag{53}$$

with

$$v_{\chi,\mathrm{in}}(t) := \vartheta_{\chi,\mathrm{in}}(t) \cdot \dot{m}_{\chi,\mathrm{in}}(t) \quad \text{and} \quad \dot{m}_{\chi,\mathrm{d}} = \dot{m}_{\chi} = \dot{m}_{\chi,\mathrm{in}} \tag{54}$$

as well as the time constants T_{G} for the subsidiary temperature control and $T_{\mathrm{SL,G}}$ for the lag behavior due to transport phenomena in the gas supply lines (SL) between the preheaters and the SOFC stack.

In (52) and (53), the desired preheater temperatures (index d, serving as control inputs in addition to the desired mass flows $\dot{m}_{\mathrm{G,d}}$), are given by

$$\vartheta_{\chi,\mathrm{d}}(t) = \begin{cases} \vartheta_{\mathrm{AG,d}}(t) & \text{for } \chi \in \{\mathrm{H_2}, \mathrm{N_2}, \mathrm{H_2O}\} \\ \vartheta_{\mathrm{CG,d}}(t) & \text{for } \chi = \mathrm{CG}. \end{cases} \tag{55}$$

Because the AG components H_2, N_2, H_2O are mixed before entering the preheater, all components of the AG have one temperature in the interior of the preheater and one in the AG supply line. Analogously, the temperatures at the preheater outlets are denoted by $\vartheta_{\chi}(t)$, while the temperatures at the inlet gas manifold of the SOFC stack are given by $\vartheta_{\chi,\mathrm{in}}(t)$. In good accuracy, it can be assumed during modeling that the AG and CG mass flows can be changed instantaneously. This leads to the definition of *virtual* control signals $v_{\chi,\mathrm{d}}(t)$ in (52). Integrator disturbance models $\dot{\tilde{d}}_{\chi}(t) = 0$ are finally included in the description of the preheaters to account for non-modeled heat losses, thermal storage effects, and imperfect behavior of the underlying temperature control of the AG and CG preheaters.

According to [27], this control-oriented modeling approach leads to a set of input-affine ODEs with the input vector

$$\mathbf{v}_{\mathrm{d}}(t) = \left[v_{\mathrm{H_2,d}}(t) \;\; v_{\mathrm{N_2,d}}(t) \;\; v_{\mathrm{H_2O,d}}(t) \;\; v_{\mathrm{CG,d}}(t) \right]^T \tag{56}$$

of the AG and CG preheaters and the complete state vector

$$\begin{aligned} \mathbf{z}(t) = \Big[& v_{\mathrm{H_2}}(t) \;\; v_{\mathrm{H_2,in}}(t) \;\; \tilde{d}_{\mathrm{H_2}}(t) \;\; v_{\mathrm{N_2}}(t) \;\; v_{\mathrm{N_2,in}}(t) \;\; \tilde{d}_{\mathrm{N_2}}(t) \\ & v_{\mathrm{H_2O}}(t) \;\; v_{\mathrm{H_2O,in}}(t) \;\; \tilde{d}_{\mathrm{H_2O}}(t) \;\; v_{\mathrm{CG}}(t) \;\; v_{\mathrm{CG,in}}(t) \;\; \tilde{d}_{\mathrm{CG}}(t) \\ & \vartheta_{(1,1,1)}(t) \;\; \ldots \;\; \vartheta_{(L,M,N)}(t) \Big]^T \in \mathbb{R}^{\mathcal{N}}, \quad \mathcal{N} = 12 + L \cdot M \cdot N. \end{aligned} \tag{57}$$

If it is assumed that the AG properties are predefined by a subsidiary controller, the ODEs (48) which are extended by the preheater dynamics in (52)–(55) turn into

$$\begin{aligned}\dot{\mathbf{z}}(t) &= \underbrace{\boldsymbol{\phi}_1\left(\mathbf{z}(t),\mathbf{p}\right)+\boldsymbol{\Phi}_{2,\mathrm{AG}}\left(\mathbf{z}(t),\mathbf{p}\right)\cdot\begin{bmatrix}v_{\mathrm{H_2,d}}(t)\\ v_{\mathrm{N_2,d}}(t)\\ v_{\mathrm{H_2O,d}}(t)\end{bmatrix}}_{=:\,\mathbf{f}_1\left(\mathbf{z}(t),\mathbf{p},v_{\mathrm{H_2,d}}(t),v_{\mathrm{N_2,d}}(t),v_{\mathrm{H_2O,d}}(t)\right)}+\underbrace{\boldsymbol{\phi}_{2,\mathrm{CG}}\left(\mathbf{z}(t),\mathbf{p}\right)\cdot v_{\mathrm{CG,d}}(t)}_{=:\,\mathbf{f}_2\left(\mathbf{z}(t),\mathbf{p},v_{\mathrm{CG,d}}(t)\right)}\\ &= \mathbf{f}\left(\mathbf{z}(t),\mathbf{p},\mathbf{v}_{\mathrm{d}}(\mathbf{t})\right).\end{aligned} \tag{58}$$

Here, $\boldsymbol{\phi}_1\left(\mathbf{z}(t),\mathbf{p}\right)$ is input-independent, whereas the expression $\mathbf{f}_1\left(\mathbf{z}(t),\mathbf{p},\ldots\right)$ further contains information about the (desired) AG preheater actuation. The control input $v_{\mathrm{CG,d}}(t)$ is related to the CG enthalpy flow which is coupled with the system dynamics by the vector

$$\boldsymbol{\phi}_{2,\mathrm{CG}}\left(\mathbf{z}(t),\mathbf{p}\right)=\left[\mathbf{0}_{1\times 9}\ \tfrac{1}{T_{\mathrm{CG}}}\ 0\ 0\ \mathbf{0}_{1\times n_x}\right]^T,\quad n_x=L\cdot M\cdot N, \tag{59}$$

where $\mathbf{0}_{i\times j}$ is a zero matrix of dimension $i\times j$ ($n_x=L\cdot M\cdot N$: number of volume elements in the SOFC stack).

If the simplifying assumptions described in [27] are exploited, the equality

$$\frac{\partial\mathbf{f}_1\left(\mathbf{z}(t),\mathbf{p}\right)}{\partial v_{\mathrm{CG,d}}}=\mathbf{0} \tag{60}$$

holds for all operating points. Moreover, choosing $v_{\mathrm{CG,d}}(t)$ as the input justifies the use of $\frac{\mathrm{d}}{\mathrm{d}t}\left[v_{\mathrm{H_2,d}}(t)\ v_{\mathrm{N_2,d}}(t)\ v_{\mathrm{H_2O,d}}(t)\right]^T\approx\mathbf{0}$ during control synthesis. Errors that are caused by this simplification can be taken into consideration by an additive *interval-bounded* disturbance variable in the state-space transformation that is introduced in the following section. This transformation replaces the ODEs (58) by a nonlinear controller canonical form in analogy to (33). Here, the Lie derivatives that are necessary for the definition of the coordinate transformation are computed using techniques for algorithmic differentiation of a corresponding C++ source code of the state equations (58). For details concerning this state-space transformation, see [2, 8, 27].

4 Interval-Based Sliding Mode Control with State and Actuator Constraints for the Thermal Behavior of SOFCs

In this section, the interval-based sliding mode procedures derived in Sect. 2.4 are employed for the robust control of the non-stationary heating phase of the SOFC and for the compensation of disturbances at high-temperature operating points.

4.1 Interval-Based Robust Variable-Structure Control

As shown in Sect. 2.4 and [21, 27], the interval-based variable-structure control approaches make use of system models that are given in a nonlinear controller canonical form.[2] To transform the ODEs (58) into this type of system representation, it is necessary to define the output variable

$$y(t) = h(\mathbf{z}(t)) = \vartheta_{\mathcal{I}^*} \tag{61}$$

as a function of the state vector $\mathbf{z}(t) \in \mathbb{R}^{\mathcal{N}}$. Throughout the remainder of this chapter, the output is defined for each point of time as the maximum segment temperature

$$y(t) = \vartheta_{\mathcal{I}^*}(t) \quad \text{with} \quad \mathcal{I}^* = \arg\max_{\mathcal{I}} \{\vartheta_{\mathcal{I}}(t)\}. \tag{62}$$

The corresponding segment index $\mathcal{I}^*$, the temperature value $y(t)$, and a sufficient number of its time derivatives are estimated in real time by a suitable state observer [24]. The goals of the following control approaches are the accurate tracking of sufficiently smooth desired temperature trajectories during the non-stationary heating phase and the guaranteed prevention of violations of state constraints. For the latter goal it is necessary that the maximum stack temperature does not exceed a predefined upper bound in the high-temperature operating phase despite variations of the electric power demand, variations of the AG properties, and uncertainties in the temperature estimation as well as in the system parameters.

Using this output definition, the system model is transformed into the nonlinear controller canonical form. For this purpose, the Lie derivatives

$$\frac{\mathrm{d}^r y(t)}{\mathrm{d}t^r} = y^{(r)}(t) = L_{\mathbf{f}}^r h(\mathbf{z}(t)) = L_{\mathbf{f}} \left(L_{\mathbf{f}}^{r-1} h(\mathbf{z}(t)) \right), \tag{63}$$

$r = 1, \ldots, \delta$, are computed using techniques for algorithmic differentiation up to the relative degree δ. The relative degree is defined as

$$\frac{\partial L_{\mathbf{f}}^r h(\mathbf{z}(t))}{\partial v_{\mathrm{CG,d}}} \equiv 0 \ \text{ for all } \ r = 0, \ldots, \delta - 1 \quad \text{with} \quad \frac{\partial L_{\mathbf{f}}^{\delta} h(\mathbf{z}(t))}{\partial v_{\mathrm{CG,d}}} \neq 0. \tag{64}$$

Here, $y^{(\delta)}(t)$ is the smallest-order time derivative of $y(t)$ that explicitly depends on the control variable $u(t) = v_{\mathrm{CG,d}}(t)$.

Using the state vector in transformed coordinates

$$\mathbf{x}(t) = \left[h(\mathbf{z}(t)) \ L_{\mathbf{f}} h(\mathbf{z}(t)) \ \ldots \ L_{\mathbf{f}}^{\delta-1} h(\mathbf{z}(t)) \right]^T \in \mathbb{R}^{\delta} \tag{65}$$

[2] Generalized sliding mode-type control procedures, which do not necessarily rely on a transformation into nonlinear controller canonical form, are, for example, described in [29, 30].

with the transformed output variable $x_1(t) := h(\mathbf{z}(t))$, Eq. (58) is rewritten as

$$\begin{aligned}\begin{bmatrix}\dot{x}(t)\\ \dot{\tilde{\mathbf{x}}}(t)\end{bmatrix} &= \left[\begin{array}{c}\left[L_{\mathbf{f}}h(\mathbf{z}) \ \ldots \ L_{\mathbf{f}}^{\delta-1}h(\mathbf{z}) \ L_{\mathbf{f}}^{\delta}h(\mathbf{z})\right]^T\\ \hdashline \left[L_{\mathbf{f}}^{\delta+1}h(\mathbf{z}) \ \ldots \ L_{\mathbf{f}}^{\mathcal{N}}h(\mathbf{z})\right]^T\end{array}\right]\\ &= \left[\begin{array}{c}\left[x_2(t) \ \ldots \ x_\delta(t) \ \ a(\mathbf{z}(t),\mathbf{p},d(t))\right]^T\\ \hdashline \mathbf{a}^{\diamondsuit}(\mathbf{z}(t),\mathbf{p},d(t))\end{array}\right]\\ &\quad + \left[\begin{array}{c}\left[0 \ \ldots \ 0 \ \ b(\mathbf{z}(t),\mathbf{p})\cdot v_{\mathrm{CG,d}}(t)\right]^T\\ \hdashline \mathbf{b}^{\diamondsuit}(\mathbf{z}(t),\mathbf{p},d(t),v_{\mathrm{CG,d}}(t),\dot{v}_{\mathrm{CG,d}}(t),\ldots,v_{\mathrm{CG,d}}^{\mathcal{N}-\delta}(t))\end{array}\right]\end{aligned} \tag{66}$$

with the constant but uncertain parameters $\mathbf{p} \in [\mathbf{p}]$ and the additive disturbance $d(t) \in [d] = [\underline{d};\ \overline{d}\,]$. Both of these quantities are assumed to be bounded by closed intervals for the robust sliding mode design.

In (66), the term $a(\mathbf{z}(t),\mathbf{p},d(t))$ is defined by splitting up the Lie derivative $L_{\mathbf{f}}^{\delta}h(\mathbf{z}(t))$ into a purely state-dependent and an input-affine term [27] according to

$$L_{\mathbf{f}}^{\delta}h(\mathbf{z}(t)) = a(\mathbf{z}(t),\mathbf{p},d(t)) + b(\mathbf{z}(t),\mathbf{p})\cdot v_{\mathrm{CG,d}}(t). \tag{67}$$

Here, the disturbance variable $d(t)$ that is included in $a(\mathbf{z}(t),\mathbf{p},d(t))$ is observed in real time. The corresponding estimate $\hat{d}(t)$ is inflated—in analogy to the values $\left[\tilde{\xi}^{(r)}\right](t)$—to the interval $[d] := \hat{d}(t) + \Delta d\cdot[-1;\ 1]$ with $\Delta d > 0$.

Now, all robust sliding mode techniques from Sect. 2.4 can be employed for the SOFC system after setting the output $y(t)$ equal to the state variable $x_1(t) = \vartheta_{\mathcal{I}^*}(t)$. Note that variations of the location at which the maximum stack temperature is expected lead to changes in the actual system output. In such a way, it is possible that non-controllable internal dynamics with the corresponding states $\tilde{\mathbf{x}}(t)$, $\dim\{\tilde{\mathbf{x}}\} > 0$, exist. These state variables are guaranteed to be bounded due to physical conservation properties. This is described in detail in [21, 27]. Since these state variables can be estimated in real time together with the controllable states, they can be treated like time-varying disturbances or parameters during the control design.

During the application of this variable-structure control strategy, the input signal $u(t) = v_{\mathrm{CG,d}}(t)$, determined according to Sect. 2.4 with (37)–(40), (42), or (43), is decomposed into the desired preheater temperature and into the CG mass flow, respectively. Both are optimal in the following sense: Unnecessarily large temporal variations are prevented by soft penalty terms in an online-minimized cost function, while bounds on the admissible minimum and maximum absolute values are treated as hard actuator constraints. A suitable optimality criterion was introduced in [24]. If the CG mass flow is predetermined by an underlying operating strategy of the test rig, the *virtual* input $v_{\mathrm{CG,d}}(t)$ is converted directly into the desired preheater temperature. An overview of the complete control structure can be found in Fig. 6.

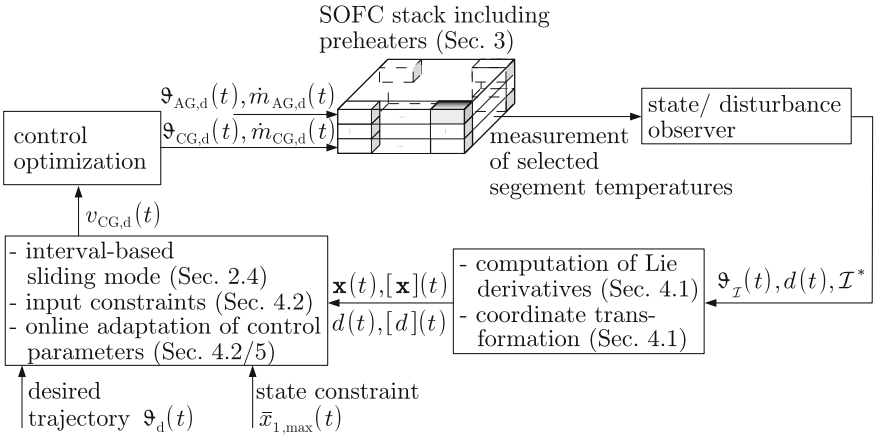


Fig. 6 Structure diagram of the complete variable-structure control law

4.2 Handling of Input and Input Rate Constraints

To make sure that the control strategies derived above can be used on a real test rig, input range as well as rate constraints have to be taken into consideration. On the one hand, this requires—according to the previous subsection—that the control variable $v_{CG,d}(t)$ is decomposed into a product of admissible gas mass flows and desired preheater temperatures. On the other hand, compatibility of the system input with actuator constraints has to be guaranteed by a suitable control parameterization.

In the offline control design of reasonable operating points, a nominal trajectory $x_{1,d}(t)$ is determined for a fixed output segment $\mathcal{I}^*$ with a predefined composition and temperature of the AG. This time-dependent trajectory is selected in such a way that control saturations are not reached for $s = 0$. In the following, this is explained in detail for the case of the first-order sliding mode.

If the nominal state trajectory is compatible with the given constraints, the interval control signal $\left[v_{CG,d}\right](t)$ is split up into a continuous and variable-structure part according to

$$\left[v_{CG,d}\right](t) = \left[v'_{CG,d}\right](t) + \tilde{\eta} \cdot \left[v''_{CG,d}\right](t) \subseteq \left[v_{CG,max}\right]. \tag{68}$$

For a suitable set of asymptotically stable eigenvalues for the dynamics on the sliding surface $s = 0$ and intervals $0 \in \left[\tilde{\xi}^{(r)}\right](t)$ for the operating range, both intervals $[v'_{CG,d}](t)$ and $\left[v_{CG,d}\right](t)$ have to be true subsets $[v'_{CG,d}](t) \subset \left[v_{CG,max}\right]$ and $\left[v_{CG,d}\right](t) \subset \left[v_{CG,max}\right]$ of the maximum possible input range $\left[v_{CG,max}\right]$. Besides an offline adaptation of the variable-structure gain $\tilde{\eta}$ (or the gains $\tilde{\eta}_1$, $\tilde{\eta}_2$ in the second-order case), also a real-time gain scheduling is possible. Both of them lead to adaptations of the parameters $\tilde{\eta}$ and α_r according to the structure diagram in Fig. 7 [22].

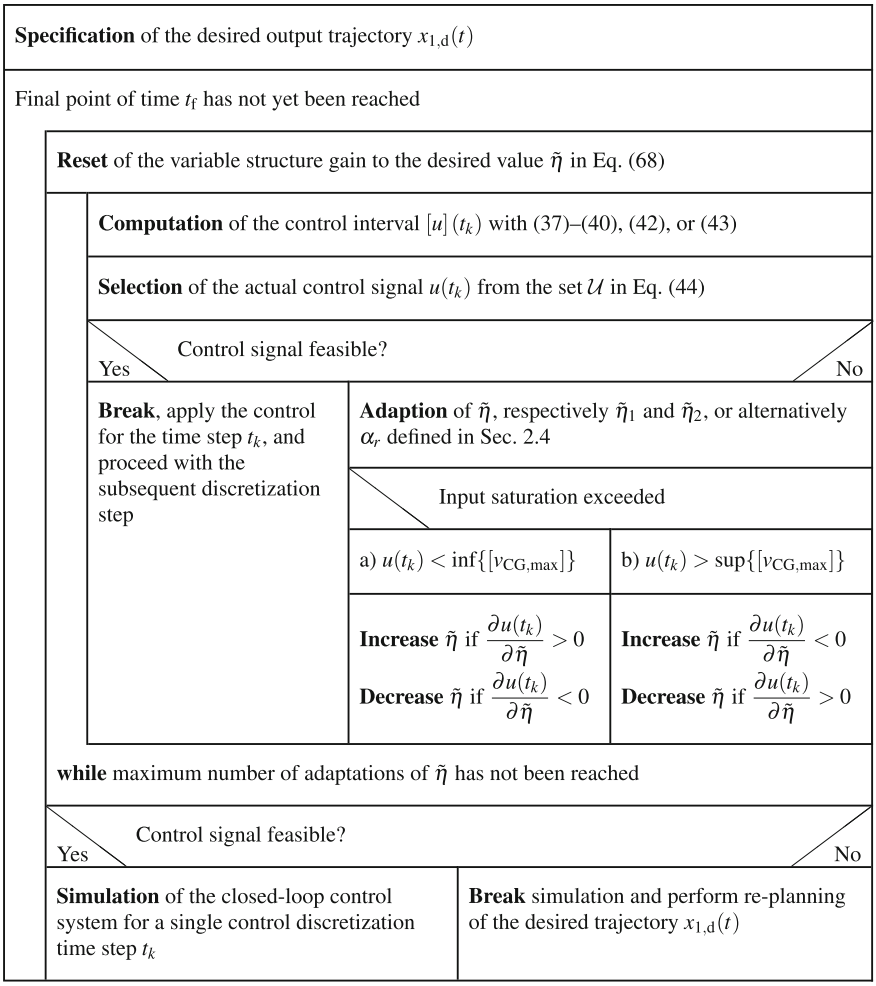


Fig. 7 Trajectory planning and gain scheduling procedure

A simultaneous treatment of hard input range and input rate constraints becomes possible if the first-order lag element

$$T_\mathrm{r} \cdot \dot{v}_{\mathrm{CG,d}}(t) + v_{\mathrm{CG,d}}(t) = \breve{v}_{\mathrm{CG,d}}(t) \quad \text{with the time constant} \quad T_\mathrm{r} > 0 \tag{69}$$

and the new system input $\breve{v}_{\mathrm{CG,d}}(t)$, is appended to the input of the ODE system (58). Equation (69) guarantees that the hard rate constraints

$$\left|\dot{v}_{\mathrm{CG,d}}(t)\right| \leq T_\mathrm{r}^{-1} \cdot \left(\sup\left\{\left[v_{\mathrm{CG,max}}\right]\right\} - \inf\left\{\left[v_{\mathrm{CG,max}}\right]\right\}\right) \tag{70}$$

are automatically satisfied for

$$\inf\left\{\left[v_{\mathrm{CG,max}}\right]\right\} \equiv \inf\left\{\left[\check{v}_{\mathrm{CG,max}}\right]\right\} \quad \text{and} \quad \sup\left\{\left[v_{\mathrm{CG,max}}\right]\right\} \equiv \sup\left\{\left[\check{v}_{\mathrm{CG,max}}\right]\right\}. \tag{71}$$

5 Simulation Results

In this section, different simulation scenarios are compared for the use of interval-based sliding mode controllers for the thermal behavior of SOFCs. In all scenarios, it is assumed that the desired temperature profile for the heating phase of the SOFC stack is given by a sufficiently smooth trajectory $\vartheta_{\mathrm{d}}(t)$ with $t \in [0;\ t^*]$. This trajectory remains constant at $\vartheta_{\mathrm{d}}(t) = \vartheta_{\mathrm{d}}(t^*)$ for $t > t^*$, $t^* = 14{,}000\,\mathrm{s}$.

Throughout the complete heating phase as well as while keeping the stationary high-temperature operating point, the AG mass flows and the AG temperature correspond to those used in [27]. Up to the point $t = t^*$, the AG consists only of preheated nitrogen, while hydrogen is included for the high-temperature phase with $\dot{m}_{\mathrm{H_2}}(t) \neq 0$ and $\frac{\mathrm{d}}{\mathrm{d}t}\dot{m}_{\mathrm{H_2}}(t) \neq 0$ for $t > t^*$. This hydrogen mass flow enables the electrochemical reaction with non-constant electric currents and leads to disturbance heat flows $\dot{Q}^{\mathcal{I}}_{\mathrm{EL}}(t)$ and $\dot{Q}^{\mathcal{I}}_{\mathrm{R}}(t)$ that need to be counteracted by the temperature control approach.

To account for the fact that the SOFC model is only an approximation of the real system dynamics, the additive interval uncertainty $[d]$ is included in the simulation according to

$$[d](t) := [-0.1;\ 0.1] \cdot \left(L^{\delta}_{\mathbf{f}} h(\mathbf{z}(t)) - b(\mathbf{z}(t), \mathbf{p}) \cdot v_{\mathrm{CG,d}}(t)\right). \tag{72}$$

Here, the term in round brackets is evaluated at each point of time $t = t_k$ with a sampling time of 0.5 s. Additionally, it is assumed for all interval-based implementations that the temperature values in the individual finite volume elements $\mathcal{I}$ are not perfectly measurable (or cannot be estimated with absolute accuracy). The corresponding errors are included in the interval $[-15;\ 15]$ K that is added to all temperature values that are involved in the computation of the Lie derivatives in (63)–(66).

Although the two-sided barrier approach included in $u^{\langle \mathrm{I,B} \rangle}(t)$ and $u^{\langle \mathrm{II,B} \rangle}(t)$ can generally be applied to the SOFC system with interval uncertainty, this approach does not allow for a guaranteed handling of hard state constraints. Therefore, only the following options for control parameterizations are considered in this section:

Case (a) $u(t) = u^{\langle \mathrm{I} \rangle}(t)$ with α_r, $\tilde{\eta} = \text{const}$ and $\vartheta_{\mathrm{d}} = 880\,\mathrm{K} = \text{const}$,
Case (b) $u(t) = u^{\langle \mathrm{I,A} \rangle}(t)$ with α_r, $\tilde{\eta} = \text{const}$ and $\vartheta_{\mathrm{d}} = 880\,\mathrm{K} = \text{const}$,
Case (c) $u(t) = u^{\langle \mathrm{I,A} \rangle}(t)$ with α_r, $\tilde{\eta} = \text{const}$ and a time-varying reference signal,
Case (d) $u(t) = u^{\langle \mathrm{I,A} \rangle}(t)$ with online adaptation of the control parameters and a time-varying reference signal.

In the Cases (a)–(c), it is assumed that all coefficients α_r and the variable-structure gain $\tilde{\eta}$ (respectively $\tilde{\eta}_1$ and $\tilde{\eta}_2$) are constant for the complete simulation horizon. The Cases (a) and (b) coincide with scenarios that were already investigated in [21]. In Case (d), an online gain scheduling approach is implemented according to the structure diagram in Fig. 7.

For the Cases (b)–(d), the maximum admissible operating temperature included in the barrier function $V^{\langle \mathrm{A} \rangle}$ is defined as the constant value $\bar{x}_{1,\max} = x_{1,\mathrm{d}}(t) + \Delta x_{1,\max}(t) \equiv 885\,\mathrm{K}$ with

$$V^{\langle \mathrm{A} \rangle} = \rho_{\mathrm{V}} \cdot \sum_{\mathcal{I}} \ln \left(\frac{\sigma_{\mathrm{V}} \cdot \bar{x}_{1,\max}}{\bar{x}_{1,\max} - \vartheta_{\mathcal{I}}(t)} \right). \tag{73}$$

If actuator constraints are violated in the Cases (a)–(c), the system input $u(t) = v_{\mathrm{CG,d}}(t)$ is set equal to the corresponding violated input constraint. Due to the fact that the control parameters α_r, $\tilde{\eta}$ (respectively $\tilde{\eta}_1$ and $\tilde{\eta}_2$) are assumed to be constant in the Cases (a)–(c), the input rate limitation introduced in Eq. (69) has not been accounted for in the computation of the corresponding control signals. However, to make sure that rate constraints are not violated at the actual plant, the control signal is filtered by (69) before applying it at the system input. Ignoring the filter time constant T_{r} during the design stage leads to some amount of chattering due to a model inaccuracy. This inaccuracy is reduced in the scenario (d) by including the filter (69) as an input rate limiter into the controller design. Moreover, the controller parameters are adapted online in the Case (d). Together with the direct inclusion of the input rate limiter (69), this leads to significantly more smooth control inputs and less chattering in the control errors. Furthermore, it yields less conservative choices of the system inputs, which becomes visible by reduced steady-state errors at high-temperature operating points.

Note that all cases in which the barrier functions were active in Figs. 8 and 9 lead to system outputs in which the maximum stack temperatures are compatible with the given state constraint.

In summary, the control implementation that was used in Case (d) of Fig. 9 is advantageous due to the following reasons from a practical point of view: First, the offline design of the desired reference trajectory helps to avoid input saturations for a nominal non-disturbed plant model. Second, the smooth desired reference trajectory for the temperature profile is the major reason why the control in Fig. 9 is generally less aggressive than for a step-like change of the reference signal that was used in Fig. 8. Finally, the online gain adaptation procedure described below for the variable-structure controller reduces chattering to a reasonable level and hence avoids unnecessary actuator wear.

The online parameter adaptation summarized in Fig. 7—which has been applied in Fig. 9—makes use of the following gain scheduling heuristics.

For the first-order sliding mode in the Cases (d), the gain adaptation is performed according to the following:

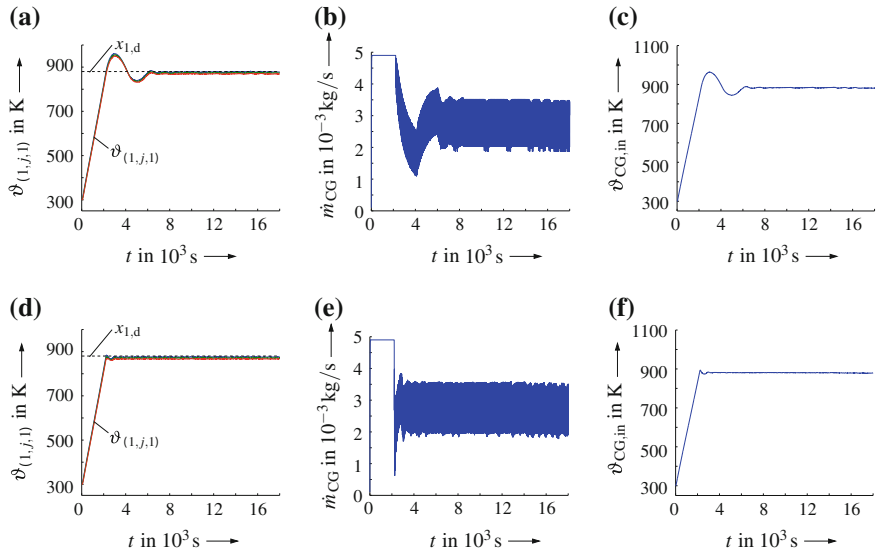


Fig. 8 Comparison of interval-based sliding mode controllers for $L = N = 1$, $M = 3$ without barrier Lyapunov function approach (Case a) and with barrier Lyapunov function (Case b) for a constant desired output $x_{1,d} = 880$ K. **a** Stack temperature (Case a). **b** Mass flow $\dot{m}_{CG}$ (Case a). **c** Temperature $\vartheta_{CG,in}$ (Case a). **d** Stack temperature (Case b). **e** Mass flow $\dot{m}_{CG}$ (Case b). **f** Temperature $\vartheta_{CG,in}$ (Case b)

Step 1. Define a desired eigenvalue λ_r of multiplicity $\delta - 1$ on the sliding surface with corresponding parameters α_r
Step 2. Initialize $\tilde{\eta}$ with the desired value
Step 3. Adapt $\tilde{\eta}$ in a line-search approach (fixed number of $N_\eta = 5$ steps) to ensure compatibility of $u(t) = v_{CG,d}(t)$ with the control constraints

- Stop, if admissible control is found;
- If no admissible control is found within N_η steps, adapt the eigenvalue λ_r and restart with Step 2; Break after at most $N_\lambda = 5$ repetitions.[3]

For the second-order sliding mode, a straightforward extension of the online gain scheduling is given basically by an extension of the previous Step 2.

Step 1. Define a desired eigenvalue λ_r of multiplicity $\delta - 1$ on the sliding surface with corresponding parameters α_r
Step 2a. Initialize both parameters $\tilde{\eta}_1$ and $\tilde{\eta}_2$ with the desired values
Step 2b. Perform one adaptation step of $\tilde{\eta}_2$ according to the sensitivities in Fig. 7
Step 3. Adapt $\tilde{\eta}_1$ in a line-search approach (fixed number of $N_\eta = 5$ steps) to ensure compatibility of $u(t) = v_{CG,d}(t)$ with the control constraints

[3]This limitation is necessary to guarantee real-time applicability of the adaptation procedure.

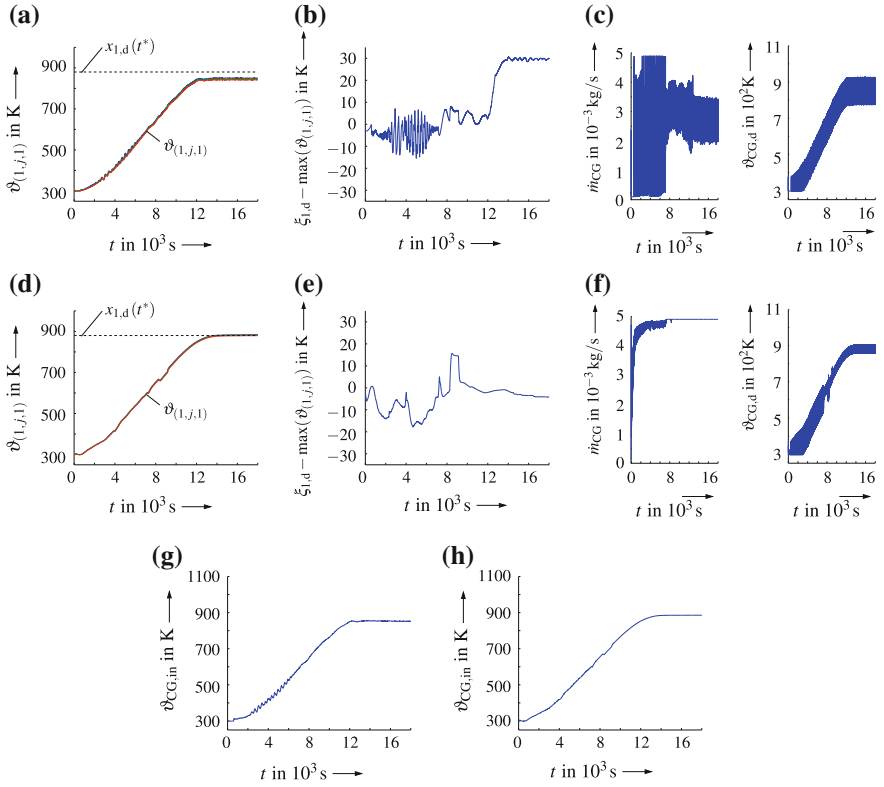


Fig. 9 Comparison of interval-based sliding mode controllers for $L = N = 1$, $M = 3$ with barrier Lyapunov function approach for constant (Case c) and variable controller gains (Case d) for a time-varying desired output $x_{1,\mathrm{d}}(t) \neq \text{const}$. **a** Stack temperature (Case c). **b** Tracking error (Case c). **c** CG control inputs (Case c). **d** Stack temperature (Case d). **e** Tracking error (Case d). **f** CG control inputs (Case d). **g** CG stack inlet temperature (Case c). **h** CG stack inlet temperature (Case d)

- Stop, if admissible control is found;
- Repeat the Steps 2b and 3 for a maximum of $M_\eta = 5$ times, where the restart of Step 3 is performed with the originally desired parameter $\tilde{\eta}_1$;
- If no admissible control is found within $M_\eta \cdot N_\eta$ steps, adapt the eigenvalue λ_r and restart with Step 2a; Break after at most $N_\lambda = 5$ repetitions.

Note that both adaptation procedures ensure stability for $\Re\{\lambda_r\} < 0$ and $\tilde{\eta} > 0$ as well as $\tilde{\eta}_1 > 0$ and $\tilde{\eta}_2 > 0$. However, simulation case studies have shown that the control performance may become worse if sign (s) changes its value during the parameter adaptation. This can be prevented by adding an additional term $-\nu(t)$ with

$$\nu(t) \begin{cases} < 0 & \text{for} \quad \inf\{-[\vartheta_{\mathcal{I}^*}](t) + x_{1,\mathrm{d}}\}(t) < 0 \\ > 0 & \text{for} \quad \sup\{-[\vartheta_{\mathcal{I}^*}](t) + x_{1,\mathrm{d}}\}(t) > 0 \\ = 0 & \text{else} \end{cases} \tag{74}$$

onto the numerators of (37) and (40), respectively. Note that the intervals on the right-hand side of (74) reflect the measurement or state reconstruction tolerances of the stack module temperatures. In cases in which the adaptation of the parameters according to Fig. 7 leads to a change of sign (s), the absolute value of $\nu(t)$ needs to be chosen in the following manner: If the Steps 1–3 yield no admissible solution with the same sign of s as the initialization, the term $-\nu(t)$ is selected in such a way that the control is set exactly to the respective bound of the input $u(t) = v_{\mathrm{CG,d}}(t)$ that was originally violated. According to the simulations in Fig. 9, this measure preserves the desired stability properties and still prevents overshooting the maximum admissible stack temperature, however, without a formal proof.

6 Conclusions and Outlook

In this chapter, various generalizations of sliding mode controllers were presented toward interval-based implementations. These implementations focus on a guaranteed stabilization of sets of ODEs describing the dynamics of closed-loop control systems with uncertain parameters and uncertain state variables. Both of these uncertain quantities are assumed to be represented by closed intervals. Despite the aim of a guaranteed stabilization of the control error dynamics in such uncertain settings, further generalizations were discussed for a reliable treatment of state and actuator constraints. Especially, the overshoot prevention of given upper state boundaries and the treatment of input range and input rate constraints were visualized for the control of the thermal behavior of a high-temperature fuel cell system.

It has been shown that the presented approaches lead to a guaranteed compatibility of the closed-loop dynamics with the before-mentioned constraints as well as with robustness and stability requirements. Moreover, an online parameter adaptation approach was validated in simulations which allows for a reduction of chattering if constant control parameters may cause the violation of input range constraints. Classically, such violations are purely avoided by corresponding saturation elements.

Future work will deal with further generalizations of the controller, e.g., with parameterizations of the system input by time-dependent polynomials. In such a way, it is desired to further smoothen the inputs when treating the polynomial coefficients as virtual inputs that are alternatively computed by sliding mode or predictive control techniques. Moreover, generalizations of the presented quasi-continuous variable-structure control approaches toward a discrete-time implementation may be investigated in the future.

References

1. Bartolini G, Pisano A, Punta E, Usai E (2003) A survey of applications of second-order sliding mode control to mechanical systems. Int J Control 76(9–10):875–892
2. Bendsten C, Stauning O (2007) FADBAD++, Version 2.1, http://www.fadbad.com
3. Bove R, Ubertini S (eds) (2008) Modeling solid oxide fuel cells. Springer, Berlin
4. Eker İ (2010) Second-order sliding mode control with experimental application. ISA Trans 49(3):394–405
5. Fliess M, Lévine J, Martin P, Rouchon P (1995) Flatness and defect of nonlinear systems: introductory theory and examples. Int J Control 61:1327–1361
6. Fortmann TE, Hitz KL (1977) An introduction to linear control systems. Marcel Dekker Inc, New York
7. Fridman L, Levant A (2002) Higher order sliding modes. In: Barbot JP, Perruguetti W (eds) Sliding mode control in engineering. Marcel Dekker, New York, pp 53–101
8. Griewank A, Walther A (2008) Evaluating derivatives: principles and techniques of algorithmic differentiation. SIAM, Philadelphia
9. Han SI, Cheong JY, Lee JM (2013) Barrier Lyapunov function-based sliding mode control for guaranteed tracking performance of robot manipulator. Mathematical problems in engineering. Hindawi Publishing Corporation, Article ID 978241, 9 p. doi:10.1155/2013/978241
10. Huang B, Qi Y, Murshed AKMM (2013) Dynamic modeling and predictive control in solid oxide fuel cells: first principle and data-based approaches. Wiley, Chichester
11. Jaulin L, Kieffer M, Didrit O, Walter É (2001) Applied interval analysis. Springer, London
12. Khalil HK (2002) Nonlinear systems, 3rd edn. Prentice-Hall, Upper Saddle River
13. Krämer W (2012) XSC languages (C-XSC, PASCAL-XSC) — scientific computing with validation, arithmetic requirements, hardware solution and language support. www.math.uni-wuppertal.de/~xsc/, C-XSC 2.5.3
14. Krasnochtanova I, Rauh A, Kletting M, Aschemann H, Hofer EP, Schoop KM (2010) Interval methods as a simulation tool for the dynamics of biological wastewater treatment processes with parameter uncertainties. Appl Math Model 34(3):744–762
15. Marquez H (2003) Nonlinear control systems. Wiley, New Jersey
16. Moore R (1966) Interval arithmetic. Prentice-Hall, Englewood Cliffs
17. Niu B, Zhao J (2013) Barrier Lyapunov functions for the output tracking control of constrained nonlinear switched systems. Syst Control Lett 62(10):963–971
18. Pukrushpan J, Stefanopoulou A, Peng H (2005) Control of fuel cell power systems: principles, modeling, analysis and feedback design, 2nd edn. Springer, Berlin
19. Rauh A, Aschemann H (2012) Interval-based sliding mode control and state estimation for uncertain systems. In: Proceedings of IEEE international conference on methods and models in automation and robotics MMAR 2012. Miedzyzdroje, Poland
20. Rauh A, Senkel L, Aschemann H (2014) Variable structure approaches for temperature control of solid oxide fuel cell stacks. In: Proceedings of 2nd international conference on vulnerability and risk analysis and management ICVRAM 2014. Liverpool, UK
21. Rauh A, Senkel L, Aschemann H (2015) Interval-based sliding mode control design for solid oxide fuel cells with state and actuator constraints. IEEE Trans Ind Electron 62(8):5208–5217
22. Rauh A, Senkel L, Aschemann H (2015) Reliable sliding mode approaches for the temperature control of solid oxide fuel cells with input and input rate constraints. In: Proceedings of 1st IFAC conference on modelling, identification and control of nonlinear systems, MICNON 2015. St. Petersburg, Russia
23. Rauh A, Senkel L, Auer E, Aschemann H (2014) Interval methods for the implementation of real-time capable robust controllers for solid oxide fuel cell systems. Math Comput Sci 8(3–4):525–542
24. Rauh A, Senkel L, Dötschel T, Auer E, Aschemann H (2014) Numerical verification and experimental validation of sliding mode control design for uncertain thermal SOFC models. Reliab Comput 19(4):330–350

25. Rauh A, Senkel L, Kersten J, Aschemann H (2013) Verified stability analysis for interval-based sliding mode and predictive control procedures with applications to high-temperature fuel cell systems. In: Proceedings of 9th IFAC symposium on nonlinear control systems. Toulouse, France
26. Rauh A, Senkel L, Kersten J, Aschemann H (2014) Interval methods for sensitivity-based model-predictive control of solid oxide fuel cell systems. Reliab Comput 19(4):361–384
27. Rauh A, Senkel L, Kersten J, Aschemann H (2014) Reliable control of high-temperature fuel cell systems using interval-based sliding mode techniques. IMA J Math Control Inf (available online). doi:10.1093/imamci/dnu051
28. Senkel L, Rauh A, Aschemann H (2013) Experimental validation of a sensitivity-based observer for solid oxide fuel cell systems. In: Proceedings of IEEE international conference on methods and models in automation and robotics MMAR 2013. Miedzyzdroje, Poland
29. Senkel L, Rauh A, Aschemann H (2014) Sliding mode techniques for robust trajectory tracking as well as state and parameter estimation. Math Comput Sci 8(3–4):543–561
30. Senkel L, Rauh A, Aschemann H (2016) Experimental and numerical validation of a reliable sliding mode control strategy considering uncertainty with interval arithmetic. In: Rauh A, Senkel L (eds) Variable-structure approaches for analysis, simulation, robust control and estimation of uncertain dynamic systems, mathematical engineering. Springer, Berlin, pp 87–122. doi:10.1007/978-3-319-31539_4
31. Shtessel Y, Edwards C, Fridman L, Levant A (2014) Sliding mode control and observation. Springer, New York
32. Slotine JJE, Li W (1991) Applied nonlinear control. Prentice Hall, Eaglewood Cliffs
33. Stiller C (2006) Design, operation and control modelling of SOFC/GT hybrid systems. Ph.D. thesis, University of Trondheim
34. Stiller C, Thorud B, Bolland O, Kandepu R, Imsland L (2006) Control strategy for a solid oxide fuel cell and gas turbine hybrid system. J Power Sources 158:303–315
35. Tee KP, Ge SS, Tay EH (2009) Barrier Lyapunov functions for the control of output-constrained nonlinear systems. Automatica 45(4):918–927
36. Utkin V (1992) Sliding modes in control and optimization. Springer, Berlin
37. Utkin V (1993) Sliding mode control design principles and applications to electric drives. IEEE Trans Ind Electron 40(1):23–36

Experimental and Numerical Validation of a Reliable Sliding Mode Control Strategy Considering Uncertainty with Interval Arithmetic

Luise Senkel, Andreas Rauh and Harald Aschemann

Abstract Real applications are often affected by uncertainty caused by, for example, unknown parameters, sensor inaccuracies, and noise processes. These effects influence control procedures in a significant way and have to be taken into consideration in simulations and experiments to ensure stability of a real system. Often, the dynamics of a considered system can be described by nonlinear equations. To control such systems, sliding mode techniques are advantageous in compensating not explicitly modeled disturbances that influence a system. In this contribution, common sliding mode controllers are extended and combined with interval arithmetic to enhance their performance. This can be achieved by an adaptive calculation of the state-dependent gain stabilizing the variable-structure part of the system—the so-called switching amplitude. Therefore, an exact knowledge of the system parameters is not necessary because their true values are assumed to be located in specified range bounds. Moreover, stochastic uncertainty is taken into consideration representing process and measurement noise that affect practically every real system.

1 Introduction

For the applicability of control theory, suitable mathematical models are necessary that describe a real system with good accuracy. Nevertheless, it is impossible to include all dynamic effects and disturbances influencing the real application into a model that behaves exactly as the real system. Thus, it is important to make a trade-off between model accuracy and computation time to overcome lack of knowledge as well as to decrease the implementation effort. Nevertheless, a suitable control strategy

L. Senkel (✉) · A. Rauh · H. Aschemann
Chair of Mechatronics, University of Rostock,
Justus-von-Liebig-Weg 6, 18059 Rostock, Germany
e-mail: luise.senkel@uni-rostock.de

A. Rauh
e-mail: andreas.rauh@uni-rostock.de

H. Aschemann
e-mail: harald.aschemann@uni-rostock.de

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*,
Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_4

ensuring the system's stability in a robust way is necessary to overcome non-modeled effects. This can be realized using sliding mode techniques for control—which are the focus of this contribution—and also for estimation tasks. To employ sliding mode techniques, the mathematical model of the considered system is split up into a linear part and a variable-structure part. The latter one compensates nonlinear dynamics as well as disturbances and non-modeled influences by a suitable calculation of a variable-structure gain, the so-called switching amplitude. This can be done using Lyapunov methods to guarantee that the system converges to a defined sliding surface on which stability of the closed-loop system can be ensured. Therefore, the switching amplitude forces the system to overcome the distance to the desired trajectory (called reaching phase). Afterward, the system will not diverge anymore and always stays in the surrounding area of the sliding surface (called sliding phase). In the following, deviations from this sliding surface are defined as the tracking error which should be as small as possible to provide an efficient control.

To increase robustness and to overcome the problem that system parameters are unknown or vary within some tolerances, a sliding mode control approach is developed including an online calculation of the switching amplitude. Therefore, uncertain parameters specified in terms of range tolerances are included as interval variables. Fundamental implementation steps are demonstrated for two control tasks: The first one is a drive-train test rig available at the University of Rostock for which the approach is realized in simulation and experiment. The second one is a model for the thermal behavior of a solid oxide fuel cell (SOFC) stack in combination with two gas preheaters with the focus on the heating phase.

The outline of this chapter is as follows: Sects. 2 and 3 summarize the basics of interval arithmetics and stability methods provided by Lyapunov approaches. Afterward, control theory for linear systems is discussed in Sect. 4 followed by two sliding mode control approaches in Sect. 5: a common way and an extended one. For the two mentioned application scenarios, the developed novel sliding mode controller taking into account intervals and stochastic processes is applied, which is the topic of Sects. 6 and 7. A summary and an outlook on further work conclude this chapter.

2 Interval Arithmetics and Stochastic Uncertainty

The characterization of bounded and stochastic uncertainty plays a major role in control tasks aiming at describing real systems by mathematical models in order to implement efficient and robust control strategies.

Bounded uncertainty affects real applications in different ways. Inaccuracy due to manufacturing tolerances is one of them. Moreover, system parameters influence the dynamics of a system and can often not be identified exactly. For this parameter identification, measurements are necessary which are affected by limited accuracy of, for example, sensors. This causes deviations between measurements and the true system states of the considered mathematical model. Such errors can be taken into consideration by specifying measurement error intervals. In addition, parameters can

be described by intervals in the sense of predefined tolerance bounds that enclose the range of a parameter around its nominal value [22].

In general, an interval is defined as

$$[a] = \left[\underline{a}\ ;\ \overline{a}\right] \tag{1}$$

with a lower bound (infimum) $\underline{a} = \inf([a])$ and an upper bound (supremum) $\overline{a} = \sup([a])$. An interval vector of length l is defined component-wise as

$$[\mathbf{a}] = \begin{bmatrix} \left[\underline{a}_1\ ;\ \overline{a}_1\right] \\ \left[\underline{a}_2\ ;\ \overline{a}_2\right] \\ \vdots \\ \left[\underline{a}_l\ ;\ \overline{a}_l\right] \end{bmatrix}. \tag{2}$$

In this contribution, intervals quantify worst-case influences of states (e.g., control errors) and parameters. Therefore, the standard arithmetic operations addition, subtraction, and multiplication can be applied to intervals as for calculations with point values [8, 9, 14]. Divisions, in which the denominator is an interval, are only allowed, if the point value zero is not included in this interval. Due to this problem, a case-sensitive evaluation of the switching amplitude is necessary as it is described in Sect. 5.3.

Unfortunately, calculating with intervals can lead to overestimation caused by two inconvenient effects: the dependency problem and the so-called wrapping effect. The dependency problem is unavoidable, if one interval occurs several times in a mathematical equation and cannot be factorized in a suitable way. An example is the difference of a given interval $[a] = [1\ ;\ 2]$ with itself according to

$$[f]([a]) = [a] - [a] = [-1\ ;\ 1] \neq [0\ ;\ 0]\,, \tag{3}$$

where the exact solution should be a point value according to $[0\ ;\ 0] = 0$. In fact, this value is only a part of the larger enclosure $[-1\ ;\ 1]$ resulting from a naive use of interval arithmetics. Because of the multiple occurrence of parameters in a mathematical model of a real system, the dependency problem may lead to large inclusions that cannot be used for further calculations or may provide too conservative solutions. The second problem, the wrapping effect, occurs because in standard interval arithmetics an interval is assumed to be an axis parallel box. In fact, such boxes may include regions that result in too conservative enclosures of non-axis-aligned solution spaces that may lead to unphysical computation results. Due to these problems, verified solution methods have to be found to reduce overestimation. The problem of overestimation in this contribution is overcome by a problem-dependent implementation of the calculation of the variable-structure gain in the control law in Sect. 5. Because this computation is repeated in each time step with the current state information, the pessimism introduced by the dependency problem can be limited.

Another class of uncertainty comprises stochastic or unknown effects on the system dynamics. On the one hand, process noise can be understood as transformation errors between two subsequent time steps. On the other hand, measurement noise may affect measurements of a system. Such stochastic uncertainty can be taken into consideration by defining standard deviations for Brownian motions[1] that can be understood as probability distributions of both mentioned random noise effects.

3 Lyapunov Methods for Stability Analysis

Of course, real applications need to be controlled such that the system does not become unstable. In fact, this means that the system states do not increase unlimited over time. The special case of asymptotic stability is characterized by the fact that the system dynamics converge over time to the system's equilibrium. To ensure stability of a system for control purposes, Lyapunov methods can be taken into consideration which are the basis for the sliding mode control described in the next section.

Consider the ordinary differential equation of a linear time-invariant system defined by a system matrix $\mathbf{A}$ and a state vector $\mathbf{x}(t)$ according to (see [7])

$$\dot{\mathbf{x}}(t) = \mathbf{A} \cdot \mathbf{x}(t) \tag{4}$$

and a positive definite quadratic Lyapunov function candidate

$$V = \frac{1}{2}\mathbf{x}^T(t) \cdot \mathbf{P} \cdot \mathbf{x}(t) > 0 \tag{5}$$

with a positive definite symmetric matrix $\mathbf{P} = \mathbf{P}^T \succ 0$. Substituting (4) into the corresponding time derivative leads to

$$\begin{aligned} \dot{V} &= \frac{1}{2}\left(\dot{\mathbf{x}}^T(t) \cdot \mathbf{P} \cdot \mathbf{x}(t) + \mathbf{x}^T(t) \cdot \mathbf{P} \cdot \dot{\mathbf{x}}(t)\right) < 0 \\ &= \frac{1}{2}\,\mathbf{x}^T(t) \cdot \left(\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A}\right) \cdot \mathbf{x}(t) < 0 \\ &\overset{!}{=} -\frac{1}{2}\,\mathbf{x}^T(t) \cdot \mathbf{Q} \cdot \mathbf{x}(t) < 0. \end{aligned} \tag{6}$$

In order to guarantee asymptotic stability in the near surrounding area of the equilibrium $\mathbf{x} = \mathbf{0}$, the term in brackets needs to be negative definite to guarantee converging behavior of the states. This results in the Lyapunov equation

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q} \tag{7}$$

[1] Normally distributed probabilities of stochastic effects are assumed as long as no further (more detailed) information is available.

with a positive definite matrix $\mathbf{Q} = \mathbf{Q}^T \succ 0$. Equation (7) has to be valid for all states $\mathbf{x}(t)$ as long as the matrices $\mathbf{A}$, $\mathbf{P}$, $\mathbf{Q}$ are state- and time-independent. The matrix $\mathbf{P}$ can be calculated if $\mathbf{Q}$ is defined (e.g., as identity matrix). If a positive definite and symmetric matrix $\mathbf{P}$ exists, the system is proven to be asymptotically stable. Such stability statements can also be made for linearized systems in order to analyze and guarantee local asymptotic stability of the corresponding nonlinear system [13]. Guaranteed stabilizing control design by the definiteness requirement in Eq. (6) is in the focus of Sect. 5 for uncertain nonlinear dynamic systems.

4 Control Theory for Linear Systems

The simplest methods to enforce trajectory tracking and steady-state accuracy for linear systems are feedback controllers. Therefore, assume a linear system that can be represented in state-space form (here, for a single-input single-output (SISO) system) without external disturbances according to

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}) = \mathbf{A}(\mathbf{p}) \cdot \mathbf{x}(t) + \mathbf{b}(\mathbf{p}) \cdot u(t) \tag{8}$$

with the system matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, the input vector $\mathbf{b} \in \mathbb{R}^{n \times 1}$, the vector of system states $\mathbf{x} \in \mathbb{R}^{n \times 1}$, the parameter vector $\mathbf{p} \in \mathbb{R}^{n_p \times 1}$, and the input $u \in \mathbb{R}^{1 \times 1}$. The output equation is given by

$$y(t) = \mathbf{c}^T(\mathbf{p}) \cdot \mathbf{x}(t) \tag{9}$$

with the output vector $\mathbf{c} \in \mathbb{R}^{n \times 1}$. Note that $\mathbf{A}$, $\mathbf{b}$ and $\mathbf{c}$ are parameter-dependent in general. A suitable control law for system (8) can be stated as

$$u(t) = u_{\mathrm{FF}}(t) - \mathbf{k}^T \mathbf{x}(t) \tag{10}$$

with the gain vector $\mathbf{k}^T$ and a feedforward control $u_{\mathrm{FF}}(t)$ if the pair $(\mathbf{A}, \mathbf{b})$ is controllable.

For the calculation of the controller gains, several methods are possible: pole placement, solving the Riccati equation, or linear matrix inequalities [5, 23].

A feedforward control u_{FF} allows for tracking a desired trajectory or to reach a stationary working point using a static or a dynamic feedforward control [27]. Optimal tracking behavior can be achieved, if the transfer function of the closed-loop system is equal to 1 according to (identity matrix $\mathbf{I} \in \mathbb{R}^{n \times n}$, Laplace variable λ)

$$G(\lambda) = \mathbf{c}^T(\lambda \cdot \mathbf{I} - \mathbf{A} + \mathbf{b}\mathbf{k}^T)^{-1} \cdot \mathbf{b} \cdot G_{\mathrm{FF}}(\lambda) \stackrel{!}{=} 1. \tag{11}$$

This means that the system output follows the desired trajectory in the ideal case.

Using a static feedforward control law leads to steady-state accuracy (all time derivatives are zero) according to

$$u_{\mathrm{FF}}(t) = u_{\mathrm{FF,stat}}(t) = S \cdot y_{\mathrm{d}}(t), \tag{12}$$

where $y_{\mathrm{d}}(t)$ is the desired output trajectory and the gain S can be calculated by

$$S = G_{\mathrm{FF}}(\lambda = 0) = \left(\mathbf{c}^T \cdot \left(\mathbf{b}\mathbf{k}^T - \mathbf{A}\right)^{-1} \cdot \mathbf{b}\right)^{-1}. \tag{13}$$

A dynamic feedforward control for the closed-loop system can be easily computed in the Laplace domain (controller gain $\mathbf{k}^T = [k_1\ k_2\ \ldots\ k_n]$) using

$$\begin{aligned} G(\lambda) = \frac{Y(\lambda)}{W(\lambda)} &= \mathbf{c}^T(\lambda \cdot \mathbf{I} - \mathbf{A} + \mathbf{b}\mathbf{k}^T)^{-1} \cdot \mathbf{b} \cdot G_{\mathrm{FF}}(\lambda) \\ &= \frac{\gamma_0 + \gamma_1\lambda + \cdots + \gamma_m\lambda^m}{a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n} \cdot G_{\mathrm{FF}}(\lambda), \end{aligned} \tag{14}$$

where the polynomial degree n of the denominator is equal to the system order and m is the numerator degree. In order to achieve perfect trajectory tracking according to $G(\lambda) = 1$, the numerator has to be extended, such that numerator and denominator degrees become equal $m \stackrel{!}{=} n$. According to

$$G(\lambda) = \frac{Y(\lambda)}{W(\lambda)} \stackrel{!}{=} \frac{v_0 + v_1\lambda + \cdots + v_n\lambda^n}{a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n}, \tag{15}$$

at least $n - m$ time derivatives of the desired trajectory have to be specified in the general case. This leads to

$$G_{\mathrm{FF}}(\lambda) = \frac{1}{\gamma_0} \cdot [a_0, a_1, \ldots, 1] \cdot [W(\lambda), \lambda W(\lambda), \ldots, \lambda^{(n)} W(\lambda)]^T \tag{16}$$

in the special case that the original numerator degree m is equal to zero. The necessary $n - m$ coefficients v_i for the dynamic feedforward control result from comparison of the numerator and denominator coefficients of the same order by

$$v_i = a_i \ \text{ for } \ i = 0, \ldots, n-1 \ \text{ if } \gamma_0 \neq 0 \text{ and all } \gamma_1, \ldots, \gamma_n = 0\ ,\ v_n = 1. \tag{17}$$

Since $\gamma_1 = \cdots = \gamma_n = 0$ holds for all applications in this chapter, the problem for the realization of

$$u_{\mathrm{FF}}(t) = u_{\mathrm{FF,dyn}}(t) = \frac{1}{\gamma_0} \cdot \mathbf{v} \cdot [w(t), \dot{w}(t), \ddot{w}(t), \ldots, w^{(n)}(t)]^T \tag{18}$$

with $\mathbf{v} = [v_0, v_1, \ldots, v_n]^T$ is on the one hand, that consistency of the time derivatives of the desired trajectory is necessary. On the other hand, the dynamic feedforward control often depends on system parameters. Here, in this chapter, it is assumed that these are not exactly known so that within the feedforward control the midpoints of the parameter intervals have to be used. If these assumptions do not describe the real system with good accuracy, the sliding mode part described in the following section will minimize the remaining deviations.

Up to this point, the control law (10) is not applicable if not all system states are measurable. Therefore, the non-measurable ones have to be estimated using an appropriate observer. According to the duality principle, an observer for the estimation of all system states $\hat{\mathbf{x}}(t)$ can be realized in a similar way as it is done for state feedback control. The adapted control law yields

$$u(t) = u_{\mathrm{FF}}(t) - \mathbf{k}^T \hat{\mathbf{x}}(t). \tag{19}$$

The estimated state vector $\hat{\mathbf{x}}(t)$ includes the reconstruction of non-measurable states and a model-based filtering of the measurable ones. Therefore, the following ordinary differential equations (ODEs)

$$\dot{\hat{\mathbf{x}}}(t) = \hat{\mathbf{A}}(\mathbf{p}) \cdot \hat{\mathbf{x}}(t) + \hat{\mathbf{b}}(\mathbf{p}) \cdot u(t) + \mathbf{h}_{\mathrm{p}} \cdot (y_{\mathrm{m}}(t) - \hat{y}(t)) \;\text{ with }\; \hat{y}(t) = \hat{\mathbf{c}}^T(\mathbf{p}) \cdot \hat{\mathbf{x}}(t) \tag{20}$$

can be used with the observer gain vector $\mathbf{h}_{\mathrm{p}} \in \mathbb{R}^{n \times 1}$. Moreover, the system matrix $\hat{\mathbf{A}}$ as well as the input and output vectors $\hat{\mathbf{b}}$ and $\hat{\mathbf{c}}$ of the observer in (20) are equal to the ones in (8) and (9). The observer gain can be calculated in analogy to the controller gain [27].

5 Sliding Mode Control Procedures

The principle of sliding mode procedures—in control and estimation tasks—is to define a stable operation mode, the so-called sliding variable or surface. Due to the design of sliding modes, the system always tends to this stable mode and stays in its surrounding area [3]. Generally, sliding mode procedures are well known to be robust against uncertainty and to stabilize the system dynamics [26]. The control law for a system in nonlinear controller canonical form is in the following divided into a linear and a variable-structure part. The first one is already described in the previous section. The latter one consists of a switching amplitude and the sign function of the sliding variable which is in the following the difference between the desired trajectories and the current states (tracking error). The sign function determines in this context on which side of the sliding surface the system states are located in the actual time step.

The following sections describe, first, a common way to implement a sliding mode controller by defining the switching amplitude which is done by trial-and-error. Second, a possibility to calculate the switching amplitude online in each time step will be proposed using Lyapunov functions. Moreover, these two approaches differ in the dimensions of the sliding surface and switching amplitude (vector).

5.1 Common Approach for First-Order Sliding Mode Control

In this section, a common sliding mode control is shown for SISO (single-input single-output) systems based on [16, 24]. For this purpose, a Lyapunov function depending on the tracking error and its time derivative follow according to

$$V = \frac{1}{2} s^2(t) \quad \text{and} \quad \dot{V} = s(t) \cdot \dot{s}(t). \tag{21}$$

In Eq. (21), s is the so-called *sliding variable* which can be defined as a Hurwitz polynomial of order $n-1$ for a system in nonlinear controller canonical form according to

$$s(t) = \kappa_0 \cdot (x_1(t) - x_{1,\mathrm{d}}(t)) + \kappa_1 \cdot (x_2(t) - x_{2,\mathrm{d}}(t)) + \cdots + (x_{n-1}(t) - x_{n-1,\mathrm{d}}(t)). \tag{22}$$

The output of the system is denoted by the first state $y = x_1$ and its time derivatives of order $i-1$ by x_i with $i \in \{2, \ldots, n\}$. Analogously, the desired trajectory of the output is $y_\mathrm{d} = x_{1,\mathrm{d}}$ and the terms $y_\mathrm{d}^{(1)} = x_{2,\mathrm{d}}, \ldots, y_\mathrm{d}^{(n-1)} = x_{n,\mathrm{d}}$ are associated with its time derivatives.

In order to guarantee asymptotic stability of the system, the condition $\dot{V} < 0$ has to be fulfilled for all $s \neq 0$. This can be achieved by defining a scalar switching amplitude η in a suitable way by enforcing proportionality of the time derivative of the Lyapunov function and the absolute value of the sliding variable $|s(t)| = s(t) \cdot \mathrm{sign}(s(t))$ according to

$$\dot{V}(t) \overset{!}{\leq} -\eta \cdot |s(t)| \text{ with } \eta > 0. \tag{23}$$

This inequality is equal to

$$s(t) \cdot \dot{s}(t) \overset{!}{\leq} -\eta \cdot s(t) \cdot \mathrm{sign}(s(t)). \tag{24}$$

After factoring out $s(t)$ on both sides and some reformulations [16], the expression

$$\dot{s}(t) + \eta \cdot \mathrm{sign}(s(t)) \overset{!}{\leq} -\varepsilon \cdot \mathrm{sign}(s(t)) \tag{25}$$

can be obtained with a desired convergence rate $\varepsilon > 0$ which then can be stated as

$$\dot{s}(t) + \underbrace{(\eta + \varepsilon)}_{:=\tilde{\eta}>0} \cdot \operatorname{sign}(s(t)) \overset{!}{\leq} 0. \tag{26}$$

In Eq. (26), the sum of the switching amplitude $\eta > 0$ and the convergence rate ε can be stated as $\tilde{\eta} > 0$.

The sign function of the sliding variable can be evaluated in each time step using the following cases in dependency of the positive or negative sign of the sliding variable according to

$$\operatorname{sign}(s(t)) = \begin{cases} 1, & \text{if } s(t) > 0 \\ -1, & \text{if } s(t) < 0 \\ 0, & \text{else.} \end{cases} \tag{27}$$

In such a way, the switching part can change its sign in the control law depending on the system dynamics instantaneously in each time step which causes chattering. However, after reaching the stable area $s(t) \approx 0$, the system will not diverge anymore and stay in a small enclosure around it.

To guarantee asymptotic stability denoted by $\dot{V} < 0$, the terms $\kappa_i > 0$ and $\tilde{\eta} > 0$ have to be defined properly.

The control law of the dynamic system $\dot{x}_\mathrm{n} = x_1^{(n)} = u$ for the common sliding mode yields

$$u(t) = x_{1,\mathrm{d}}^{(n)}(t) - \sum_{i=0}^{n-1} \kappa_i \cdot (x_1^{(i)}(t) - x_{1,\mathrm{d}}^{(i)}(t)) - \tilde{\eta} \cdot \operatorname{sign}(s(t)) \tag{28}$$

with the constant scalar switching amplitude $\tilde{\eta}$ and the scalar sliding variable $s(t)$. Note that in the control law (28), which is used for the common sliding mode control, a dynamic feedforward control is inherently included due to the choice of $s(t)$ in Eq. (22).

To replace the intelligent guess of $\tilde{\eta}$ by a systematic design procedure, the next section describes a sliding mode controller that calculates a switching amplitude vector—one switching amplitude for every state in each time step—in order to reduce chattering. This is done adaptively by taking into account bounded uncertainty in the sense of defined range enclosures (intervals, see next subsection) as well as by considering stochastic noise processes in the system model.

5.2 Interval Definitions in Sliding Mode Approaches

Intervals can be understood as range bounds for several purposes. First, not exactly or also varying parameters can be defined as intervals. Then, the parameter-dependent system matrix $\mathbf{A}(\mathbf{p}(t))$ is extended to an interval evaluation $\mathbb{A} := [\mathbf{A}(\mathbf{p}(t))] = \mathbf{A}([\mathbf{p}](t))$. The same also holds for the input vector $\mathbb{B} := [\mathbf{b}(\mathbf{p}(t))] = \mathbf{b}([\mathbf{p}](t))$. In the presented application scenarios, the output vector is parameter-independent, so that $\mathbb{C} := \mathbf{c}^T$ holds. Analog enclosures are taken into consideration in this chapter for control errors (e.g., if the system states have not yet reached the desired trajectory), estimation errors in an observer for the estimation of parameters and states (e.g., if the reconstructed states do not coincide with the measured ones) and measurement errors (due to sensor inaccuracies that provide measurements only with a limited accuracy or resolution) as it can be seen in Fig. 1.

Taking into account such intervals, the sliding surfaces in control (and in analogy also estimation) become range bounds around the value 0 (see Fig. 2). This aims at reducing the calculated switching amplitude in regions, where the tracking error is sufficiently small to avoid unnecessary large switching amplitudes and actuator effort. Taking into account these definitions, the interval-based sliding mode controller (ISMC) is described in the next subsection.

5.3 Sliding Mode Control Using Intervals (ISMC)

In the following, the control law (19) is extended to

$$u(t) = u_{\mathrm{FF}}(t) - \mathbf{k}^T \cdot \hat{\mathbf{x}}(t) + \boldsymbol{\eta}^T \cdot \mathrm{sign}(\mathbf{s}(t)) \tag{29}$$

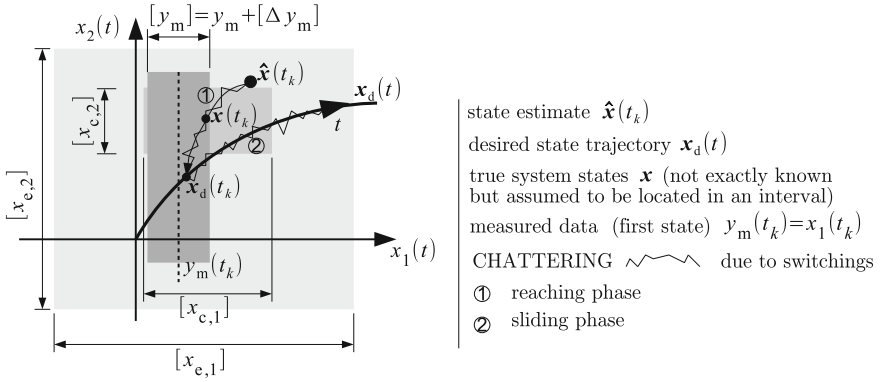


Fig. 1 Interval definitions of control, estimation, and measurement errors illustrated for a system with two states

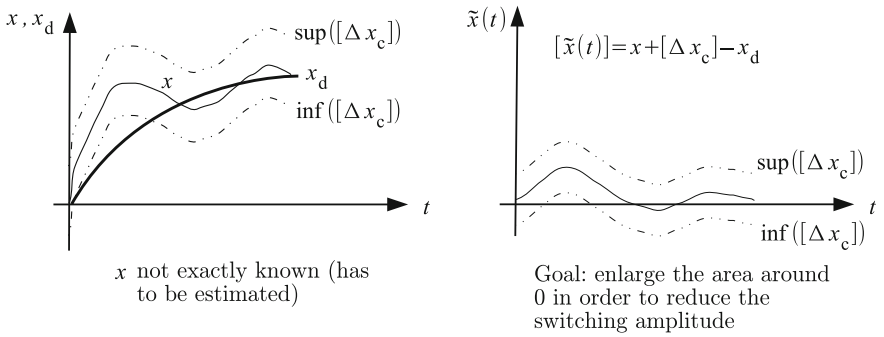


Fig. 2 Interval definition of the tracking error

with the vector of switching amplitudes $\boldsymbol{\eta} \in \mathbb{R}^{n\times 1}$ and a vector-valued calculation of the sign function of the sliding surface vector $\mathbf{s}(t) = \tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \mathbf{x}_\mathrm{d}(t) \in \mathbb{R}^{n\times 1}$ which includes the tracking errors between the state variables and the corresponding desired trajectories in a component-wise manner.

The extended controller can be applied to an uncertain system described by the following stochastic differential equations for a SISO system [24]:

$$\mathrm{d}\mathbf{x}(t) = \mathbf{F}(\mathbf{x}\,(t)\,, [\mathbf{p}], \mathbf{u}(t))\mathrm{d}t + \mathbf{G}_\mathrm{p} \cdot \mathrm{d}\mathbf{w}_\mathrm{p}(t) \;\text{ and} \tag{30}$$

$$y(t) = \mathbf{c}^T\,([\mathbf{p}]) \cdot \mathbf{x}\,(t) + G_\mathrm{m} \cdot \mathrm{d}w_\mathrm{m}(t). \tag{31}$$

In Eq. (30), $\mathrm{d}\mathbf{w}_\mathrm{p} \in \mathbb{R}^{n\times 1}$ is the standard Brownian motion of the process noise and $\mathbf{G}_\mathrm{p} \in \mathbb{R}^{n\times n_w}$ the corresponding matrix of standard deviation ($n_w = \dim(\mathrm{d}\mathbf{w}_\mathrm{p})$). In Eq. (31), $\mathrm{d}w_\mathrm{m}$ is the standard Brownian motion that represents the measurement errors with G_m as the related standard deviation for the system output (which is scalar in this contribution, $\dim(y) = \dim(\mathrm{d}w_\mathrm{m}) = 1$).

The uncertain SISO system

$$\mathbf{F}(\mathbf{x}\,(t)\,, [\mathbf{p}], \mathbf{u}(t)) = \mathbf{f}(\mathbf{x}\,(t)\,, [\mathbf{p}], \mathbf{u}(t)) + \mathbf{g}(t) \tag{32}$$

can be stated into the following form:

$$\begin{aligned}\mathbf{f}([\mathbf{x}](t), [\mathbf{p}], u(t)) &= \mathbb{A} \cdot [\mathbf{x}](t) + \mathbb{B} \cdot u(t) \\ &= \mathbb{A} \cdot [\mathbf{x}](t) + \mathbb{B} \cdot (u_\mathrm{FF}(t) - \mathbf{k}^T \cdot [\mathbf{x}](t) + \boldsymbol{\eta}^T \cdot \mathrm{sign}([\tilde{\mathbf{x}}](t))),\end{aligned} \tag{33}$$

where $\mathbf{g}\,(t)$ includes all influences, which do not fit into the linear form in Eq. (33). In such a way, the term $\mathbf{g}\,(t)$ contains nonlinear dynamics and not explicitly modeled terms, for example, deterministic parts of friction.

Moreover, $\mathbb{A}$ and $\mathbb{B}$ are the interval notations for the system matrix and the input vector (see Sect. 5.2). Analogously, the system states $[\mathbf{x}](t)$ are extended by some tolerance bounds to include control errors due to Fig. 2.

For the guaranteed stabilizing calculation of the switching amplitude vector, the Itô differential operator [12]

$$L([V](t)) = \frac{\partial V}{\partial t} + \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T \cdot (\mathbf{F}([\mathbf{x}](t), [\mathbf{p}], \mathbf{u}(t)) - \dot{\mathbf{x}}_\mathrm{d}(t)) + \frac{1}{2}\mathrm{trace}\left\{\mathbf{G}_\mathrm{p}^T \frac{\partial^2 V}{\partial \tilde{\mathbf{x}}^2}\mathbf{G}_\mathrm{p}\right\} \tag{34}$$

is employed with the Lyapunov function candidate

$$V(t) = \frac{1}{2}\tilde{\mathbf{x}}^T(t) \cdot \mathbf{P} \cdot \tilde{\mathbf{x}}(t) \tag{35}$$

and its partial derivative with respect to the tracking error $\frac{\partial V}{\partial \tilde{\mathbf{x}}}$ (for detailed information, refer to the appendix). The tracking error $\tilde{\mathbf{x}}(t)$ serves as a generalization of the sliding variable

$$\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \mathbf{x}_\mathrm{d}(t) = [x_1(t) - x_{1,\mathrm{d}}(t)\ ,\ x_2(t) - x_{2,\mathrm{d}}(t), \ldots, x_n(t) - x_{n,\mathrm{d}}(t)]^T. \tag{36}$$

According to the definition of V, it is obvious that no explicit time dependency occurs so that the term $\frac{\partial V}{\partial t} = 0$ can be removed from (34). Moreover, the system equations $\mathbf{F}([\mathbf{x}](t), [\mathbf{p}], \mathbf{u}(t))$ and the sliding variable are evaluated using a control error interval $[\Delta\mathbf{x}_\mathrm{c}]$ for the deviation of the true unknown system states from the desired trajectories according to $\mathbf{x}(t) \in [\mathbf{x}](t) = \mathbf{x}(t) + [\Delta\mathbf{x}_\mathrm{c}]$ and with the parameter intervals $\mathbf{p} \in [\mathbf{p}]$. The term

$$T_\mathrm{S} = \frac{1}{2}\mathrm{trace}\left\{\mathbf{G}_\mathrm{p}^T \cdot \left(\frac{\partial^2 V}{\partial \tilde{\mathbf{x}}^2}\right) \cdot \mathbf{G}_\mathrm{p}\right\} = \frac{1}{2}\mathrm{trace}\left\{\mathbf{G}_\mathrm{p}^T \cdot \mathbf{P} \cdot \mathbf{G}_\mathrm{p}\right\} \tag{37}$$

denotes the trace of the stochastic processes, in fact process noise for control purposes. Therefore, the matrix of standard deviation of the process noise $\mathbf{G}_\mathrm{p}$ is taken into consideration as well as the solution of the Lyapunov equation given by (7).

Inserting (33) into (34) yields

$$\begin{aligned} L([V](t)) = &\underbrace{[\tilde{\mathbf{x}}]^T(t) \cdot \mathbf{P} \cdot \left(\mathbb{A} - \mathbb{B} \cdot \mathbf{k}^T\right) \cdot [\mathbf{x}](t) + [\tilde{\mathbf{x}}]^T(t) \cdot \mathbf{P} \cdot \mathbb{B} \cdot u_\mathrm{FF}(t) - [\tilde{\mathbf{x}}]^T(t) \cdot \mathbf{P} \cdot \dot{\mathbf{x}}_\mathrm{d}(t)}_{[\dot{V}_\mathrm{a}](t)} \\ &+ [\tilde{\mathbf{x}}]^T(t) \cdot \mathbf{P} \cdot \mathbb{B} \cdot \boldsymbol{\eta}^T \cdot \mathrm{sign}([\tilde{\mathbf{x}}](t)) + T_\mathrm{S} \end{aligned} \tag{38}$$

with $[\tilde{\mathbf{x}}](t) = [\mathbf{x}](t) - \mathbf{x}_\mathrm{d}(t)$, $\mathbf{g}(t) = \mathbf{0}$.

Solving this relation with respect to the switching amplitude vector $\boldsymbol{\eta}$ by employing the condition $L(V(t)) \overset{!}{<} -\mathbf{q}^T \cdot \text{abs}\,([\tilde{\mathbf{x}}](t)) \in \mathbb{R}^{n\times 1}$ equals

$$[\dot{V}_\text{a}](t) + [\tilde{\mathbf{x}}]^T(t) \cdot \mathbf{P} \cdot \mathbb{B} \cdot \boldsymbol{\eta}^T \cdot \text{sign}([\tilde{\mathbf{x}}](t)) + T_\text{S} < -\mathbf{q}^T \cdot \text{abs}\,([\tilde{\mathbf{x}}](t)) \tag{39}$$

which can be reformulated according to

$$[\dot{V}_\text{a}](t) + \boldsymbol{\eta}^T \cdot \underbrace{(\text{sign}([\tilde{\mathbf{x}}](t)) \cdot [\tilde{\mathbf{x}}]^T(t))}_{|[\tilde{\mathbf{x}}]|^T(t) \in \mathbb{R}^{n\times n}} \cdot \mathbf{P} \cdot \mathbb{B} + T_\text{S} < -\mathbf{q}^T \cdot \text{abs}\,([\tilde{\mathbf{x}}](t)) \tag{40}$$

with a user-defined piecewise non-negative convergence rate vector $\mathbf{q} \in \mathbb{R}^{n\times 1} > \mathbf{0}$. Moreover, the vector of the absolute values abs $([\tilde{\mathbf{x}}](t))$ is defined as

$$\text{abs}\,([\tilde{\mathbf{x}}](t)) = \begin{bmatrix} \left|[x_1](t) - x_{1,\text{d}}(t)\right| \\ \left|[x_2](t) - x_{2,\text{d}}(t)\right| \\ \vdots \\ \left|[x_n](t) - x_{n,\text{d}}(t)\right| \end{bmatrix}. \tag{41}$$

The matrix of the absolute values $|[\tilde{\mathbf{x}}]|\,(t) \in \mathbb{R}^{n\times n}$ with the definition of the tracking error interval $[\tilde{x}_i](t) = [x_i](t) - x_{i,\text{d}}(t)$ for all $i \in \{1, \ldots, n\}$ is defined as

$$|[\tilde{\mathbf{x}}](t)| = \begin{bmatrix} [\tilde{x}_1](t) \cdot \text{sign}([\tilde{x}_1](t)) & [\tilde{x}_1](t) \cdot \text{sign}([\tilde{x}_2](t)) & \ldots & [\tilde{x}_1](t) \cdot \text{sign}([\tilde{x}_n](t)) \\ [\tilde{x}_2](t) \cdot \text{sign}([\tilde{x}_1](t)) & [\tilde{x}_2](t) \cdot \text{sign}([\tilde{x}_2](t)) & \ldots & [\tilde{x}_2](t) \cdot \text{sign}([\tilde{x}_n](t)) \\ \vdots & \vdots & \ddots & \vdots \\ [\tilde{x}_n](t) \cdot \text{sign}([\tilde{x}_1](t)) & [\tilde{x}_n](t) \cdot \text{sign}([\tilde{x}_2](t)) & \ldots & [\tilde{x}_n](t) \cdot \text{sign}([\tilde{x}_n](t)) \end{bmatrix}. \tag{42}$$

Additionally, the sign function of the tracking error components is given by

$$\text{sign}([\tilde{x}_i](t)) = \begin{cases} 1, & \text{if } \inf([\tilde{x}_i](t)) > 0 \\ -1, & \text{if } \sup([\tilde{x}_i](t)) < 0 \\ 0, & \text{else.} \end{cases} \tag{43}$$

This sign function causes the mentioned chattering phenomenon: if $\text{sign}([\tilde{x}_i](t)) < 0$ in time step t_k, it is possible that $\text{sign}([\tilde{x}_i](t)) > 0$ may occur in the following time step t_{k+1}. Sliding mode theory assumes that the switching part can change arbitrarily between two subsequent time steps [6]. In practice, this is not reasonable, because a large switching amplitude vector involves a large value of the input signal. To prevent wear or even damage of the actuator of the real system, the input signal is usually limited, e.g., an electric motor with a limited motor torque and a corresponding variation rate. Therefore, the frequency of the switching amplitude has to be limited to a maximum value $\boldsymbol{\eta}_\text{max}$ to ensure that its values do not vary arbitrarily between two time steps.

From Eq. (40), the expression

$$[\dot{V}_\mathrm{a}](t) + \boldsymbol{\eta}^T \cdot \underbrace{|[\tilde{\mathbf{x}}](t)|^T \cdot \mathbf{P} \cdot \mathbb{B}}_{:=[\mathbf{M}]} + T_\mathrm{S} < -\mathbf{q}^T \cdot \mathrm{abs}\,([\tilde{\mathbf{x}}](t)) \tag{44}$$

can then be solved with respect to the switching amplitude vector $\boldsymbol{\eta}$. Its components $i \in \{1, \ldots, n\}$ can be calculated in each time step by one of the following cases

$$\eta_i = \begin{cases} \sup\left([\mathbf{M}]_i^+ \cdot \left(-[\dot{V}_\mathrm{a}](t) - \mathbf{q}^T \cdot \mathrm{abs}\,([\tilde{\mathbf{x}}](t)) - T_\mathrm{S}\right)\right) + \varepsilon, & \text{if } \sup([\mathbf{M}]_i) < 0 \\ \inf\left([\mathbf{M}]_i^+ \cdot \left(-[\dot{V}_\mathrm{a}](t) - \mathbf{q}^T \cdot \mathrm{abs}\,([\tilde{\mathbf{x}}](t)) - T_\mathrm{S}\right)\right) - \varepsilon, & \text{if } \inf([\mathbf{M}]_i) > 0 \\ 0, & \text{else} \end{cases} \tag{45}$$

with the pseudoinverse $(\ldots)^+$ of the matrix $[\mathbf{M}]^T := \mathbb{B}^T \cdot \mathbf{P} \cdot |[\tilde{\mathbf{x}}](t)|$ which turns out to be a row vector for SISO systems. Note, ε is a small value guaranteeing the strict inequality (44) after solving for the switching amplitude. These three cases are necessary to determine on which side of the sliding surface the system states are located or if they have already reached the small area around it (case 3).

The i-th component of the switching amplitude depends on the corresponding entry $[\mathbf{M}]_i$, where either the infimum or the supremum is relevant. As a consequence, the switching amplitude entries are calculated separately as small as possible. Moreover, the switching amplitude is evaluated at each time step t_k with actual information about current states and desired trajectories. The case $\eta_i = 0$ means that the sliding surface is practically reached or the system stays in its near surrounding area.

In general, there are different ways to calculate the pseudoinverse of $[\mathbf{M}]$ in (45). If $[\mathbf{M}]$ is a matrix, then the minimum norm solution according to $[\mathbf{M}]^+ = \left([\mathbf{M}]^T \cdot [\mathbf{M}]\right)^{-1} \cdot [\mathbf{M}]^T$ has to be used. If $[\mathbf{M}]$ is a vector as it is in both application scenarios discussed in this chapter, the pseudoinverse can also be calculated by general-purpose routines such as `minv` implemented in C- XSC [11]. However, for the application scenario 2, the minimum norm solution provides too conservative calculations so that the third case of (45) occurs unnecessarily often. Therefore, the pseudoinverse for SISO systems can also be calculated component-wise according to

$$[M]_i^+ = \frac{[M]_i}{||[\mathbf{M}]||^2} = \frac{[M]_i}{\sum_{i=1}^{n} [M]_i^2} \tag{46}$$

which typically leads to less conservative results.

As already mentioned, the described approach ensures the system's stability as long as $L(V(t)) < 0$ or $L(V(t)) < -\mathbf{q}^T \cdot \mathrm{abs}(|[\tilde{\mathbf{x}}](t)|)$ hold. The case $L(V(t)) = 0$ describes the boundary of the provable stability domain. Due to the chosen quadratic Lyapunov function with the weighting matrix $\mathbf{P}$, the boundary is shaped like an ellipsoid. In order to reduce the non-stabilizable area, in which the positive or negative

sign of the tracking error cannot be determined, the volume of $L(V(t))$ should be as small as possible [23].

As already mentioned, in cases in which not all system states are measurable, it is necessary to reconstruct them by a state observer. Alternatively, the control approach discussed in this contribution can be applied to estimation tasks, as it is described in [24] where not only non-measurable system states are estimated but simultaneously also selected uncertain system parameters of the presented application scenario.

The described approach is implemented using MATLAB/SIMULINK. Due to the fact that the switching amplitude vector is calculated in each discretization step, s-functions calling a C++ code are used to make the approach real-time applicable. For the calculation with intervals, the toolbox C- XSC [11] is used in simulation. However, this cannot be used in experiments of the first application scenario, because the test rig is actuated by a real-time target system which does not support the directed rounding mode needed in C- XSC. Therefore, a self-defined structure is written in C++ that is able to handle interval matrices and the standard operations $\{+, -, \cdot\}$. The / operator, which is in this approach only needed for the calculation of the pseudoinverse, is defined separately for both applications as it is already described in this section. Due to the fact that the considered intervals for parameters and control errors are at least ten orders of magnitude larger than rounding errors, the self-defined data type provides results that practically cannot be distinguished from calculations using C- XSC, where directed rounding is employed [11].

6 Application Scenario 1: Modeling and Experimental Setup of a Drive-Train Test Rig

The first application scenario is a drive-train test rig representing the longitudinal dynamics of a vehicle which is available at the Chair of Mechatronics at the University of Rostock, as it can be seen in Fig. 3 [24].

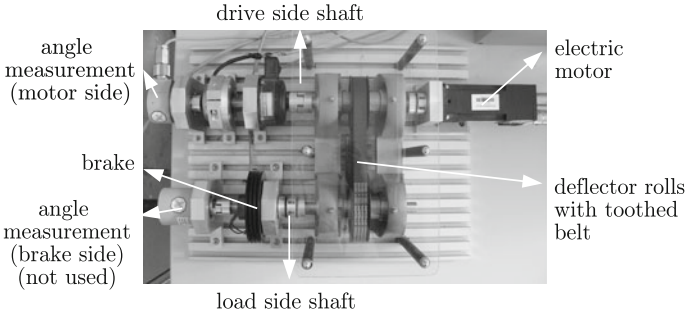


Fig. 3 Photo of the test rig

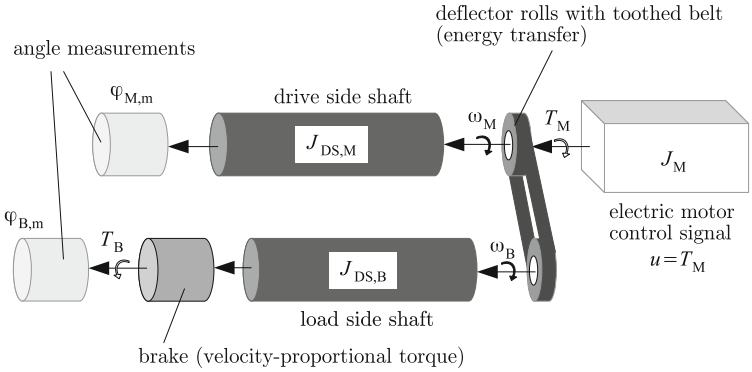


Fig. 4 Schematic visualization of the test rig

The principle is quite simple: an electric motor drives a shaft which is connected via two deflector rolls and one toothed belt with another shaft. On this second shaft, a velocity-proportional braking torque can be activated. Only the available measurement of the angle of the drive side shaft is used for control purposes. Figure 4 shows the schematic interpretation from the experimental setup in Fig. 3 with the most important variables for the mathematical model.

If the mass moments of inertia of the motor J_M, both shafts $J_{DS,M}$ and $J_{DS,B}$ as well as of the brake J_B are summarized into one common mass moment of inertia J, and if a velocity-proportional friction coefficient d is considered, the test rig can be modeled with two parameters, two system states, one input and one output (SISO system). Friction has an important influence on the system, because it occurs between the deflector rolls and the toothed belt as well as in all bearings. Moreover, static friction has to be compensated in order to start a desired trajectory causing rotation of the shafts without stick-slip effects.

Under these assumptions, the system $\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t), [\mathbf{p}], u(t)) = [\dot{x}_1(t) \quad \dot{x}_2(t)]^T$ can be described by a set of two ordinary differential equations according to

$$\dot{\mathbf{x}}(t) = \mathbf{A} \cdot \mathbf{x}(t) + \mathbf{b} \cdot u(t) = \begin{bmatrix} 0 & 1 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \beta \end{bmatrix} u(t) \tag{47}$$

with the system output

$$y(t) = \mathbf{c}^T \cdot \mathbf{x}(t) = x_1(t). \tag{48}$$

Physically, the first system state is equal to the angle of the drive side shaft $x_1(t) = \varphi_M(t)$ and the second state represents its angular velocity $x_2(t) = \dot{x}_1(t) = \dot{\varphi}_M(t) = \omega_M(t)$. Note, only the first state is measurable $y_m(t) = x_{1,m}(t) = \varphi_{M,m}(t)$. The input signal is defined as the commanded motor torque $u(t) = T_M(t)$ and the two parameters are defined by

$$\alpha = -\frac{d}{J} \in [\alpha] \quad \text{and} \quad \beta = \frac{1}{J} \in [\beta], \tag{49}$$

both influencing the system dynamics in a significant way [24]. It is assumed that the nominal system parameters are located in some intervals specified as range enclosures $[\alpha]$ and $[\beta]$ around nominal values, because they are uncertain, however, known within some tolerance bounds. Moreover, the multiplicative coupling between parameters with the time-varying states and the system input is in the focus of the described control procedure using sliding mode techniques with interval arithmetics and uncertainty. Furthermore, the influence of friction is not explicitly modeled and can be overcome by such robust control procedures.

6.1 Trajectory Tracking by Common Sliding Mode Control

The model of the application scenario described above is in controller canonical form. Consequently, the sliding surface for the common sliding mode control (see Sect. 5.1) is chosen according to

$$\begin{aligned} s(t) &= \kappa_0 \cdot (x_1(t) - x_{1,\mathrm{d}}(t)) + (x_2(t) - x_{2,\mathrm{d}}(t)) \\ &= \kappa_0 \cdot (\varphi_{\mathrm{M}}(t) - \varphi_{\mathrm{M,d}}(t)) + (\omega_{\mathrm{M}}(t) - \omega_{\mathrm{M,d}}(t)). \end{aligned} \tag{50}$$

The corresponding time derivative yields

$$\begin{aligned} \dot{s}(t) &= \kappa_0 \cdot (\dot{x}_1(t) - \dot{x}_{1,\mathrm{d}}(t)) + (\dot{x}_2(t) - \dot{x}_{2,\mathrm{d}}(t)) \\ &= \kappa_0 \cdot (\omega_{\mathrm{M}}(t) - \omega_{\mathrm{M,d}}(t)) + (\dot{\omega}_{\mathrm{M}}(t) - \dot{\omega}_{\mathrm{M,d}}(t)). \end{aligned} \tag{51}$$

After inserting the second state equation from (47)

$$\dot{\omega}_{\mathrm{M}}(t) = \alpha \cdot \omega_{\mathrm{M}}(t) + \beta \cdot u(t), \tag{52}$$

in (51), the expression

$$\dot{s}(t) = \kappa_0 \cdot (\omega_{\mathrm{M}}(t) - \omega_{\mathrm{M,d}}(t)) + (\alpha \cdot \omega_{\mathrm{M}}(t) + \beta \cdot u(t) - \dot{\omega}_{\mathrm{M,d}}(t)) \tag{53}$$

results. Inserting (47) and (53) into (26) results in the control law

$$u(t) = \frac{-\kappa_0(\omega_M(t) - \omega_{M,d}(t)) - (\alpha \cdot \omega_M(t) - \dot{\omega}_{M,d}(t)) - \tilde{\eta} \cdot \mathrm{sign}(s(t))}{\beta} \tag{54}$$

including the switching part $\tilde{\eta} \cdot \mathrm{sign}(s(t))$. Unfortunately, this common form of sliding mode control does not work in the experiment. This might be caused by the not modeled influence of the varying static as well as sliding friction, and the two degrees of freedom that have to be predefined and cannot be calculated by

model-based systematic procedures, namely the constant switching amplitude $\tilde{\eta}$ as well as κ_0. Moreover, noise processes cannot be considered explicitly. All these facts cause too conservative results, which make large switching amplitudes necessary and make the system unstable in experiments, where such large switching amplitudes are not applicable due to the limited control input represented by the motor torque.

6.2 Trajectory Tracking by Interval-Based Sliding Mode Control

The presented system model (47) is also applied to the interval-based sliding mode controller for the two kinds of static and dynamic feedforward control. For the dynamic one, the transfer function of (47) yields

$$\frac{Y(\lambda)}{U(\lambda)} = \frac{X_1(\lambda)}{U(\lambda)} = \frac{\beta}{\lambda^2 + \alpha\lambda}. \tag{55}$$

The dynamic feedforward control can easily be computed with the control gain $\mathbf{k}^T = [k_1 \;\; k_2]$ using

$$G(\lambda) = \mathbf{c}^T(\lambda \cdot \mathbf{I} - \mathbf{A} + \mathbf{b} \cdot \mathbf{k}^T)^{-1} \cdot \mathbf{b} \cdot G_{\mathrm{FF}}(\lambda) = \frac{\beta}{\lambda^2 + \lambda \cdot (\beta k_2 - \alpha) + \beta k_1} \cdot G_{\mathrm{FF}}(\lambda). \tag{56}$$

Because the order of the denominator is $n = 2$ and the order of the numerator $m = 0$, the numerator has to be adapted such that both orders are equal. Due of the fact that the output of the given system model is also the flat one,[2] this corresponds to solving (56) with respect to the desired trajectories $W(\lambda)$ and transforming back into original coordinates. This leads to the dynamic feedforward control

$$u_{\mathrm{FF}}(t) = u_{\mathrm{FF,dyn}}(t) = \frac{\ddot{x}_{1,\mathrm{d}}(t) + \dot{x}_{1,\mathrm{d}}(t) \cdot (\beta k_2 - \alpha) + \beta k_1 x_{1,\mathrm{d}}(t)}{\beta} \tag{57}$$

with the desired trajectory $x_{1,\mathrm{d}}(t) = \varphi_{\mathrm{M,d}}(t)$ and the corresponding two derivatives $\dot{x}_{1,\mathrm{d}}(t) = \dot{\varphi}_{\mathrm{M,d}}(t)$ and $\ddot{x}_{1,\mathrm{d}}(t) = \ddot{\varphi}_{\mathrm{M,d}}(t)$. For the parameters, the following assumptions are included $\alpha = \mathrm{mid}([\alpha])$ and $\beta = \mathrm{mid}([\beta])$ as the midpoints of the defined range bounds.

[2]The output of a SISO system can be stated as flat output, if the system input affects explicitly the n-th time derivative of the output but no smaller order ones.

As already mentioned, also a static feedforward control which leads to steady-state accuracy (if all time derivatives are zero) is possible according to

$$u_{\mathrm{FF}}(t) = u_{\mathrm{FF,stat}}(t) = S \cdot x_{1,\mathrm{d}}(t) \tag{58}$$

with the gain $S = k_1$ from Eq. (13).

Because only the angle and not the angular velocity is measurable, the duality principle of controllers can be applied to estimate $\dot{\varphi}_{\mathrm{M}}(t) = \omega_{\mathrm{M}}(t)$ using (20). The adapted control law for the linear part yields again

$$u(t) = u_{\mathrm{FF}}(t) - \mathbf{k}^T \cdot \hat{\mathbf{x}}(t) \tag{59}$$

with $\hat{\mathbf{x}}(t) = [\hat{x}_1(t)\ \ \hat{x}_2(t)]^T = [\hat{\varphi}_{\mathrm{M}}(t)\ \ \dot{\hat{\varphi}}_{\mathrm{M}}(t)]^T = [\hat{\varphi}_{\mathrm{M}}(t)\ \ \hat{\omega}_{\mathrm{M}}(t)]^T$ denoting the estimated state vector.

In simulation and experiment, the measurement noise is defined by the measurement error of $\Delta\varphi_{\mathrm{M}} = 0.1167\,\mathrm{rad}$ according to $\mathbf{G}_{\mathrm{m}} = 3 \cdot \Delta\varphi_{\mathrm{M}} = 0.35\,\mathrm{rad}$. The process noise is set to a diagonal matrix $\mathbf{G}_{\mathrm{p}} = \mathrm{diag}([0.7, 0.7])$ which allows for deviations of the angular velocity of about 5 % and of the angular acceleration of 28 %.

For the interval-based sliding mode control, the following interval notations of the system matrix and the input vector are used:

$$\mathbb{A} = \begin{bmatrix} 0 & 1 \\ 0 & [\alpha] \end{bmatrix} \text{ as well as } \mathbb{B} = \begin{bmatrix} 0 \\ [\beta] \end{bmatrix}. \tag{60}$$

The point values of the parameters are defined as $\alpha^* = -4.22$ and $\beta^* = 64.10$ in the linear control parts. The corresponding point-valued intervals are chosen according to $[\alpha] = [0.2\ ;\ 1.2] \cdot \alpha^* = [-5.07\ ;\ -0.84]$ and $[\beta] = [0.2\ ;\ 1.2] \cdot \beta^* = [12.82\ ;\ 76.92]$. Moreover, the vector-valued sliding variable equals

$$[\tilde{\mathbf{x}}](t) = [\mathbf{x}](t) - \mathbf{x}_{\mathrm{d}}(t) = [[\varphi_{\mathrm{M}}](t) - \varphi_{\mathrm{M,d}}(t),\ [\omega_{\mathrm{M}}](t) - \omega_{\mathrm{M,d}}(t)]^T \tag{61}$$

with $[\varphi_{\mathrm{M}}](t) = \varphi_{\mathrm{M}}(t) + [\Delta x_{\mathrm{c},\varphi}]$, $[\omega_{\mathrm{M}}](t) = \omega_{\mathrm{M}}(t) + [\Delta x_{\mathrm{c},\omega}]$, and the control error intervals for both states $[\Delta x_{\mathrm{c},\varphi}]$ and $[\Delta x_{\mathrm{c},\omega}]$. Note that ω_{M} cannot be measured and therefore needs to be estimated in experiment, so that it will be replaced in the following by $\hat{\omega}$. The next section shows simulative and experimental results of the interval-based sliding mode controller Eq. (29) as well as simulative results from the common sliding mode controller.

6.3 Numerical and Experimental Results

At first, numerical simulations are shown in this section for the following reason. Due to the fact that the electric drive is affected by actuator constraints, the switching amplitude in the control law has to be limited. Otherwise, it could happen that the value of the calculated control law is larger than the actuator limit. The consequence is that it would be no longer valid, because the exact value cannot be realized by the described quasi-continuous approach. This would cause errors in trajectory tracking which may even lead to an unstable closed-loop behavior. The calculated switching amplitude is significantly influenced by the control error interval vector. In dependency of the choice of this interval, the area in which the switching amplitude is active or not varies. It needs to be taken into consideration that the switching control part is based on the tracking errors of both states and not only of the controlled variable (here the angle). This means that a trade-off between the choices of the two components of the control error interval vector has to be made: a small control error interval for the angle leads also to a small tracking error of the angle but to a larger tracking error of the angular velocity. The same holds vice versa. To determine the best choice of the control error interval, the root mean square error of the two states can be calculated separately according to

$$\overline{\mu}_{\varphi} = \sqrt{\frac{1}{N} \cdot \sum_{k=1}^{N} (\varphi_{\mathrm{d}}(t_k) - \varphi_{\mathrm{M,m}}(t_k))^2} \quad \text{and} \quad \overline{\mu}_{\omega} = \sqrt{\frac{1}{N} \cdot \sum_{k=1}^{N} (\omega_{\mathrm{d}}(t_k) - \hat{\omega}(t_k))^2}. \tag{62}$$

Figures 5, 6, 7 and 8 show the simulation results for the common first-order sliding mode controller in comparison with the interval-based approach ($\eta_{\max} = 0.8$, dynamic feedforward control) assuming that both system states are measurable (no state observer is included). Figures 5 and 6 are obtained if the system is simulated with

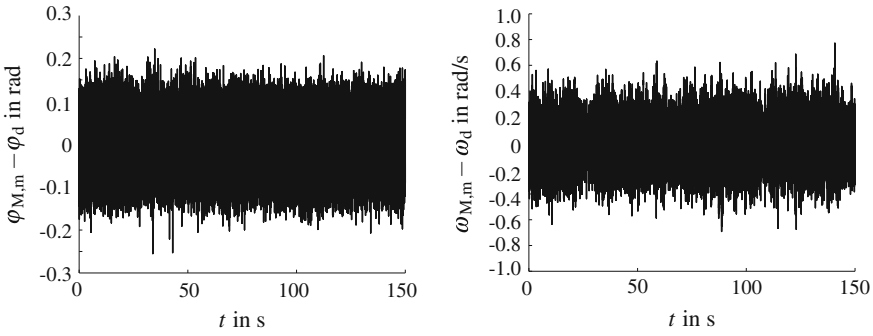


Fig. 5 Numerical results for the common sliding mode controller assuming all states are measurable with nominal parameters α^* and β^* as well as $\tilde{\eta} = 40$ and $\kappa_0 = 15$

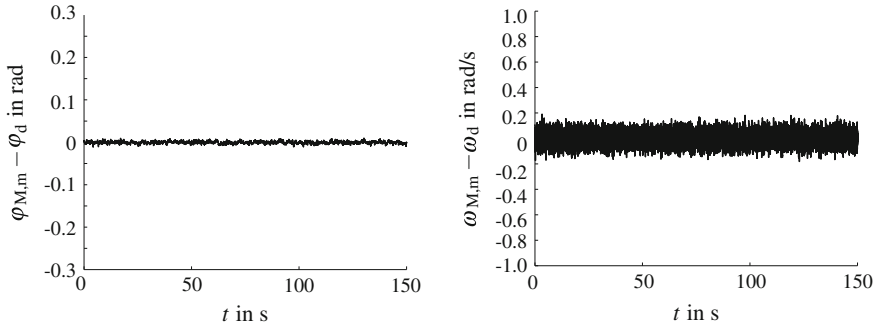


Fig. 6 Numerical results for the interval-based sliding mode controller assuming all states are measurable with nominal parameters α^* and β^* as well as $\eta_{max,i} = 0.8$

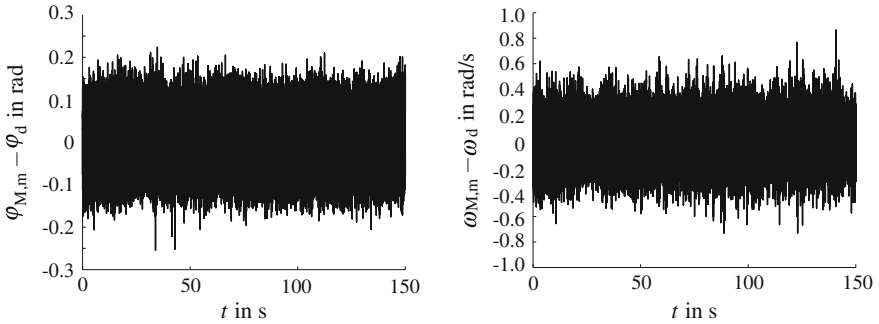


Fig. 7 Numerical results for the common sliding mode controller assuming all states are measurable with disturbed parameters $\breve{\alpha}$ and $\breve{\beta}$ as well as $\tilde{\eta} = 40$ and $\kappa_0 = 15$

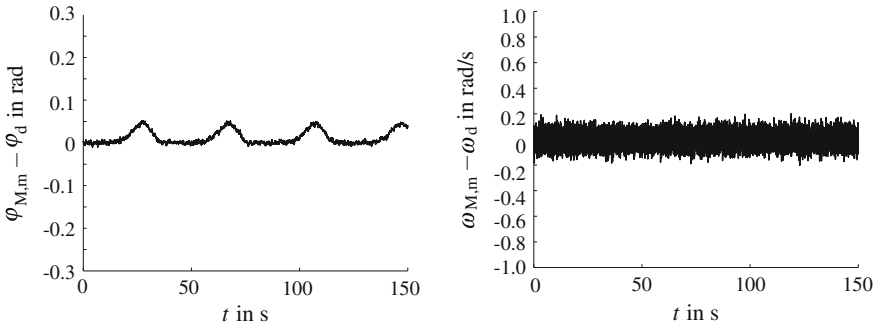


Fig. 8 Numerical results for the interval-based sliding mode controller assuming all states are measurable with disturbed parameters $\breve{\alpha}$ and $\breve{\beta}$ and $\eta_{max,i} = 0.8$

undisturbed parameters, whereas Figs. 7 and 8 show the results from simulations in which the system model is disturbed using $\breve{\alpha} = -3.37 \in [\alpha]$ and $\breve{\beta} = 51.28 \in [\beta]$. The linear part of the controller is still evaluated with the nominal values $\alpha^* = -4.22$ and $\beta^* = 64.10$. Obviously, the interval-based sliding mode provides much better results, in fact the tracking errors of both states can be decreased significantly not only in the undisturbed case but also in the disturbed one compared to the common sliding mode control.

To show how the choice of the control error interval influences the tracking errors of the system states, several simulations were performed. The corresponding root mean square errors can be seen in Table 1 taking into account dynamic feedforward control and a limitation of the switching amplitude with $\eta_{\max} = 0.8$. It can be seen that the smallest root mean square of the tracking error of the angle can be achieved with $[\Delta x_{\mathrm{c},\varphi}] = [\Delta x_{\mathrm{c},1}] = [-0.05\ ;\ 0.05]$. Simultaneously, the root mean square of the tracking error of the angular velocity with $[\Delta x_{\mathrm{c},\omega}] = [\Delta x_{\mathrm{c},2}] = [-1\ ;\ 1]$ is larger than in other specifications, as it can be seen in Table 1.

Experimental results show a similar behavior: depending on the choice of the control error interval, the root mean squares of the tracking errors interact as it can be seen in Table 2. Therefore, for comparison also the root mean square errors resulting from pure state feedback control without switching part are included. Note that the trajectory of the desired angle can be understood as a periodically repeated driving cycle consisting of an acceleration phase, a constant velocity phase, and a deceleration phase. One driving cycle has a duration of 40 s and a final angle of 150 rad. Additionally, the common sliding mode controller leads to unstable behavior in experiment. Therefore, only experimental results with the ISMC are shown.

In comparison to a control law with pure state feedback and without sliding mode part, the best improvement is ≈27 % with static and 16 % with dynamic feedforward control. Including dynamic feedforward control, Fig. 9 shows the desired angle with the corresponding measurement and the tracking error. In Fig. 10, the desired angular velocity (resulting from differentiation of the desired trajectory of the angle) with

Table 1 Numerical root mean square errors for angle and angular velocity with dynamic feedforward control and different control error intervals

$\Delta x_{\mathrm{c},\varphi}$	$\Delta x_{\mathrm{c},\omega}$	$\overline{\mu}_{\varphi}$ in rad	$\overline{\mu}_{\omega}$ in rad·s^{-1}
±0.05	±0.5	1.1003	0.5802
±0.05	±0.3	1.4288	0.5533
±0.05	±1.0	1.0677	0.5905
±0.05	±0.1	2.8211	0.5069
±0.5	±1.0	1.0770	0.5907
±0.5	±0.5	1.1207	0.5773
±0.1	±0.1	2.7414	0.5088

Desired angle $\varphi_{\mathrm{d}} = 150$ rad at the end of each driving cycle (repeated periodically), $\eta_{\max,i} = 0.8$, angular velocity estimated using a state observer

Table 2 Experimental root mean square errors for angle and angular velocity with static (top) as well as dynamic (bottom) feedforward control, different control error intervals and comparison to pure state feedback control (index SF)

$\Delta x_{\mathrm{c},\varphi}$	$\Delta x_{\mathrm{c},\omega}$	$\overline{\mu}_{\varphi}$ in rad	$\overline{\mu}_{\omega}$ in rad·s^{-1}	$\overline{\mu}_{\varphi,\mathrm{SF}}$	$\overline{\mu}_{\omega,\mathrm{SF}}$	Improvement for angle in %
±0.05	±10	1.11	4.11	1.53	4.12	22.67
±1	±10	1.17	4.00	1.53	4.12	26.99

$\Delta x_{\mathrm{c},\varphi}$	$\Delta x_{\mathrm{c},\omega}$	$\overline{\mu}_{\varphi}$ in rad	$\overline{\mu}_{\omega}$ in rad·s^{-1}	$\overline{\mu}_{\varphi,\mathrm{SF}}$	$\overline{\mu}_{\omega,\mathrm{SF}}$	Improvement for angle in %
±0.5	±10	8.14	3.55	9.77	4.22	16.63

Desired angle $\varphi_{\mathrm{d}} = 150$ rad at the end of each driving cycle (repeated periodically), limitation of the switching amplitude $\eta_{\max,i} = 0.8$. Angular velocity estimated using a state observer

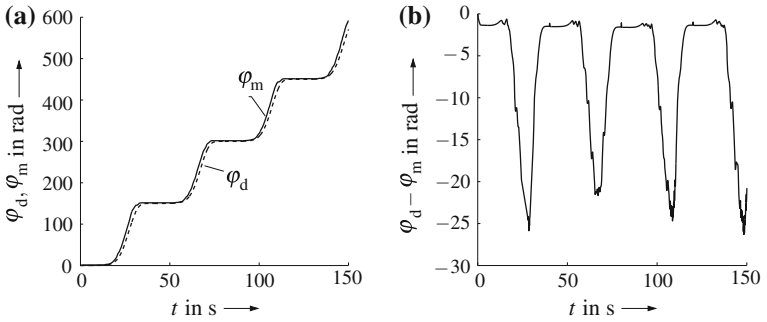


Fig. 9 Experimental results provided by the ISMC for the angle and the corresponding deviation including dynamic feedforward control. **a** Desired and measured angle. **b** Control error of the angle

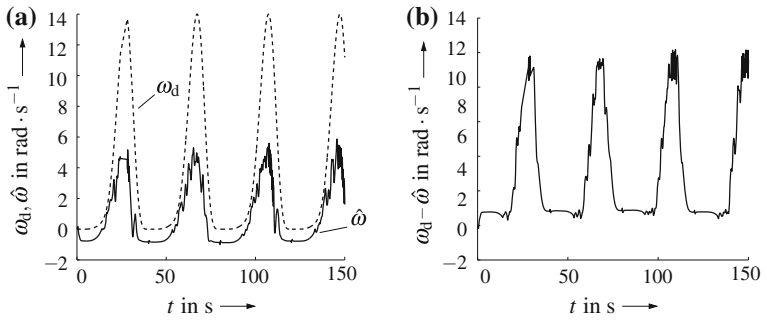


Fig. 10 Experimental results provided by the ISMC for the angular velocity and the corresponding deviation including dynamic feedforward control. **a** Desired and estimated angular velocity. **b** Control error of the angular velocity

the estimated angular velocity can be seen as well as the corresponding deviation. The estimated velocity is determined using a state observer.

In experiment, the ISMC using a static feedforward control provides better results than with dynamic feedforward control according to Table 2 which can also be seen in Figs. 11 and 12 compared with Figs. 9 and 10. This is caused by the assumption that the parameters are assumed to be constant over the whole driving cycle. Especially, the parameter α including velocity-proportional friction is—in all likelihood—not constant because of the high influence of static friction after each standstill. At these points, it is important at which position the toothed belt comes to a standstill before moving again. This effect is only indirectly included in the simple system description in terms of the process noise. In future work, this effect and the influence of time discretization errors have to be examined. In case that time discretization effects play a mayor role, an interval evaluation of the ordinary differential equations describing the system dynamics between two subsequent time steps would be necessary. Moreover, a discrete-time Lyapunov function could be taken into consideration instead of neglecting the time discretization completely. The reason for this possible extension

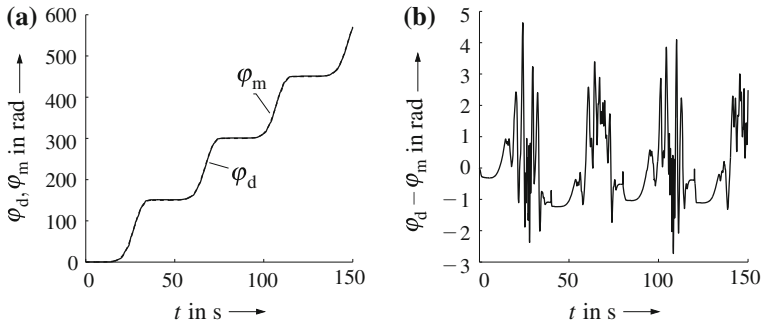


Fig. 11 Experimental results provided by the ISMC for the angle and the corresponding deviation including static feedforward control. **a** Desired and measured angle. **b** Control error of the angle

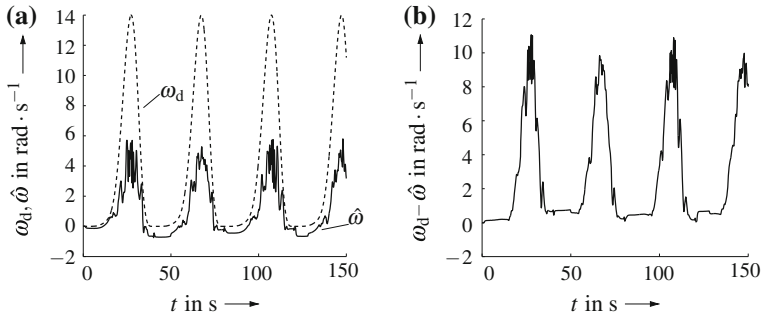


Fig. 12 Experimental results provided by the ISMC for the angular velocity and the corresponding deviation including static feedforward control. **a** Desired and estimated angular velocity. **b** Control error of the angular velocity

is that the switching amplitude is calculated in the experiment every 0.1 s to decrease the computational effort while the control sampling time is 0.01 s. This means that the values of $\boldsymbol{\eta}$ are assumed to be constant for 10 subsequent time steps.

Independently from the used type of feedforward control, knowledge about the system parameters is necessary. Therefore, the duality principle of the interval-based sliding mode control—namely an interval-based sliding mode observer—was used in previous work for the simultaneous estimation of states and parameters. Especially, the parameter identification provides better results than a state-of-the-art least squares approach [24, 25]. To show that the ISMC is also applicable to more complex systems, the temperature control of a solid oxide fuel cell (SOFC) system is considered in the next section.

7 Application Scenario 2: Thermal Subsystem of a Solid Oxide Fuel Cell System

In this section, the applicability of the interval-based sliding mode control approach is presented for the thermal subsystem of a high-temperature solid oxide fuel cell system available at the Chair of Mechatronics at the University of Rostock. In this chapter, only numerical results are presented. However, in future work an experimental validation will be realized.

Usually, for the industrial control of such a complex system, PI (proportional, integral) controllers are employed that only work for a specified operating point [20] at which the model is linearized for design purposes. To enlarge the controllable domain, sliding mode procedures provide the possibility to ensure robust control performance also in regions that deviate from the operating point. Moreover, the problems that parameters are not exactly known and non-modeled influences have to be compensated in a robust way can be overcome using sliding mode approaches.

In Fig. 13, the thermal subsystem which consists of two preheaters (one for the anode, one for the cathode gas) and the stack module is shown in a schematic way.

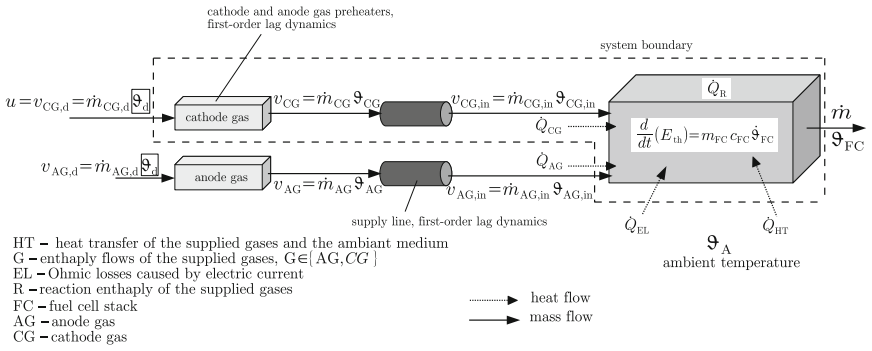


Fig. 13 Schematic visualization of the thermal subsystem of the SOFC stack with preheaters

In the stack, several planar fuel cells are collocated in electric series connection. For a complete modeling of the SOFC system, the reader is referred to [20]. The aim in the following is the temperature control of the stack. This is necessary to prevent over-temperatures in the stack material that may destroy the stack and to ensure optimal operating temperatures for energy-efficient chemical reactions. Therefore, the mathematical model considered in this contribution is based on [17, 20].

In general, the considered system is affected by several thermal effects and therefore by some distributed and not exactly known parameters. Usually, such a system is described by a nonlinear partial differential equation due to large ranges of operating temperatures [20]. However, these models are hardly suitable for real-time control approaches. Therefore, a set of ordinary differential equations is applied in the following. For its derivation, it is assumed that all dynamic states are homogeneously distributed in space. Because the ISMC is up to this point only applicable to systems whose mathematical model can be represented in nonlinear controller canonical form for SISO systems, the nonlinear ordinary differential equations have to be transformed using appropriate Lie derivatives of the system output. Because of the fact that the time delay of the preheaters cannot be neglected, they have to be taken into consideration in the system model as well. Since, in this chapter, only the heating phase is considered, the influence of the anode gas (here pure nitrogen) is interpreted as an external disturbance because the anode gas mass flow is significantly smaller than the cathode gas mass flow. Consequently, only the cathode gas serves as direct stack input. In future work, this special case of the heating phase will be combined with the high-temperature exothermic phase with active chemical reactions in order to take into account Ohmic losses and chemical reactions inside the stack [20].

The components depicted in Fig. 13 are briefly described in the following.

7.1 Nonlinear Dynamic Model

The preheater dynamics of the two supply gases are each described by two first-order lag dynamics. To obtain a SISO system,[3] the cathode gas can be described by

$$\dot{v}_{\mathrm{CG}}(t) = \frac{1}{T_{\mathrm{CG}}} \cdot \left(v_{\mathrm{CG,d}}(t) - v_{\mathrm{CG}}(t)\right) \quad \text{and} \tag{63}$$

$$\dot{v}_{\mathrm{CG,in}}(t) = \frac{1}{T_{\mathrm{CG,SL}}} \cdot \left(v_{\mathrm{CG}}(t) - v_{\mathrm{CG,in}}(t)\right) \tag{64}$$

with the time constant T_{CG} of the preheater and the time constant $T_{\mathrm{CG,SL}}$ of the supply line between the preheater and the stack. The second preheater for the anode

[3] In [20], the mathematical modeling was done as MIMO (multi-input multi-output) system, where both the anode and cathode gases are independent inputs into the stack. Because this chapter only considers the heating phase, the anode gas plays only a minor role and therefore can be interpreted as an external disturbance.

gas usually employs the gas mixture of nitrogen, hydrogen, methane, and carbon monoxide (all mixed before entering the preheater) to regulate the produced energy of the system. It is considered in analogy to the cathode gas during high-temperature phases. During the heating phase, it is treated as an external disturbance-like influence on the stack according to

$$\dot{v}_{\mathrm{AG}}(t) = \frac{1}{T_{\mathrm{AG}}} \cdot (v_{\mathrm{AG,d}}(t) - v_{\mathrm{AG}}(t)) \quad \text{and} \tag{65}$$

$$\dot{v}_{\mathrm{AG,in}}(t) = \frac{1}{T_{\mathrm{AG,SL}}} \cdot (v_{\mathrm{AG}}(t) - v_{\mathrm{AG,in}}(t)) \tag{66}$$

with the time constant T_{AG} of the preheater and the time constant $T_{\mathrm{AG,SL}}$ of the supply line between the preheater and the stack. Physically, all v_i with the indices $i \in \{\mathrm{CG,d};\ \mathrm{CG};\ \mathrm{CG,in};\ \mathrm{AG,d};\ \mathrm{AG};\ \mathrm{AG,in}\}$ are products consisting of a mass flow and a temperature because these two values have to be specified at the considered test rig for an experiment. Note that all mass flows are assumed to be constant in the whole thermal system because the electrochemical reaction phase does not take place in the heating phase. Therefore, the thermal behavior of the stack results from the energy balance for the stack

$$\dot{\vartheta}_{\mathrm{FC}}(t) = \frac{1}{c_{\mathrm{FC}} m_{\mathrm{FC}}} \left(\dot{Q}_{\mathrm{HT}}(t) + \dot{Q}_{\mathrm{AG}}(t) + \dot{Q}_{\mathrm{CG}}(t) + \dot{Q}_{\mathrm{EL}}(t) + \dot{Q}_{\mathrm{R}}(t) \right) \tag{67}$$

with Ohmic losses $\dot{Q}_{\mathrm{EL}}(t) = 0$, exothermic heat production due to deactivated chemical reactions $\dot{Q}_{\mathrm{R}}(t) = 0$, the mass of the stack m_{FC}, and its heat capacity c_{FC}. The temperature of the stack is assumed to be homogeneous. It is not discretized into several volume elements each with own temperatures as it was done in [20]. In fact, here the stack is assumed as one finite volume element. Therefore, heat convection between the stack and the ambient temperature $\vartheta_{\mathrm{A}}(t)$, conduction inside the stack material, and heat radiation are lumped into

$$\dot{Q}_{\mathrm{HT}}(t) = \alpha \cdot (\vartheta_{\mathrm{FC}}(t) - \vartheta_{\mathrm{A}}(t)) \tag{68}$$

with the lumped heat transfer coefficient α. In addition, the anode and cathode gases flowing through the stack lead to the enthalpy flows

$$\dot{Q}_{\mathrm{CG}}(t) = C_{\mathrm{CG}}(\vartheta_{\mathrm{FC}}(t)) \cdot (\vartheta_{\mathrm{CG,in}} - \vartheta_{\mathrm{FC}}(t)) \tag{69}$$

$$\dot{Q}_{\mathrm{AG}}(t) = C_{\mathrm{AG}}(\vartheta_{\mathrm{FC}}(t)) \cdot (\vartheta_{\mathrm{AG,in}} - \vartheta_{\mathrm{FC}}(t)) \tag{70}$$

with $\vartheta_{\mathrm{AG,in}} = \frac{v_{\mathrm{AG,in}}}{\dot{m}_{\mathrm{AG}}}$ and $\vartheta_{\mathrm{CG,in}} = \frac{v_{\mathrm{CG,in}}}{\dot{m}_{\mathrm{CG}}}$. Moreover, the heat capacities result in

$$C_{\mathrm{AG}}(\vartheta_{\mathrm{FC}}(t)) = c_{\mathrm{AG}}(\vartheta_{\mathrm{FC}}(t)) \cdot \dot{m}_{\mathrm{AG}}(t) \tag{71}$$

$$C_{\mathrm{CG}}(\vartheta_{\mathrm{FC}}(t)) \ = \ c_{\mathrm{CG}}(\vartheta_{\mathrm{FC}}(t)) \cdot \dot{m}_{\mathrm{CG}}(t) \tag{72}$$

Table 3 Overview of the nominal system parameters for the thermal subsystem of the SOFC system

Parameter		Point value (P) or Interval (I)
Heat transfer coefficient	α	I[a]
Time constant of the CG preheater	T_{CG}	P
Time constant of the CG supply line	$T_{\mathrm{CG,in}}$	P
Time constant of the AG preheater	T_{AG}	P
Time constant of the AG supply line	$T_{\mathrm{AG,in}}$	P
Product of mass and heat capacity of the stack	$c_{\mathrm{FC}} m_{\mathrm{FC}}$	P
Polynomial coefficients for heat capacity of CG	$c_{\mathrm{CG},0}$	I[a]
	$c_{\mathrm{CG},1}$	P
	$c_{\mathrm{CG},2}$	P
Polynomial coefficients for heat capacity of AG	$c_{\mathrm{AG},0}$	I[a]
	$c_{\mathrm{AG},1}$	P
	$c_{\mathrm{AG},2}$	P

Note the number of interval parameters can also be extended to all parameters
[a] $\pm 10\,\%$ uncertainty of the corresponding nominal value

for the heating phase with the temperature-dependent second-order polynomials

$$c_{\chi}\left(\vartheta_{\mathrm{FC}}(t)\right) = \sum_{\nu=0}^{2} \gamma_{\chi,\nu} \cdot \vartheta_{\mathrm{FC}}^{\nu}(t) \tag{73}$$

for both $\chi \in \{\mathrm{AG}, \mathrm{CG}\}$.

The thermal system is then given by

$$\dot{\mathbf{z}}(t) = \begin{bmatrix} \dot{v}_{\mathrm{CG}}(t) \\ \dot{v}_{\mathrm{CG,in}}(t) \\ \dot{\vartheta}_{\mathrm{FC}}(t) \end{bmatrix} = \mathbf{F}(\mathbf{z}(t), [\mathbf{p}], v_{\mathrm{CG,d}}(t), v_{\mathrm{AG,in}}(t)) \tag{74}$$

with the state vector $\mathbf{z}(t) = [v_{\mathrm{CG}}(t), v_{\mathrm{CG,in}}(t), \vartheta_{\mathrm{FC}}(t)]^T$ and the input variable $v_{\mathrm{CG,d}}(t)$.

Table 3 gives an overview about the included system parameters which were identified experimentally in previous works [2, 10, 18, 19] and for which interval enclosures are used in the ISMC.

7.2 Application of the ISMC

In order to apply the ISMC to the SOFC model, the system has to be described in nonlinear controller canonical form so that Eq. (74) has to be transformed using Lie

derivatives as it is done in [20]. These are defined for the output

$$y(t) = h(\mathbf{z}(t)) = \vartheta_{\mathrm{FC}}(t) \tag{75}$$

according to

$$\frac{\mathrm{d}^r y(t)}{\mathrm{d}t^r} = y^{(r)}(t) = L_{\mathbf{f}}^r h(\mathbf{z}(t)) = L_{\mathbf{f}}\left(L_{\mathbf{f}}^{r-1} h(\mathbf{z}(t))\right), \quad r \in \{1, 2, \ldots, \delta\} \tag{76}$$

where $\delta = 3$ holds for the relative degree in this application (because exactly the third-order and no smaller order derivatives of the output depends on the system input).

Using the new state vector

$$\mathbf{x}(t) = \begin{bmatrix} y(t) \\ \dot{y}(t) \\ \ddot{y}(t) \end{bmatrix} = \begin{bmatrix} h(\mathbf{x}(t)) \\ L_{\mathbf{f}}^1 h(\mathbf{x}(t)) \\ L_{\mathbf{f}}^2 h(\mathbf{x}(t)) \end{bmatrix} \in \mathbb{R}^{\delta}, \tag{77}$$

the set of ordinary differential equations $\dot{\mathbf{z}}(t) = \mathbf{F}\left(\mathbf{z}(t), \mathbf{p}, v_{\mathrm{CG,d}}(t), v_{\mathrm{AG,in}}(t)\right)$ can be transformed into the following state-space representation using

$$\dot{\mathbf{x}}(t) = \left[L_{\mathbf{f}}^1 h(\mathbf{z}(t)),\ L_{\mathbf{f}}^2 h(\mathbf{z}(t)),\ L_{\mathbf{f}}^3 h(\mathbf{z}(t))\right]^T \tag{78}$$

according to

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} x_2(t) \\ x_3(t) \\ \tilde{a}(\mathbf{z}(t), \mathbf{p}) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \tilde{b}_{\mathrm{CG}}(\mathbf{z}(t), \mathbf{p}) \cdot v_{\mathrm{CG,d}}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \tilde{b}_{\mathrm{AG}}(\mathbf{z}(t), \mathbf{p}) \cdot v_{\mathrm{AG,in}}(t) \end{bmatrix} \tag{79}$$

$$= \underbrace{\begin{bmatrix} 0\ 1\ 0 \\ 0\ 0\ 1 \\ 0\ 0\ 0 \end{bmatrix}}_{\mathbf{A}} \mathbf{x}(t) + \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{b}} u(t) + \underbrace{\begin{bmatrix} 0 \\ 0 \\ g(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \tilde{b}_{\mathrm{AG}}(\mathbf{z}(t), \mathbf{p}) \cdot v_{\mathrm{AG,in}}(t) \end{bmatrix}}_{\mathbf{g}(t)} \tag{80}$$

with $\mathbf{p} \in [\mathbf{p}]$, the assumption that $\frac{\mathrm{d}v_{\mathrm{AG,in}}}{\mathrm{d}t} \approx 0$, and with the scalar output

$$y(t) = \underbrace{\begin{bmatrix} 1\ 0\ 0 \end{bmatrix}}_{\mathbf{c}^T} \mathbf{x}(t) = \vartheta_{\mathrm{FC}}(t). \tag{81}$$

Moreover, the linear control law for the transformed system is given by

$$u(t) = u_{\mathrm{FF}}(t) - \mathbf{k}^T \mathbf{x}(t) \tag{82}$$

and the overall system input containing the variable-structure part according to

$$v_{\mathrm{CG,d}}(t) = \frac{1}{\tilde{b}_{\mathrm{CG}}(\mathbf{z}(t), \mathbf{p})}(-\tilde{a}(\mathbf{z}(t), \mathbf{p})u_{\mathrm{FF}}(t) - \mathbf{k}^T \mathbf{x}(t) + \boldsymbol{\eta}^T \operatorname{sign}(\mathbf{x}(t) - \mathbf{x}_{\mathrm{d}}(t))). \tag{83}$$

Note that the Lie derivatives are evaluated using the C++ toolbox FADBAD++ [4] for algorithmic differentiation. Moreover, the static feedforward control $u_{\mathrm{FF}}(t) = S \cdot \vartheta_{1,1,1,\mathrm{d}}(t)$ is used and the control gain $\mathbf{k}$ results from solving the Riccati equation [27] for the nominal parameters $\mathbf{p} = \mathrm{mid}([\mathbf{p}])$.

Figure 14 shows the overall structure of the ISMC for the described scenario assuming the original system states $\mathbf{z}(t)$ can be measured without using an observer. The implementation was done in MATLAB/SIMULINK using C++ templates in s-functions as well as the toolbox C- XSC for interval arithmetic. Operator overloading techniques are employed to compute the required derivatives of the output. As it can be seen, the Lie derivatives are calculated twice: once with and once

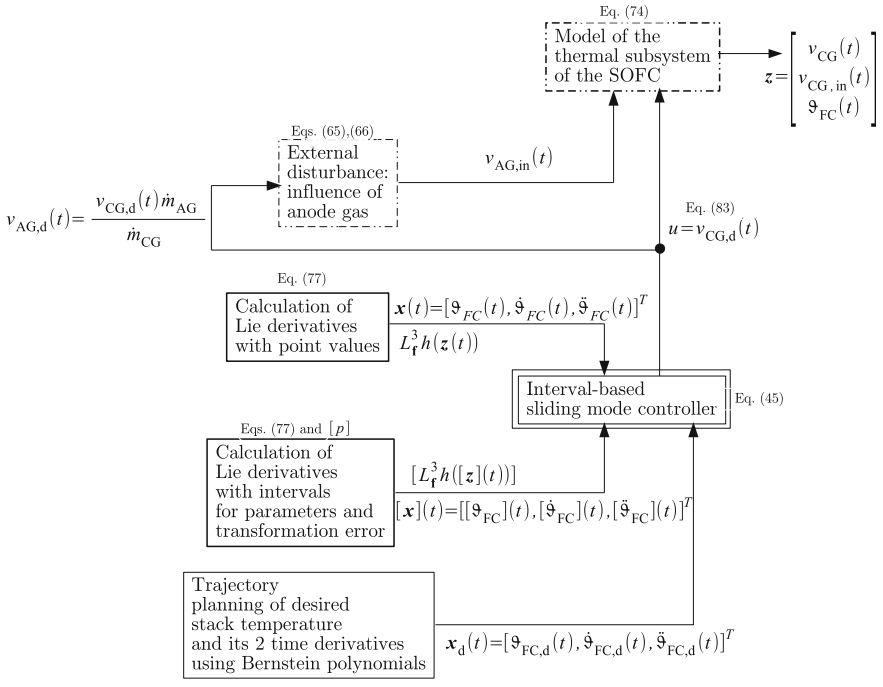


Fig. 14 Structure diagram of the control procedure for the thermal subsystem of the SOFC system

without intervals. The difference compared to the first application scenario is that the system matrix $\mathbb{A} = \mathbf{A}$ and input vector $\mathbb{B} = \mathbf{b}$ do not include any interval parameters and therefore are only point-valued terms. Therefore, it is assumed that uncertainty which is not expressed by parameter intervals is unknown from a structural point of view and consequently summarized in the vector $\mathbf{g}(t)$. It represents for example the following model assumptions that

- heat conduction in the interior of the separate fuel cells is not taken explicitely into consideration,
- the anode gas serves as an external disturbance,
- heat radiation is not modeled,
- the transformation using Lie derivatives is affected by rounding errors, and
- the approach is quasi-continuous, where time discretization errors are neglected.

These effects have been neglected in the model to find a compromise between the real model and the mathematical model for the control approach with an appropriate trade-off between model accuracy and computational complexity. Therefore, the unknown term $\mathbf{g}(t)$ is approximated by the previously mentioned process noise with the interval-valued standard deviation

$$[\mathbf{G}_p] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -[\tilde{a}(\mathbf{z}(t), [\mathbf{p}])] + \mathrm{mid}([\tilde{a}(\mathbf{z}(t), [\mathbf{p}])]) \end{bmatrix}. \tag{84}$$

Then, the control law is calculated according to Eq. (29) using Eq. (45) by taking into account a control error for the states in transformed coordinates $\mathbf{x}(t)$ as it is also done for the first application scenario. Note that the pseudoinverse in Eq. (45) is calculated according to the less conservative possibility described in Sect. 5.2.

7.3 Numerical Results

In Figs. 15 and 16, the simulation results are depicted. It can be seen that the influence of the anode gas is one order of magnitude smaller than the cathode gas which justifies the assumption that the stack temperature can be controlled by the cathode gas in the heating phase. Moreover, the resulting trajectory of the controlled stack temperature follows its desired one with good accuracy which can be seen due to the maximum control error of $\pm 0.2\,\mathrm{K}$ in Fig. 16. The trajectory planning for the stack temperature was done with the help of Bernstein polynomials in order to ensure smooth time derivatives $\dot{\vartheta}_{\mathrm{FC,d}}$ and $\ddot{\vartheta}_{\mathrm{FC,d}}$ [21].

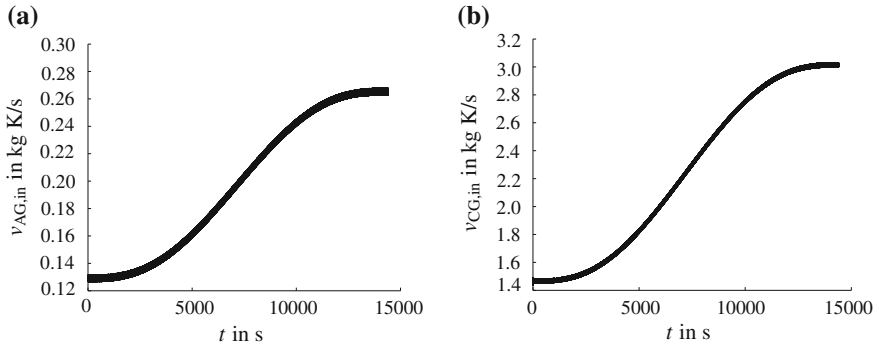


Fig. 15 Simulation results provided by the interval-based sliding mode controller for the supply lines of the gases which serve as input into the SOFC stack. **a** Anode gas supply line. **b** Cathode gas supply line

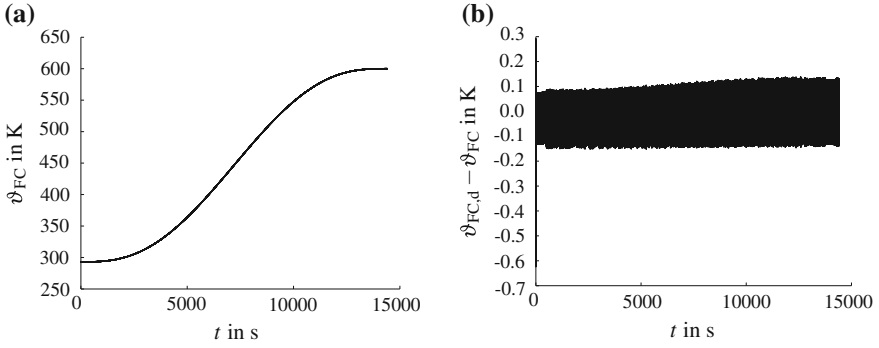


Fig. 16 Simulation results provided by the interval-based sliding mode controller for the SOFC stack. **a** Stack temperature. **b** Tracking error (stack)

8 Conclusions and Outlook

In this chapter, an interval-based sliding mode controller was described for two different application scenarios. The advantage compared with a common approach is the adaptive calculation of the variable-structure gain which reduces chattering and unnecessarily large gains that may violate actuator constraints. Moreover, intervals and stochastic influences are included to make the controller robust against uncertainty.

In future work, the ISMC for the SOFC system will be applied in real time on the test rig at the University of Rostock. Moreover, it will be extended to MIMO (multi-input multi-output) systems in general. Another point is to couple the ISMC with an online state and parameter estimation approach from previous work in terms of a closed-loop control to include a more detailed knowledge about system parameters. Furthermore, a discrete Lyapunov function candidate will be considered using

specified bounding boxes to quantify the time discretization error. Therefore, the state equations will be solved in a verified way with the help of a Picard iteration or using verified solvers for initial value problems [15].

Appendix

A Derivation of the Itô Differential Operator

Consider the stochastic differential equation

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + \mathbf{G}_p(\mathbf{x}, t)\mathrm{d}\mathbf{w}, \tag{85}$$

where $\mathbf{f}(\mathbf{x}, t)\mathrm{d}t$ can be understood as the deterministic part and $\mathbf{G}_p(\mathbf{x}, t)\mathrm{d}\mathbf{w}$ as the stochastic part. The latter one is described by a Wiener process $\mathrm{d}\mathbf{w}$ from which it is assumed that its finite increments are proportional to $\sqrt{\mathrm{d}t}$ [1]. Taking into account the aim to guarantee stability of such a stochastic differential equation, a suitable candidate of a Lyapunov function is used. In the case of control procedures, it depends on the tracking error $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{x}_\mathrm{d}$. In the following, time arguments in states are neglected. The Lyapunov function candidate is chosen as $V = \frac{1}{2}\tilde{\mathbf{x}}^T \cdot \mathbf{P} \cdot \tilde{\mathbf{x}}$. Its time derivative can be calculated for finite time increments Δt using the Taylor series expansion of order 2 according to

$$\Delta V = \frac{\partial V}{\partial t} \cdot \Delta t + \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T \cdot \Delta\tilde{\mathbf{x}} + \frac{1}{2}\Delta\tilde{\mathbf{x}}^T \cdot \left(\frac{\partial^2 V}{\partial \tilde{\mathbf{x}}^2}\right) \cdot \Delta\tilde{\mathbf{x}} + h.o.t. \tag{86}$$

Here, higher order terms ($h.o.t$) can be neglected because their influence tends to zero if the finite time increment Δt is replaced by an infinitesimally short time span $\mathrm{d}t$ in the limit $\Delta t \to 0$.

The incremental differentiation of the tracking error leads to

$$\Delta\tilde{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t)\Delta t + \mathbf{G}_p(\mathbf{x}, t)\Delta\mathbf{w} - \dot{\mathbf{x}}_d \Delta t \tag{87}$$

with the time derivative $\dot{\mathbf{x}}_d$ of the desired trajectories.

Inserting (87) into (86) yields

$$\begin{aligned}\Delta V = {} & \frac{\partial V}{\partial t} \cdot \Delta t + \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T \cdot (\mathbf{f}(\mathbf{x}, t)\Delta t + \mathbf{G}_p(\mathbf{x}, t)\Delta\mathbf{w} - \dot{\mathbf{x}}_\mathrm{d}\Delta t) \\ & + \frac{1}{2}(\mathbf{f}(\mathbf{x}, t)\Delta t + \mathbf{G}_p(\mathbf{x}, t)\Delta\mathbf{w} - \dot{\mathbf{x}}_\mathrm{d}\Delta t)^T \\ & \cdot \left(\frac{\partial^2 V}{\partial \tilde{\mathbf{x}}^2}\right) \cdot (\mathbf{f}(\mathbf{x}, t)\Delta t + \mathbf{G}_p(\mathbf{x}, t)\Delta\mathbf{w} - \dot{\mathbf{x}}_\mathrm{d}\Delta t)\end{aligned} \tag{88}$$

which can be summarized according to

$$\begin{aligned}
\Delta V &= \frac{\partial V}{\partial t}\cdot\Delta t + \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T\cdot(\mathbf{f}(\mathbf{x},t)\Delta t + \mathbf{G}_p(\mathbf{x},t)\Delta\mathbf{w} - \dot{\mathbf{x}}_\mathrm{d}\Delta t)\\
&\quad + \frac{1}{2}\left(\mathbf{f}^T\Delta t\left(\frac{\partial^2 V}{\partial\tilde{\mathbf{x}}^2}\right)\mathbf{f}\Delta t + (\mathbf{G}_p(\mathbf{x},t)\Delta\mathbf{w})^T\left(\frac{\partial^2 V}{\partial\tilde{\mathbf{x}}^2}\right)(\mathbf{G}_p(\mathbf{x},t)\Delta\mathbf{w})\right)\\
&\quad + \left(\mathbf{f}^T\Delta t\left(\frac{\partial^2 V}{\partial\tilde{\mathbf{x}}^2}\right)(\mathbf{G}_p(\mathbf{x},t)\Delta\mathbf{w})\right)\\
&= \frac{\partial V}{\partial t}\cdot\Delta t + \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T\mathbf{f}(\mathbf{x},t)\Delta t - \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T\dot{\mathbf{x}}_\mathrm{d}\Delta t\\
&\quad + \left(\mathbf{f}^T\Delta t\left(\frac{\partial^2 V}{\partial\tilde{\mathbf{x}}^2}\right)(\mathbf{G}_p(\mathbf{x},t)\Delta\mathbf{w})\right) + \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T\mathbf{G}_p(\mathbf{x},t)\Delta\mathbf{w}\\
&\quad + \frac{1}{2}\left(\mathbf{f}^T\Delta t\left(\frac{\partial^2 V}{\partial\tilde{\mathbf{x}}^2}\right)\mathbf{f}\Delta t + (\mathbf{G}_p(\mathbf{x},t)\Delta\mathbf{w})^T\left(\frac{\partial^2 V}{\partial\tilde{\mathbf{x}}^2}\right)(\mathbf{G}_p(\mathbf{x},t)\Delta\mathbf{w})\right) \qquad (89)
\end{aligned}$$

Due to the fact that the noise process is a zero mean process, its expectation value follows to $E\{\Delta\mathbf{w}\} = \mathbf{0}$ while the expectation value of the scalar product of this Brownian motion with itself yields the process covariance $E\{\Delta\mathbf{w}\cdot\Delta\mathbf{w}^T\} = \mathbf{I}\Delta t$ (identity matrix $\mathbf{I}$). Moreover, the calculation of the trace of matrices can be reformulated as $\mathrm{tr}\{\mathbf{A}\cdot\mathbf{B}\cdot\mathbf{C}\} = \mathrm{tr}\{\mathbf{C}\cdot\mathbf{A}\cdot\mathbf{B}\}$ in general. Taking into account these conditions, the expression

$$\begin{aligned}
E\{\Delta V\} = E\Bigg\{&\frac{\partial V}{\partial t}\cdot\Delta t + \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T\mathbf{f}(\mathbf{x},t)\Delta t - \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T\dot{\mathbf{x}}_\mathrm{d}\Delta t\\
&+ \left(\mathbf{f}^T\Delta t\left(\frac{\partial^2 V}{\partial\tilde{\mathbf{x}}^2}\right)(\mathbf{G}_p(\mathbf{x},t)\Delta\mathbf{w})\right) + \frac{1}{2}\left(\mathbf{f}^T\Delta t\left(\frac{\partial^2 V}{\partial\tilde{\mathbf{x}}^2}\right)\mathbf{f}\Delta t\right)\\
&+ \frac{1}{2}\mathrm{trace}\left\{\underbrace{(\Delta\mathbf{w})(\Delta\mathbf{w})^T}_{E\{(\Delta\mathbf{w})(\Delta\mathbf{w})^T\}=\mathbf{I}\Delta t}\left(\mathbf{G}_p(\mathbf{x},t)^T\left(\frac{\partial^2 V}{\partial\tilde{\mathbf{x}}^2}\right)\mathbf{G}_p(\mathbf{x},t)\right)\right\}
\end{aligned} \qquad (90)$$

and consequently

$$E\left\{\frac{\Delta V}{\Delta t}\right\} = E\left\{\frac{\partial V}{\partial t} + \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T \mathbf{f}(\mathbf{x},t) - \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T \dot{\mathbf{x}}_\mathrm{d} + \left(\mathbf{f}^T\left(\frac{\partial^2 V}{\partial \tilde{\mathbf{x}}^2}\right)\mathbf{G}_p(\mathbf{x},t)\right)\Delta\mathbf{w} + \frac{1}{2}\left(\mathbf{f}^T\left(\frac{\partial^2 V}{\partial \tilde{\mathbf{x}}^2}\right)\mathbf{f}\Delta t\right)\right\} + \frac{1}{2}\mathrm{trace}\left\{\mathbf{G}_p(\mathbf{x},t)^T\left(\frac{\partial^2 V}{\partial \tilde{\mathbf{x}}^2}\right)\mathbf{G}_p(\mathbf{x},t)\right\} \tag{91}$$

results. If $\Delta t \to 0$, then $E\left\{\frac{\Delta V}{\Delta t}\right\}$ can be replaced by $E\left\{\frac{\mathrm{d}V}{\mathrm{d}t}\right\}$. Consequently, the Itô differential operator $L(V) = \lim_{\Delta t \to 0}\left(E\left\{\frac{\Delta V}{\Delta t}\right\}\right)$ results according to

$$L(V) = \frac{\partial V}{\partial t} + \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T \mathbf{f}(\mathbf{x},t) - \left(\frac{\partial V}{\partial \tilde{\mathbf{x}}}\right)^T \dot{\mathbf{x}}_\mathrm{d} + \frac{1}{2}\mathrm{trace}\left\{\mathbf{G}_p(\mathbf{x},t)^T\left(\frac{\partial^2 V}{\partial \tilde{\mathbf{x}}^2}\right)\mathbf{G}_p(\mathbf{x},t)\right\}. \tag{92}$$

In the same way, the Itô differential operator can be derived for the dual task of state estimation using the estimation error in a suitable candidate of a Lyapunov function.

References

1. Åström KJ (1970) Introduction to stochastic control theory. Mathematics in science and engineering. Academic Press, New York
2. Auer E, Kiel S, Rauh A (2012) Verified parameter identification for solid oxide fuel cells. In: Proceedings of 5th international conference on reliable engineering computing. Brno, Czech Repulic, pp 41–55. http://rec2012.fce.vutbr.cz/documents/proceedings/REC2012_proceedings.pdf
3. Bartoszewicz A, Nowacka-Leverton A (2009) Time-varying sliding modes for second and third order systems, vol 382. Lecture notes in control and information sciences. Springer, Berlin
4. Bendsten C, Stauning O (2007) FADBAD++, Version 2.1. http://www.fadbad.com
5. Boyd S, El Ghaoui L, Feron E, Balakrishnan V (1994) Linear matrix inequalities in system and control theory. SIAM, Philadelphia
6. Brandtstädter H (2008) Sliding mode control of electromechanical systems. Dissertation, Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München
7. Datta BN (2003) Numerical methods for linear control systems. Elsevier, Amsterdam
8. Heeks J (2001) Charakterisierung unsicherer Systeme mit intervallarithmetischen Methoden. Fortschritt-Berichte VDI, Reihe 8 Mess-, Steuerungs- und Regelungstechnik, Nr. 919
9. Jaulin L, Kieffer M, Didrit O, Walter É (2001) Applied interval analysis. Springer, London
10. Kiel S, Auer E, Rauh A (2013) Uses of GPU powered interval optimization for parameter identification in the context of SO fuel cells. In: Proceedings of 9th IFAC symposium on nonlinear control systems. Toulouse, France
11. Krämer W, XSC Languages (C-XSC, PASCAL-XSC) — scientific computing with validation, arithmetic requirements, hardware solution and language support, n.a. www.math.uni-wuppertal.de/~xsc/

12. Kushner H (1967) Stochastic stability and control. Academic Press, New York
13. Marquez HJ (2003) Nonlinear control systems, analysis and design. Wiley, Hoboken
14. Moore RE (1966) Interval arithmetic. Prentice-Hall, Englewood Cliffs
15. Nedialkov NS (2007) Interval tools for ODEs and DAEs. In: CD-proceedings of the 12th GAMM-IMACS international symposium on scientific computing, computer arithmetic, and validated numerics SCAN 2006. IEEE Computer Society, Duisburg, Germany
16. Rauh A, Aschemann H (2012) Interval-based sliding mode control and state estimation of uncertain systems. In: Proceedings of IEEE international conference on methods and models in automation and Robotics MMAR 2012. Miedzyzdroje, Poland
17. Rauh A, Senkel L (2016) Interval methods for robust sliding mode control synthesis of high-temperature fuel cells with state and input constraints. In: Rauh A, Senkel L (eds) Variable-structure approaches for analysis, simulation robust control and estimation of uncertain dynamic systems, mathematical engineering. Springer, Berlin, pp 53–85. doi: 10.1007/978-3-319-31539-3_3
18. Rauh A, Dötschel T, Aschemann H (2011) Experimental parameter identification for a control-oriented model of the thermal behavior of high-temperature fuel cells. In: CD-proceedings of IEEE international conference on methods and models in automation and robotics MMAR 2011. Miedzyzdroje, Poland
19. Rauh A, Senkel L, Aschemann H (2012) Sensitivity-based state and parameter estimation for fuel cell systems. In: Proceedings of 7th IFAC symposium on robust control design. Aalborg, Denmark
20. Rauh A, Senkel L, Kersten J, Aschemann H (2014) Reliable control of high-temperature fuel cell systems using interval-based sliding mode techniques. IMA J Math Control Inf. http://imamci.oxfordjournals.org/content/early/2014/12/30/imamci.dnu051.abstract
21. Senkel L, Rauh A, Aschemann H (2013) Optimal input design for online state and parameter estimation using interval sliding mode observers. In: Proceedings of 52nd IEEE conference on decision and control, CDC 2013. Florence, Italy
22. Senkel L, Rauh A, Aschemann H (2014) Sliding mode techniques for robust trajectory tracking as well as state and parameter estimation. Mathematics in computer science, pp 543–561. http://dx.doi.org/10.1007/s11786-014-0208-7
23. Senkel L, Rauh A, Aschemann H (2014) Robust sliding mode techniques for control and state estimation of dynamic systems with bounded and stochastic uncertainty. In: Proceedings of second international conference on vulnerability and risk analysis and management. Liverpool, UK
24. Senkel L, Rauh A, Aschemann H (2014) Sliding mode approaches considering uncertainty for reliable control and computation of confidence regions in state and parameter estimation. In: Proceedings of the 16th GAMM-IMACS international symposium on scientific computing, computer arithmetic, and validated numerics SCAN 2014. Würzburg, Germany. Lecture notes in computer science LNCS, vol 9553. Springer-Verlag, pp 77–96
25. Senkel L, Rauh A, Aschemann H (2014) Experimental validation of state and parameter estimation using sliding-mode-techniques with bounded and stochastic disturbances. In: ECMI 2014, proceedings of 18th European conference on mathematics for industry. Taormina, Italy (in print)
26. Shtessel Y, Edwards C, Fridman L, Levant A (eds) (2014) Sliding mode control and observation. Springer series. Birkhäuser, New York
27. Sontag ED (1998) Mathematical control theory – deterministic finite dimensional systems. Springer, New York

Part II
Sliding Mode State Estimation for Control Purposes

The second part of this book is concerned with sliding mode state estimation procedures. These estimators are implemented in such a way that their results are directly applicable within closed-loop control structures. In Chap. 5 by Manuel Schimmack and Paolo Mercorelli, a sliding mode control strategy is presented which makes use of a bang-bang observer for the detection of particle pollutions. As in the beforementioned contribution, Hao Sun and Harald Aschemann present a sliding mode-based closed-loop control strategy in Chap. 6. The focus of this chapter is on the practical implementation and the experimental validation of the corresponding observer-based control structure for a hydrostatic transmission system. Chapter 7, authored by Tristan Braun and Johannes Reuter, focuses on the development of a sliding mode observer that includes an iterative parameter adaption approach for the characterization of a fast-switching solenoid valve. Finally, Chap. 8 by Horst Schulte and Florian Pöschke thoroughly compares linear parameter-varying and Takagi–Sugeno model approaches for the implementation of sliding mode observers that can be employed in the frame of a reliable fault diagnosis.

A Sliding Mode Control with a Bang–Bang Observer for Detection of Particle Pollution

Manuel Schimmack and Paolo Mercorelli

Abstract This chapter presents a single-input single-output (SISO) adaptive sliding mode control combined with an adaptive bang–bang observer to improve a metal–polymer composite sensor system. The proposed techniques improve the disturbance rejection of a sensor system and thus their reliability in an industrial environment. The industrial application is based on the workplace particulate pollution of welding fumes. Breathing welding fumes is extremely detrimental to human health and exposes the lungs to great hazards, therefore an effective ventilation system is essential. Typically, sliding mode control is applied in actuator control. In this sense, the proposed application is an innovative one. It seeks to improve the performance of sensors in terms of robustness with respect to parametric uncertainties and in terms of insensibility with respect to disturbances. In particular, a sufficient condition to obtain an asymptotic robustness of the estimation of the proposed bang–bang observer is designed and substantiated. The whole control scheme is designed using the well-known Lyapunov approach. A particular sliding surface is defined to obtain the inductive voltage as a controlled output. The adaptation is performed using scalar factors of the input–output data with the assistance of an output error model. A general identification technique is obtained through scaling data. To obtain this data, recursive least squares (RLS) methods are used to estimate the parameters of a linear model using input–output scaling factors. In order to estimate the parametric values in the small-scale range, the input signal requires a high frequency and thus a high sampling rate is needed. Through this proposed technique, a broader sampling rate and input signal with low frequency can be used to identify the small-scale parameters that characterise the linear model. The results indicate that the proposed algorithm is practical and robust.

M. Schimmack (✉) · P. Mercorelli (✉)
Institute of Product and Process Innovation, Leuphana University of Lueneburg, Volgershall 1, 21339 Lueneburg, Germany
e-mail: schimmack@uni.leuphana.de

P. Mercorelli
e-mail: mercorelli@uni.leuphana.de

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*, Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_5

Nomenclature

$\mathbf{A}_0$	Nominal dynamic matrix
C_0	Nominal capacity of the system
$\hat{C}_0$	Estimated capacity of the system
$d(t)$	Voltage disturbance
$\mathbf{e}(t)$	Error vector
f_m	Maximal available frequency
f_M	Maximal value of the bandwidth
$\mathbf{G}$	Observer matrix
h	Exponential scaling factor
$\mathbf{H}$	Output observer matrix
H_u	Scaling factor of the input signal
H_y	Scaling factor of the output signal
$i(t)$	Current of the system
$\hat{i}(t)$	Observed current of the system
K_s	Steady-state factor
L_0	Nominal inductance of the system
$\hat{L}_0$	Estimated inductance of the system
$\mathbf{L}_\mathrm{s}(k)$	Discrete gain matrix
$\mathbf{P}_\mathrm{s}(k)$	Discrete gain matrix
R_0	Nominal resistance of the system
$\hat{R}_0$	Estimated resistance of the system
$s(t)$	Sliding surface
t_s	Sampling rate
$t_{\mathrm{s_m}}$	Scaled sampling rate
T	Calculate factor
$u_\mathrm{s}(k)$	Discrete scaled input voltage of the model
$u_C(t)$	Capacitive voltage
$u_\mathrm{in}(t)$	Input voltage
$u_L(t)$	Inductance voltage
$\hat{u}_L(t)$	Observed inductance voltage
$\hat{u}_{L_\mathrm{max}}$	Maximal output voltage of the system
$u_\mathrm{out}(t)$	Output voltage of the system
$x_e(t)$	Magnetic flux error
$\hat{x}_2(t)$	Observed current
$x_{2\mathrm{d}}(t)$	Desired current
$y_\mathrm{s}(k)$	Discrete scaled current of the model
λ_f	Forgetting factor
$\theta_\mathrm{s}(k)$	Discrete parameter vector of scaled system
$\theta_{u_\mathrm{s}}(k)$	Discrete parameter vector of scaled input signal
$\theta_{y_\mathrm{s}}(k)$	Discrete parameter vector of scaled output signal

1 Introduction

Sliding mode control is one of the most popular control approaches. Because of its versatile ability to be applied in disturbance rejection, along with its ability to handle uncertainty and its robust stability, sliding mode control is among the most suitable control strategies in industrial applications.

In general, the theoretical literature in the field of sliding mode control points to possible applications in mechanical systems and actuators for trajectory tracking (e.g. [1]). More recently, there has been a notable interest in the applications of sliding mode control for actuators. For example, in [2–4], position controls using a sliding mode technique are proposed for various different actuator structures. The robustness of this approach against parametric uncertainty is demonstrated. In more recent publications, intelligent control designs have been proposed for electromagnetic systems; for example the development of a robust adaptive sliding mode controller in [5] and the proposal of a cascade control system, which could be used in a maglev train [6]. In [7], the authors proposed an optimal design with multiple objectives and energy compensation controls to achieve a soft landing. The landing velocity can be greatly reduced by adjusting the duty cycle of the landing current. It is worth noting that the trend in controlling electromagnetic actuators with fast dynamics is to avoid the systems using switching modalities, which can be attributed to two main factors. The first is electromagnetic compatibility. In fact, switching signals can generate dangerous interference. The second is that, if there is a high inductance in the electrical circuit in which the switching signals are involved, it is known to be difficult to switch the current quickly. This kind of problem has been considered in recent literature. Particularly in sliding mode control, the phenomena associated with a high switching frequency are referred to as chattering. Chattering properties of various control approaches must be considered. These phenomena are classified into three types. The first chattering phenomenon is classified as infinitesimal, which is harmless and cannot be avoided. However, the second one is classified as bounded and the third as unbounded. Both of them are dangerous, but it was demonstrated in [8] that they can be eliminated by proper use of high-order sliding modes (HOSM).

In general, the variable of the sliding controller should be switched at a high frequency to achieve a soft landing in actuator positioning problems. An application concerning actuators is presented in [9, 10], where the author implemented a switching extended Kalman filter in order to observe switching dynamics. Sliding mode control is one possible technique for the guaranteed stabilisation of the system by means of feedback control. This design approach makes use of a suitable Lyapunov function with which parametric uncertainties can also be handled reliably during the parameterisation of the control law. A classical design of sliding mode control laws needs the adaption of the model's estimation parameters in order to influence uncertain variables, which commonly leads to large amplitudes in the switching part of the sliding mode control law [11]. Unfortunately, this results in

undesirable noise. In [11], an arithmetic interval extension of sliding mode control is presented. This allows for a reduction in both chattering phenomena and control amplitudes, but still stabilises the system dynamics in a demonstrable way. A huge number of dynamic systems in control engineering are characterised by nonlinear behaviour, which often include both state variables, that are not directly measurable, and unknown or uncertain parameters. In [12], a novel interval sliding mode observer providing point-valued estimates is described in this article, where an optimal input design is explored in order to find an efficient operation. In recent literature, as in [13], applications of sliding mode control are utilised to achieve robustness in the sensor system against faults and disturbances. In [13], a new scheme for detection and isolation of incipient sensor faults for a class of uncertain systems is proposed by combining sliding mode observers with a Luenberger observer.

In a similar way to [14], two sliding mode observers are designed to estimate actuator and sensor faults. An analogous idea is used in this contribution to make a sensor insensitive to disturbances. Therefore, a sinusoidal voltage reference signal with a resonance frequency of the sensor is given in order to be tracked independently of disturbances. This permits the procurement of an axis intersect, which should be insensitive to disturbances. In a very recent contribution [15], the author addresses the problem of sensor fault diagnosis. The proposed solution is based on a non-homogeneous, high-order, sliding mode observer used to estimate the faults, theoretically in finite time and in the presence of defined disturbances.

In investigations related to that study, it became clear that an important prerequisite for obtaining good control performance would be increased robustness against noise and uncertainties in general. All these applications are related to the actuator control or sensor fault detection and monitoring. In this context, the proposed application is innovative. By applying a sliding mode control, it attempts to improve the performance of sensors in terms of robustness, with respect to parametric uncertainties and in terms of insensitivity to disturbances.

For an adaptive sliding mode control strategy to be obtained so that a robust control scheme is guaranteed, an identification procedure using a scaled sampling rate for the identification of the model parameters is used in [16].

This contribution intends to apply this control technique in order to improve its versatility in harsh industrial environments. A metal–polymer composite sensor is developed for the estimation of particle pollution in the study of welding fumes. Breathing welding fumes is extremely harmful and exposes the lungs to great hazards. The welding fumes consist of a mixture of dust, gases and water molecules, which, due to their small size, enter the respiratory system with ease. During the welding of stainless steel, among others, chromium and nickel are released, which are harmful to humans.

2 Measuring System and Model of the Metal–Polymer Composite Sensor

The sensory system is modelled as consisting of two coils, in which the magnetic flux from the primary to the secondary coil is amplified by the ferromagnetic particles in the welding fumes, which is depicted in Fig. 1. The convolutions of the secondary coil are based on a polymer–metal composite fibre. The manufactured fibre is composed of multiple filaments. A single filament of the conductive fibre consists of an organic polyamide kernel, which is surrounded by a thin, metallic cover.

Let us consider that the induced terminal voltage of the sensor can be described as

$$u_L(t) = -N\frac{\mathrm{d}\Phi(t)}{\mathrm{d}t} = -NA\frac{\mathrm{d}B(t)}{\mathrm{d}t}, \tag{1}$$

with number of windings N and inner surface A, which is influenced by the magnetic flux $\Phi(t)$. With the integration of (1), the relationship between the magnetic flux densities $B(t)$ at $B(0) = 0$, so the induced terminal voltage $u_L(t)$ can be formulated, thus

$$B(t) = -\frac{1}{N\,A}\int_0^t u_L(\tau)\,\mathrm{d}\tau. \tag{2}$$

In Fig. 1, the general schematic of the magnetometer circuit is depicted. The output voltage is achieved by an inverter in the following manner:

$$u_1(t) = -R_\mathrm{N}\,\frac{u_L(t)}{R_1}. \tag{3}$$

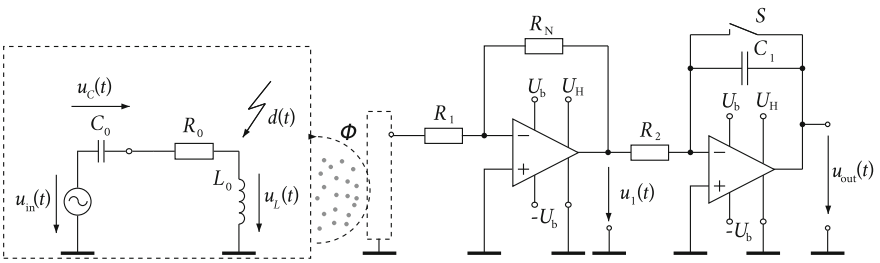


Fig. 1 Schematic of the metal–polymer composite sensor

Considering the transfer function of the subsequent integrator, which is combined with (3), the expression

$$u_{\text{out}}(t) = -\frac{1}{C_1\ R_2}\int_0^t -R_{\text{N}}\ \frac{u_L(\tau)}{R_1}\ \mathrm{d}\tau, \tag{4}$$

with $u_{\text{out}}(0) = 0$ is obtained. According to (2) and (4), it follows that

$$u_{\text{out}}(t) = -N\ A\ \frac{R_{\text{N}}}{R_1\ R_2\ C_1}\ B(t). \tag{5}$$

Thus, a linear relationship between the magnetic flux $\Phi(t)$ of the sensory element of the secondary side and the output voltage $u_{\text{out}}(t)$ of the operational amplifier is evident. Finally, let us take a look at magnetic coupling and the maximum output voltage of the sensor of the secondary side. Assuming that the sensor head is detected by a sinusoidal magnetic flux of

$$\Phi(t) = \Phi_{\max}\ \sin(\omega\, t), \tag{6}$$

which is included in (1), it follows after differentiation that the maximum output voltage is

$$\hat{u}_{L_{\max}} = N\ \Phi_{\max}\ \omega. \tag{7}$$

The sensor can be modelled using a linear system of the second order, which represents a standard RLC circuit as follows:

$$\begin{bmatrix} \frac{\mathrm{d}u_C(t)}{\mathrm{d}t} \\ \frac{\mathrm{d}i(t)}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{C_0} \\ -\frac{1}{L_0} & -\frac{R_0}{L_0} \end{bmatrix} \begin{bmatrix} u_C(t) \\ i(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{L_0} \end{bmatrix} u_{\text{in}}(t) + \begin{bmatrix} 0 \\ \frac{1}{L_0} \end{bmatrix} d(t), \tag{8}$$

in which the informative state is represented by the voltage $u_C(t)$ and the controlled state is the inductance voltage $u_L(t)$. This leads to the equations

$$u_C(t) = \begin{bmatrix} 1\ 0 \end{bmatrix} \begin{bmatrix} u_C(t) \\ i(t) \end{bmatrix} \tag{9}$$

and

$$u_L(t) = \begin{bmatrix} -1\ -R_0 \end{bmatrix} \begin{bmatrix} u_C(t) \\ i(t) \end{bmatrix} + u_{\text{in}}(t) + d(t), \tag{10}$$

where $d(t)$ represents an external voltage disturbance, which is a very common sensory problem.

3 Design of a Sliding Mode Controller

Considering the following controlled magnetic flux error as

$$x_e(t) = \Phi_{2\mathrm{d}}(t) - \Phi(t), \tag{11}$$

if the inductance is considered a constant, this means that the sensor works in the linear region of its hysteresis characteristic. In this case, the equation

$$x_e(t) = \hat{L}_0\Big(x_{2\mathrm{d}}(t) - \hat{x}_2(t)\Big) \tag{12}$$

suggests that where $\hat{L}_0$ represents the estimated inductance, $x_{2\mathrm{d}}(t)$ the desired current and $\hat{x}_2(t)$ indicates the observed current. Even though this analysis being the one of the control schemes without including the observer, the observed current is nevertheless considered. This is done for sake of notation to be used in Sect. 6 in which the analysis of the closed-loop control scheme includes also the observer. Figure 2 shows the general block diagram of the system with a scaling process of the input and output signal of the sensory system, which is described in Sect. 5.
If the following sliding surface is defined as

$$s(t) = x_e(t) + k_e \int_0^t x_e(\tau)\mathrm{d}\tau, \tag{13}$$

where $k_e \in \mathbb{R}$ with $k_e > 0$ and $u_{\mathrm{Ld}}(t)$ and $\hat{u}_L(t)$ represent the desired and the observed inductance voltages, then, by combining (14) and (12),

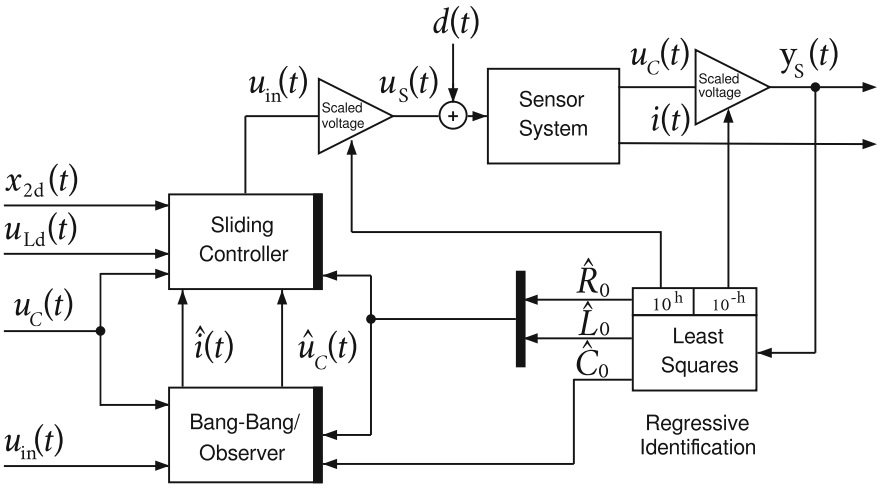


Fig. 2 General block diagram of the system and its components

$$\dot{s}(t) = \dot{x}_e(t) + k_e x_e = \hat{L}_0 \dot{x}_{2\mathrm{d}}(t) - \hat{L}_0 \dot{\hat{x}}_2(t) + k_e \hat{L}_0 \Big(x_{2\mathrm{d}}(t) - \hat{x}_2(t) \Big) \tag{14}$$

can be used. Equation (14) with $\dot{s}(t) = 0$ states the dynamics for the tracking of the controlled system in terms of inductance voltage and k_e states the rate of convergence. By solving $\dot{s}(t) = 0$ formally for the control input, an expression for the input known as *equivalent control* can be obtained, which in turn can be interpreted as the continuous control law that would maintain $\dot{s}(t) = 0$ if the exact dynamics were known. The case $s(t) = 0$ means that as the system state approaches the sliding surface, recalling that $x_e(t)$ represents the magnetic flux error, one obtains the following:

$$x_e(t) + k_e \hat{L}_0 \int_0^t \Big(x_{2\mathrm{d}}(\tau) - \hat{x}_2(\tau) \Big) \mathrm{d}\tau = 0. \tag{15}$$

Because of the derivation of (15), it follows that

$$\dot{x}_e(t) + k_e x_e(t) = 0, \tag{16}$$

which permits the solution

$$x_e(t) = \eta \exp^{-k_e t}. \tag{17}$$

The constant η is determined by the initial conditions. In order to create a region of attraction around this sliding surface, a Lyapunov function can be considered. The system model described in (8) turns out to be both controllable and observable. Consequently, a Lyapunov function can be chosen to design the sliding mode controller, as is typical, to which the following equation pertains:

$$V_{\mathrm{sl}}(t) = \frac{1}{2} s^2(t). \tag{18}$$

The Lyapunov function defined in (18) considers the error in terms of magnetic flux, which is essentially the error in terms of the current and does not consider the error of the capacitor voltage. This is due to the fact that the matter of interest is the control of the current. In this case the error of the current approaches zero, according to (12), (17) and from (15), thanks to the *equivalent control* of (21), which consists of a limited output, the inductance voltage error approaches zero. Therefore, it is guaranteed that the capacitor voltage remains limited. After differentiation of (18), it follows that

$$\dot{V}_{\mathrm{sl}}(t) = s(t)\dot{s}(t) = s(t) \Big(\hat{L}_0 \dot{x}_{2\mathrm{d}}(t) - \hat{L}_0 \dot{\hat{x}}_2(t) + k_e \hat{L}_0 \big(x_{2\mathrm{d}}(t) - \hat{x}_2(t) \big) \Big). \tag{19}$$

Furthermore, using the second equation of the system model described in (8), it follows that

$$\dot{V}_{\mathrm{sl}}(t) = s(t)\Big(u_{\mathrm{Ld}}(t) + u_C(t) + \hat{R}_0\hat{x}_2(t) - u_{\mathrm{in}}(t) - d(t) + k_e\hat{L}_0 x_{2\mathrm{d}}(t) - k_e\hat{L}_0\hat{x}_2(t)\Big). \tag{20}$$

The control loop can be designed through the use of the equation

$$u_{\mathrm{in}}(t) = u_{\mathrm{Ld}}(t) + u_C(t) + \hat{R}_0\hat{x}_2(t) + k_e\hat{L}_0 x_{2\mathrm{d}}(t) - k_e\hat{L}_0\hat{x}_2(t) + \lambda s(t) + \beta\,\mathrm{sign}\big(s(t)\big), \tag{21}$$

where $\lambda \in \mathbb{R}$, $\lambda > 0$, $\beta \in \mathbb{R}$ and $\beta > 0$. If (21) is inserted into (20), then

$$\dot{V}_{\mathrm{sl}}(t) = s(t)\dot{s}(t) = s(t)\Big(-d(t) - \lambda s(t) - \beta\,\mathrm{sign}\big(s(t)\big)\Big) < s(t)\Big(-d(t) - \beta\,\mathrm{sign}\big(s(t)\big)\Big). \tag{22}$$

Therefore, it follows that

$$\dot{V}_{\mathrm{sl}}(t) < -d(t)s(t) - \beta|s(t)|. \tag{23}$$

The sufficient condition for the global asymptotic convergence, or

$$\dot{V}_{\mathrm{sl}}(t) < 0,\ \forall t \tag{24}$$

is guaranteed, if

$$\beta > |d(t)|. \tag{25}$$

Remark 1 In the case of no correct estimation of the model's parameters being found, the following expression can be considered for the asymptotic stability

$$\beta > \max|d(t)| + \max|\Delta(R_0, L_0)|, \tag{26}$$

where $\max|\Delta(R_0, L_0)|$ is a function which represents the estimated maximal margin of uncertainty due to the variations of the parameters, which influences condition (23).

A specific method is needed in order to accomplish such a sliding mode controller. The same results are obtained if a specific version of the Lyapunov approach is used. In particular, if a forward Euler approximation is used with the sampling time t_s, the control law in (21) is transformed into

$$u_s(k-1) = u_{\mathrm{in}}(k-1) = \frac{1}{(1+k_e)}\Big((1+k_e)\big(u_{\mathrm{Ld}}(k-1) + u_C(k-1)\big) + (1 + \hat{R}_0 k_e)\hat{x}_2(k-1) + \lambda s(k-1) + \beta\,\mathrm{sign}\big(s(k-1)\big)\Big). \tag{27}$$

Therefore, it is possible to see that condition (22) becomes

$$\begin{aligned} V_{\mathrm{sl}}(k) - V_{\mathrm{sl}}(k-1) &= t_{\mathrm{s}} s(k-1)\Big(- d(k-1) - \lambda s(k-1) - \beta \operatorname{sign}\big(s(k-1)\big)\Big) \\ &< t_{\mathrm{s}} s(k-1)\Big(- d(k-1) - \beta \operatorname{sign}\big(s(k-1)\big)\Big) \end{aligned} \tag{28}$$

and also that (25) still holds true. In fact, if (23) is specified, then

$$V_{\mathrm{sl}}(k) - V_{\mathrm{sl}}(k-1) < t_{\mathrm{s}}\Big(- d(k-1)s(k-1) - \beta |s(k-1)|\Big), \quad \text{where } k \in \mathbb{N}. \tag{29}$$

It follows that the necessary condition for the discrete-time global asymptotic stability is

$$V_{\mathrm{sl}}(k) < V_{\mathrm{sl}}(k-1), \ \forall k \tag{30}$$

which is ensured by

$$\beta > \max |d(k-1)|. \tag{31}$$

4 Bang–Bang Observer to Estimate the Inductance Voltage

The presented sliding mode control strategy needs, at the very least, the availability of the inductance voltage of the scheme's primary coil reported in Fig. 1 and above. As already explained, the sensor can be modelled using a linear system of second order which represents an RLC circuit with the following uncertainty:

$$\begin{bmatrix} \frac{\mathrm{d}u_C(t)}{\mathrm{d}t} \\ \frac{\mathrm{d}i(t)}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{C_0+\delta_C} \\ -\frac{1}{L_0+\delta_L} & -\frac{R_0+\delta_R}{L_0+\delta_L} \end{bmatrix} \begin{bmatrix} u_C(t) \\ i(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{L_0+\delta_L} \end{bmatrix} u_{\mathrm{in}}(t) + \begin{bmatrix} 0 \\ \frac{1}{L_0+\delta_L} \end{bmatrix} d(t), \tag{32}$$

$$u_L(t) = \begin{bmatrix} -1 & -(R_0+\delta_R) \end{bmatrix} \begin{bmatrix} u_C(t) \\ i(t) \end{bmatrix} + u_{\mathrm{in}}(t) + d(t). \tag{33}$$

Figure 1 shows the general block diagram of the sensor system with the uncertainty of the RLC parameters. If the diagram of Fig. 3 is considered, which is also presented in [17], then the following mathematical structure can be derived

$$\begin{aligned} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} &= \begin{bmatrix} 0 & \frac{1}{C_0} \\ -\frac{1}{L_0} & -\frac{R_0}{L_0} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} q_1(t) \\ q_2(t) \\ q_3(t) \end{bmatrix} \\ &\quad + \begin{bmatrix} 0 \\ \frac{1}{L_0} \end{bmatrix} u_{\mathrm{in}}(t) + \begin{bmatrix} 0 \\ \frac{1}{L_0} \end{bmatrix} d(t). \end{aligned} \tag{34}$$

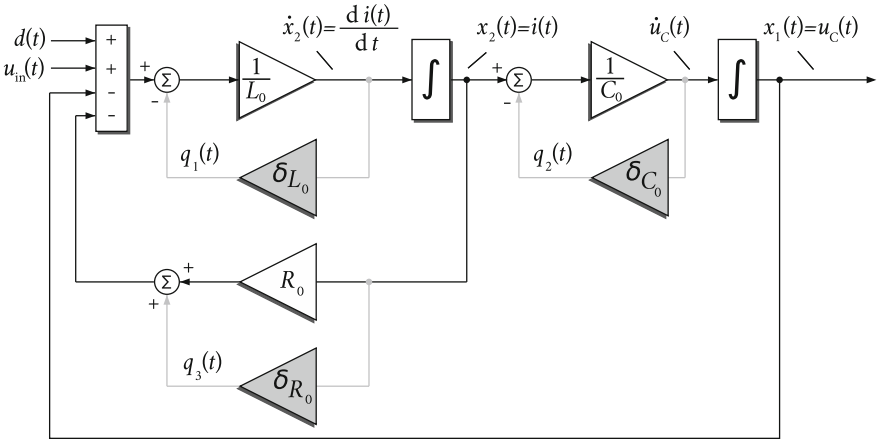


Fig. 3 Block diagram of the system with the uncertainty of the parameters

The bang–bang observer can be written as follows:

$$\begin{bmatrix}\dot{\hat{x}}_1(t)\\ \dot{\hat{x}}_2(t)\end{bmatrix} = \underbrace{\begin{bmatrix}0 & \frac{1}{C_0}\\ -\frac{1}{L_0} & -\frac{R_0}{L_0}\end{bmatrix}}_{\mathbf{A_0}}\begin{bmatrix}\hat{x}_1(t)\\ \hat{x}_2(t)\end{bmatrix} + \begin{bmatrix}0 & 0 & 0\\ -1 & -1 & -1\end{bmatrix}\begin{bmatrix}\hat{q}_1(t)\\ \hat{q}_2(t)\\ \hat{q}_3(t)\end{bmatrix} + \begin{bmatrix}0\\ \frac{1}{L_0}\end{bmatrix}u_{\text{in}}(t)$$
$$+ \underbrace{\begin{bmatrix}g_{11}\\ g_{12}\end{bmatrix}}_{\mathbf{G}}\underbrace{\begin{bmatrix}0 & 1\end{bmatrix}}_{\mathbf{H}}\begin{bmatrix}x_1(t)-\hat{x}_1(t)\\ x_2(t)-\hat{x}_2(t)\end{bmatrix} + \underbrace{\begin{bmatrix}K_1\\ K_2\end{bmatrix}}_{\mathbf{K}}\text{sign}\Big(x_2(t)-\hat{x}_2(t)\Big), \quad (35)$$

where $\mathbf{H}$ is the output observer matrix and $\mathbf{A_0}$ represents the nominal dynamic matrix. The error of the current $e_2(t)$ is obtained through the derivation of the capacity voltage error. The proposed observer consists of two parts. The first one is a Luenberger observer characterised by the matrix $\mathbf{G}$ and the second one is a switching signal, which compensates errors in both states through the matrix $\mathbf{K}$.

Theorem 1 *Let us consider the bang–bang observer defined in (35) for the system described in (34). There is a real constant $K_2^\star < +\infty$ such that, if $K_2 > K_2^\star$, then*

$$\lim_{t\to+\infty}\mathbf{e}(t) = 0, \quad \forall K_1, \text{ where } \mathbf{e}(t) = \begin{bmatrix}e_1(t)\\ e_2(t)\end{bmatrix} = \begin{bmatrix}x_1(t)-\hat{x}_1(t)\\ x_2(t)-\hat{x}_2(t)\end{bmatrix}.$$

Proof Let us define the following Lyapunov function:

$$V_{\text{ob}}(t) = \frac{e_1^2(t)+e_2^2(t)}{2}, \quad (36)$$

then it should be shown that

$$\dot{V}_{\text{ob}}(t) = e_1(t)\,\dot{e}_1(t) + e_2(t)\,\dot{e}_2(t) < 0. \tag{37}$$

The error dynamics result in the following:

$$\dot{\mathbf{e}}(t) = \mathbf{A}_0\mathbf{e}(t) + \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} q_1(t) - \hat{q}_1(t) \\ q_2(t) - \hat{q}_2(t) \\ q_3(t) - \hat{q}_3(t) \end{bmatrix} - \underbrace{\begin{bmatrix} g_{11} \\ g_{12} \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} 0 & 1 \end{bmatrix}}_{\mathbf{H}} \mathbf{e}(t)$$
$$- \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} \text{sign}\Big(x_2(t) - \hat{x}_2(t)\Big) + \begin{bmatrix} 0 \\ \frac{1}{L_0} \end{bmatrix} d(t) \tag{38}$$

and thus

$$\begin{bmatrix} \dot{e}_1(t) \\ \dot{e}_2(t) \end{bmatrix} = \Big(\mathbf{A}_0 - \mathbf{G}\,\mathbf{H}\Big) \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} q_1(t) - \hat{q}_1(t) \\ q_2(t) - \hat{q}_2(t) \\ q_3(t) - \hat{q}_3(t) \end{bmatrix}$$
$$- \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} \text{sign}\Big(x_2(t) - \hat{x}_2(t)\Big) + \begin{bmatrix} 0 \\ \frac{1}{L_0} \end{bmatrix} d(t). \tag{39}$$

Multiplying (39) from the left part by $\mathbf{e}^T(t)$ it follows that

$$\mathbf{e}^T(t) \begin{bmatrix} \dot{e}_1(t) \\ \dot{e}_2(t) \end{bmatrix} = \mathbf{e}^T(t) \Bigg(\Big(\mathbf{A}_0 - \mathbf{G}\,\mathbf{H}\Big) \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} q_1(t) - \hat{q}_1(t) \\ q_2(t) - \hat{q}_2(t) \\ q_3(t) - \hat{q}_3(t) \end{bmatrix}$$
$$- \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} \text{sign}\Big(x_2(t) - \hat{x}_2(t)\Big) + \begin{bmatrix} 0 \\ \frac{1}{L_0} \end{bmatrix} d(t) \Bigg). \tag{40}$$

According to the well-known sufficient Lyapunov condition stated in (37), it should be demonstrated that

$$\begin{bmatrix} e_1(t) & e_2(t) \end{bmatrix} \begin{bmatrix} \dot{e}_1(t) \\ \dot{e}_2(t) \end{bmatrix} = e_1(t)\,\dot{e}_1(t) + e_2(t)\,\dot{e}_2(t) < 0. \tag{41}$$

It is known that if σ_i, where $i = 1, 2$, represent the eigenvalues of $(\mathbf{A}_0 - \mathbf{G}\,\mathbf{H})$ and if the max $\big(\Re(\sigma_i)\big) < 0$ for $i = 1, 2$, then the following inequality is true

$$\mathbf{e}^T(t)\Big(\mathbf{A}_0 - \mathbf{G}\,\mathbf{H}\Big)\mathbf{e}(t) < 0. \tag{42}$$

Let us define $e_{2_{q1}}(t) = q_1(t) - \hat{q}_1(t)$, $e_{2_{q2}}(t) = q_2(t) - \hat{q}_2(t)$ and $e_{2_{q3}}(t) = q_3(t) - \hat{q}_3(t)$, then it remains to be shown that

$$e_2(t)\left(e_{2_{q1}}(t) + e_{2_{q2}}(t) + e_{2_{q3}}(t)\right) - \underbrace{e_2(t)\, K_2\, \text{sign}\left(e_2(t)\right)}_{K_2|e_2(t)|} + e_2(t)\frac{1}{L_0}d(t) < 0. \tag{43}$$

If

$$K_2 > K_2^{\star} = \max\left(\|e_{2_{q1}}(t) + e_{2_{q2}}(t) + e_{2_{q3}}(t) + \frac{1}{L_0}d(t)\|\right), \tag{44}$$

then (41) has been proven and the Lyapunov condition stated in (37) satisfied. This estimation of the norm is performed offline and the selected values based on experimental observations. □

Remark 2 A possible interpretation of such an observer structure is represented in Fig. 4, in which it is possible to see the effect of the bang–bang action. In fact, the estimation of the error due to parametric uncertainties,

$$\mathbf{e}_q(t) = \begin{bmatrix} q_1(t) - \hat{q}_1(t) \\ q_2(t) - \hat{q}_2(t) \\ q_3(t) - \hat{q}_3(t) \end{bmatrix}, \tag{45}$$

turns out to be orthogonal, in a asymptotic way, to the dynamic plan of $e_1(t)$ and $e_2(t)$ using only the condition $K_2 > K_2^{\star}$.

Remark 3 After the implementation, one is left with a discrete observer. The same results are obtained if a discrete version of the Lyapunov approach is used. This

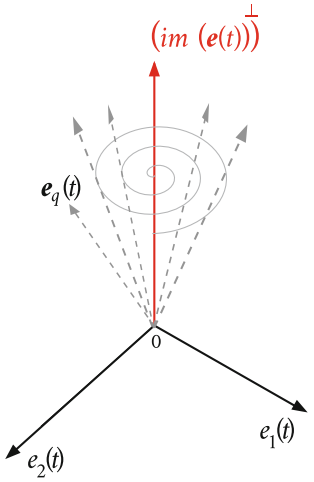


Fig. 4 Graphical interpretation of the bang–bang observer

results in the same method as presented in Sect. 3. In particular, it is possible to see that if the presented observer in (35) is specified using forward Euler with a sampling time t_s, condition (44) holds. In fact, if considering (43) and (44), it follows that

$$V_{ob}(k) - V_{ob}(k-1) < t_s\Big[e_2(k-1)\Big(e_{2_{q1}}(k-1) + e_{2_{q2}}(k-1) + e_{2_{q3}}(k-1)\Big) \\ \underbrace{- e_2(k-1)\, K_2\, \text{sign}\Big(e_2(k-1)\Big)}_{K_2|e_2(k-1)|} + e_2(k-1)\frac{1}{L_0}d(k-1)\Big] < 0, \tag{46}$$

then

$$K_2 > K_2^{\star} = \max\left(\|e_{2_{q1}}(k-1) + e_{2_{q2}}(k-1) + e_{2_{q3}}(k-1) + \frac{1}{L_0}d(k-1)\|\right). \tag{47}$$

5 Adaptive Control Through an Identification Procedure to Estimate Model Parameters

Obtaining an adaptive sliding mode control strategy in order to guarantee a robust control scheme, which is depicted in Fig. 1 and an identification procedure proposed, using a scaled sampling rate for the identification of the model parameters in (8) is necessary.

Definition 1 The scaled input and output signal of the system can be described as follows:

$$u_s(k) = H_u u(k) \qquad \text{and} \qquad y_s(k) = H_y y(k), \tag{48}$$

where H_u and H_y represent known input and output scaling factors, respectively.

Consider that the SISO system is described through the scaled process model, which is depicted in Fig. 2 with

$$A(q^{-1})H_y^{-1}y_s(k) = B(q^{-1})H_u^{-1}u_s(k). \tag{49}$$

The scaled input sequence of the system is $u_s(k)$ and the scaled output sequence of the system is $y_s(k)$. $A(z^{-1})$ and $B(z^{-1})$ are the polynomials in the unit backward shift operator and are defined by

$$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots a_{n_a} q^{-n_a}, \tag{50}$$
$$B(q^{-1}) = 1 + b_1 q^{-1} + b_2 q^{-2} + \cdots b_{n_b} q^{-n_b}. \tag{51}$$

Assuming that the degrees n_a and n_b are known and that $u_s(k) = 0$ and $y_s(k) = 0$ if $k \leq 0$, the parameter vector $\boldsymbol{\theta}(k)$ is defined as follows:

$$\boldsymbol{\theta}(k) \in \mathbb{R}^n,\ n = n_a + n_b, \tag{52}$$

where

$$\boldsymbol{\theta}(k) = \left[\, a_1\ a_2\ \ldots\ a_{n_a}\ b_1\ b_2\ \ldots\ b_{n_b} \,\right]^T. \tag{53}$$

If $\boldsymbol{\theta}(k)$ represents the parameter vector of the system, then $\boldsymbol{\theta}_{\mathrm{s}}(k)$ is the parameter vector of a scaled system such that

$$y_{\mathrm{s}}(k) = \boldsymbol{\varphi}_{\mathrm{s}}^T(k)\boldsymbol{\theta}_{\mathrm{s}}(k). \tag{54}$$

If $H_u = H_y^{-1}$, with $H_u(z^{-1}) = 10^h$ and $h \in \mathbb{N}$ being a great enough integer, then

$$\boldsymbol{\theta}_{y}(k) = \boldsymbol{\theta}_{y_{\mathrm{s}}}(k). \tag{55}$$

It follows for the scaled parameter vector that

$$\boldsymbol{\theta}_{u_{\mathrm{s}}}(k) = b_i H_u^{-2}, \quad \text{with}\ \ i = 1, 2, \ldots, n_b, \tag{56}$$

which can be simplified to

$$\boldsymbol{\theta}_{\mathrm{s}}(k) = \left[\, \boldsymbol{\theta}_{y}^T(k)\ \boldsymbol{\theta}_{u_{\mathrm{s}}}^T(k) \,\right]^T. \tag{57}$$

The filtered information vector can be written as follows:

$$\boldsymbol{\varphi}_{\mathrm{s}}(k) = \begin{bmatrix} \varphi_{y_{\mathrm{s}}}(k) \\ \varphi_{u_{\mathrm{s}}}(k) \end{bmatrix} \in \mathbb{R}^n, n = n_a + n_b, \tag{58}$$

with

$$\boldsymbol{\varphi}_{y_{\mathrm{s}}}(k) = \left[-y_{\mathrm{s}}(k-1)\ -y_{\mathrm{s}}(k-2)\ \ldots\ -y_{\mathrm{s}}(k-n_a)\right]^T, \tag{59}$$

and

$$\boldsymbol{\varphi}_{u_{\mathrm{s}}}(k) = \left[u_{\mathrm{s}}(k-1)\ u_{\mathrm{s}}(k-2)\ \ldots\ u_{\mathrm{s}}(k-n_b)\right]^T. \tag{60}$$

Transforming the scaled model, which can be seen as an equation error model, in the following form results in

$$H_y^{-1} y_{\mathrm{s}}(k) = [1 - A(q^{-1})] H_y^{-1} y_{\mathrm{s}}(k) + B(q^{-1}) H_u^{-1} u_{\mathrm{s}}(k). \tag{61}$$

Considering that $H_y = H_u^{-1}$, Eq. (61) can be written as follows:

$$y_{\mathrm{s}}(k) = -\sum_{i=1}^{n_a} a_i y_{\mathrm{s}}(k-i) + H_u^{-2} \sum_{i=1}^{n_b} b_i u_{\mathrm{s}}(k-i). \tag{62}$$

According to (53) and (59), it can be concluded that

$$\boldsymbol{\varphi}_{\mathrm{s}}^{T}(k)\boldsymbol{\theta}_{y}(k) = -\sum_{i=1}^{n_a} a_i y_{\mathrm{s}}(k-i), \tag{63}$$

and it can be deduced from (53) and (60) that

$$\boldsymbol{\varphi}_{\mathrm{s}}^{T}(k)\boldsymbol{\theta}_{u_{\mathrm{s}}}(k) = H_u^{-2}\sum_{i=1}^{n_b} b_i u_{\mathrm{s}}(k-i). \tag{64}$$

Having proved (55)–(57), it can be established that

$$y_{\mathrm{s}}(k) = \boldsymbol{\varphi}_{\mathrm{s}}^{T}(k)\left[\,\boldsymbol{\theta}_{y}(k)\;\boldsymbol{\theta}_{u_{\mathrm{s}}}(k)\,\right]^{T} = \boldsymbol{\varphi}_{\mathrm{s}}^{T}(k)\boldsymbol{\theta}_{\mathrm{s}}(k). \tag{65}$$

If the values of the system dynamics are constant, a classical RLS algorithm is suitable. However, if the system is given by

$$y_{\mathrm{s}}(k) = \boldsymbol{\varphi}_{\mathrm{s}}^{T}(k)\boldsymbol{\theta}_{\mathrm{s}}(k), \tag{66}$$

where $\boldsymbol{\theta}_{\mathrm{s}}(k)$ describes a time-varying parameter vector, a classical RLS method is unsuitable. One way to obtain an estimation algorithm, which tracks parametric changes, is to change the least squares criterion to

$$J = \sum_{k=1}^{N} \lambda_{\mathrm{f}}^{N-k}(y_{\mathrm{s}}(k) - \hat{y}_{\mathrm{s}}(k))^2. \tag{67}$$

In this criterion, a so-called forgetting factor λ_{f} ($0 < \lambda_{\mathrm{f}} \leq 1$) is introduced. A forgetting factor less than one emphasises old prediction errors less than new ones (for $\lambda_{\mathrm{f}} = 1$ the standard least squares criterion is retrieved). The smaller the forgetting factor is, the faster old prediction errors will be "forgotten". The modified RLS based on the model (66) and on criterion (67) is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(k) = \hat{\boldsymbol{\theta}}_{\mathrm{s}}(k-1) + \mathbf{L}_{\mathrm{s}}(k)\Big(y_{\mathrm{s}}(k) - \boldsymbol{\varphi}_{\mathrm{s}}^{T}(k)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(k-1)\Big), \tag{68}$$

where

$$\mathbf{L}_{\mathrm{s}}(k) = \frac{\mathbf{P}_{\mathrm{s}}(k-1)\boldsymbol{\varphi}_{\mathrm{s}}^{T}(k)}{\lambda_{\mathrm{f}} + \boldsymbol{\varphi}_{\mathrm{s}}^{T}(k)\mathbf{P}_{\mathrm{s}}(k-1)\boldsymbol{\varphi}_{\mathrm{s}}(k)} \tag{69}$$

and

$$\mathbf{P}_{\mathrm{s}}(k) = \frac{1}{\lambda_{\mathrm{f}}}\Big([\mathbf{I} - \mathbf{L}_{\mathrm{s}}(k)\boldsymbol{\varphi}_{\mathrm{s}}^{T}(k)]\mathbf{P}_{\mathrm{s}}(k-1)\Big). \tag{70}$$

For $\lambda_f < 1$, the gain matrix $\mathbf{P}_s(k)$ will not reach zero as time continues to infinity, thereby allowing the algorithm to update the parameters when the system is time-varying. A typical choice for $\lambda_f < 1$ is in the range of 0.94–0.999. More details on this matter can be found in [18–20].

It follows the implementation of the scaled least square method. A system of the second order is considered as follows:

$$\frac{Y(s)}{U(s)} = \frac{K_s}{1 + 2\frac{\xi_1}{\omega_{n_1}}s + \frac{s^2}{\omega_{n_1}^2}}. \tag{71}$$

Let frequency f_M be the maximal bandwidth value of a system described in (71), and frequency f_m be a possible scaled frequency, then consider

$$H_u = \frac{f_M}{f_m} \tag{72}$$

and

$$H_y^{-1} = \frac{f_m}{f_M}. \tag{73}$$

A scaled sampling rate is considered equal to

$$t_{s_m} = \frac{1}{2pf_m}, \tag{74}$$

where the parameter p is typically between 1 and 10, and factor 2 guarantees an anti-aliasing effect (Nyquist–Shannon Theorem). If (71) is scaled in the following way:

$$U_s(s) = \frac{f_M}{f_m}U(s) \quad \text{and} \quad Y_s(s) = \frac{f_m}{f_M}Y(s), \tag{75}$$

the following expression is obtained

$$\frac{Y_s(s)}{U_s(s)} = \frac{\frac{f_m}{f_M}K_s}{\frac{f_M}{f_m}\left(1 + 2\frac{\xi_1}{\omega_{n_1}}s + \frac{s^2}{\omega_{n_1}^2}\right)}. \tag{76}$$

If we consider the feedforward Euler approximation using a sample rate t_s, it follows that

$$y_s(t) = y_s(k-1), \quad u_s(t) = u_s(k-1), \qquad k = 1, 2, \ldots, q, \tag{77}$$

with $q \in \mathbb{N}$ and structuring the derivation based on Newton's notation of differentials for the output signal

$$\dot{y}_s(t) \approx \frac{y_s(k) - y_s(k-1)}{t_s} \tag{78}$$

and

$$\ddot{y}_{\mathrm{s}}(t) \approx \frac{\frac{y_{\mathrm{s}}(k)-y_{\mathrm{s}}(k-1)}{t_{\mathrm{s}}} - \frac{y_{\mathrm{s}}(k-1)-y_{\mathrm{s}}(k-2)}{t_{\mathrm{s}}}}{t_{\mathrm{s}}}. \tag{79}$$

If the system is represented by (76) and is transformed back in time, the following expression is obtained

$$\frac{f_{\mathrm{m}}}{f_{\mathrm{M}}}\omega_{n_1}^2 K_{\mathrm{s}} u_{\mathrm{s}}(t) = \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}}\ddot{y}_{\mathrm{s}}(t) + \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}} 2\xi_1 \omega_{n_1} \dot{y}(t) + \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}}\omega_{n_1}^2 y_{\mathrm{s}}(t). \tag{80}$$

The discrete form of (80) using feedforward Euler approximation is as follows:

$$\begin{aligned}\frac{f_{\mathrm{m}}}{f_{\mathrm{M}}}\omega_{n_1}^2 K_{\mathrm{s}} u_{\mathrm{s}}(k-1) = {} & \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}}\left(\frac{y_{\mathrm{s}}(k)-y_{\mathrm{s}}(k-1)}{t_{\mathrm{s}}^2} - \frac{y_{\mathrm{s}}(k-1)-y_{\mathrm{s}}(k-2)}{t_{\mathrm{s}}^2}\right) \\ & + \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}} 2\xi_1 \omega_{n_1} \frac{y_{\mathrm{s}}(k)-y_{\mathrm{s}}(k-1)}{t_{\mathrm{s}}} + \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}}\omega_{n_1}^2 y_{\mathrm{s}}(k-1)\end{aligned} \tag{81}$$

and

$$\begin{aligned}\frac{f_{\mathrm{m}}}{f_{\mathrm{M}}}\omega_{n_1}^2 K_{\mathrm{s}} t_{\mathrm{s}}^2 u_{\mathrm{s}}(k-1) = {} & \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}} y_{\mathrm{s}}(k) - \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}} y_{\mathrm{s}}(k-1) - \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}}\Big(y_{\mathrm{s}}(k-1) - y_{\mathrm{s}}(k-2)\Big) \\ & + \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}} 2\xi_1 \omega_{n_1} t_{\mathrm{s}} \Big(y_{\mathrm{s}}(k) - y_{\mathrm{s}}(k-1)\Big) + \frac{f_{\mathrm{M}}}{f_{\mathrm{m}}} t_{\mathrm{s}}^2 \omega_{n_1}^2 y_{\mathrm{s}}(k-1)\end{aligned} \tag{82}$$

which finally leads to

$$\begin{aligned}y_{\mathrm{s}}(k) = {} & \frac{1}{1+2\xi_1\omega_{n_1}t_{\mathrm{s}}}\Big(\Big[2+2\xi_1\omega_{n_1}t_{\mathrm{s}}+\omega_{n_1}^2 t_{\mathrm{s}}^2\Big] y_{\mathrm{s}}(k-1) - y_{\mathrm{s}}(k-2) \\ & + \frac{f_{\mathrm{m}}^2}{f_{\mathrm{M}}^2} K_1 \omega_{n_1}^2 t_{\mathrm{s}}^2 u_{\mathrm{s}}(k-1)\Big).\end{aligned} \tag{83}$$

Let us isolate the estimation parameters

$$y(k) = -\boldsymbol{\theta_1} y(k-1) + \boldsymbol{\theta_2} y(k-2) + \boldsymbol{\theta_3} u(k-1). \tag{84}$$

According to (84), the parameters Θ_1, Θ_2 and Θ_3 can be identified as

$$\boldsymbol{\theta_1} = -\frac{2+2\xi_1\omega_{n_1}t_{\mathrm{s}}+\omega_{n_1}^2 t_{\mathrm{s}}^2}{1+2\xi_1\omega_{n_1}t_{\mathrm{s}}}, \tag{85}$$

$$\boldsymbol{\theta_2} = -\frac{1}{1+2\xi_1\omega_{n_1}t_{\mathrm{s}}} \tag{86}$$

and

$$\boldsymbol{\theta_3} = \frac{f_{\mathrm{m}}^2}{f_{\mathrm{M}}^2} \frac{K_{\mathrm{s}} \omega_{n_1}^2 t_{\mathrm{s}}^2}{1 + 2\xi_1 \omega_{n_1} t_{\mathrm{s}}}. \tag{87}$$

6 Tracking Dynamics of the Closed-Loop System and Discussion of the Results

The following result guarantees the convergence of the closed control loop consisting of the system described in (34), the observer defined in (35) and the controller in (21). Considering the error dynamics of the whole control structure in which the *equivalent control* as in (21) and the bang–bang observer are used, then it follows

$$\begin{bmatrix} \dot{e}_{x_1}(t) \\ \dot{e}_{x_2}(t) \\ \dot{e}_i(t) \\ \dot{e}_{\hat{x}_1}(t) \\ \dot{e}_{\hat{x}_2}(t) \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{C_0} & 0 & 0 & 0 \\ 0 & -(k_e + \lambda) & -\lambda k_e & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{C_0} - g_{11} \\ 0 & 0 & 0 & -\frac{1}{L_0} & -\frac{R_0}{L_0} - g_{21} \end{bmatrix} \begin{bmatrix} e_{x_1}(t) \\ e_{x_2}(t) \\ e_i(t) \\ e_{\hat{x}_1}(t) \\ e_{\hat{x}_2}(t) \end{bmatrix}$$
$$+ \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \Delta q_1(t) \\ \Delta q_2(t) \\ \Delta q_3(t) \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ K_1 \\ K_2 \end{bmatrix} \operatorname{sign}\big(e_{\hat{x}_2}(t)\big)$$
$$+ \begin{bmatrix} 0 \\ \frac{1}{L_0} \\ 0 \\ 0 \\ \frac{1}{L_0} \end{bmatrix} d(t) - \begin{bmatrix} 0 \\ \beta \\ 0 \\ 0 \\ 0 \end{bmatrix} \operatorname{sign}\big(s(t)\big), \tag{88}$$

where $e_{x_1}(t) = x_{1d}(t) - \hat{x}_1(t)$, $e_{x_1}(t)$ represents the voltage capacitor error, $e_{x_2}(t) = x_{2d}(t) - \hat{x}_2(t)$ represents the current capacitor error and the auxiliary variable $e_i(t)$ is the integral variable of the current error as stated in (88), to be more precise it results in $\dot{e}_i(t) = e_{x_2}(t)$. State variables $e_{\hat{x}_1}(t)$ and $e_{\hat{x}_2}(t)$ represent the observation errors as follows: $e_{\hat{x}_1}(t) = x_1(t) - \hat{x}_1(t)$, and $e_{\hat{x}_2}(t) = x_2(t) - \hat{x}_2(t)$. And it is $\Delta q_1(t) = q_1(t) - \hat{q}_1(t)$, $\Delta q_2(t) = q_2(t) - \hat{q}_2(t)$ together with $\Delta q_3(t) = q_2(t) - \hat{q}_3(t)$.
If

$$\mathbf{A_C} = \begin{bmatrix} 0 & \frac{1}{C_0} & 0 \\ 0 & -(k_e + \lambda) & -\lambda k_e \\ 0 & 1 & 0 \end{bmatrix}, \tag{89}$$

the error dynamics results to be

$$\underbrace{\begin{bmatrix} \dot{e}_{x_1}(t) \\ \dot{e}_{x_2}(t) \\ \dot{e}_i(t) \\ \dot{e}_{\hat{x}_1}(t) \\ \dot{e}_{\hat{x}_2}(t) \end{bmatrix}}_{\dot{\mathbf{e}}(t)} = \underbrace{\begin{bmatrix} \mathbf{A}_{\mathrm{C}} & \mathbf{0}_{3\times 2} \\ \mathbf{0}_{2\times 3} & (\mathbf{A}_0 - \mathbf{GH}) \end{bmatrix}}_{\mathbf{A}_{\mathrm{CO}}} \underbrace{\begin{bmatrix} e_{x_1}(t) \\ e_{x_2}(t) \\ e_i(t) \\ e_{\hat{x}_1}(t) \\ e_{\hat{x}_2}(t) \end{bmatrix}}_{\mathbf{e}(t)} + \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \Delta q_1(t) \\ \Delta q_2(t) \\ \Delta q_3(t) \end{bmatrix}$$

$$- \begin{bmatrix} 0 \\ 0 \\ 0 \\ K_1 \\ K_2 \end{bmatrix} \operatorname{sign}\left(e_{\hat{x}_2}(t)\right) + \begin{bmatrix} 0 \\ \frac{1}{L_0} \\ 0 \\ 0 \\ \frac{1}{L}0 \end{bmatrix} d(t) - \begin{bmatrix} 0 \\ \beta \\ 0 \\ 0 \\ 0 \end{bmatrix} \operatorname{sign}\left(s(t)\right). \tag{90}$$

It is known that if ρ_i, where $i = 1, 2, \ldots, 5$ represents the eigenvalues of $\mathbf{A}_{\mathrm{CO}}$ and if the max $\left(\Re(\rho_i)\right) \leq 0$ for $i = 1, 2, \ldots, 5$, then the following inequality is true:

$$\mathbf{e}^T(t)\, \mathbf{A}_{\mathrm{CO}}\, \mathbf{e}(t) \leq 0. \tag{91}$$

Through parameter k_e and λ, it is possible to place in an arbitrary way just two eigenvalues of the matrix $\mathbf{A}_{\mathrm{C}}$ defined by (89). In fact, the eigenvalues of the matrix $\mathbf{A}_{\mathrm{C}}$ are $\rho_1 = 0$, $\rho_2 = -\lambda$ and $\rho_3 = -k_e$. The eigenvalue which is not assignable using this controller corresponds to the dynamics of the voltage capacitor. The eigenvalue $\rho_2 = -\lambda$ states the velocity of $\lim_{t\to+\infty} s(t) = 0$, where $s(t)$ is defined in Sect. 3 *velocity of attraction of the sliding surface* as explained later and $\rho_3 = -k_e$ states the dynamics of the current in case $s(t) \approx 0$.

The dynamic of the observer, which is characterised by the eigenvalues of the matrix $\mathbf{A}_0 - \mathbf{GH}$, can be totally determined by placing its eigenvalues. For the analysis of the error convergence the state variable $s(t)$ defined in Sect. 3, which for reason of notation is renamed $s_{x_2}(t)$, can be suitably used. At the end of the calculation a *similar matrix* expression is obtained. A *similar matrix* is obtained through *similar transformations* and the final obtained matrix consists of the same eigenvalues as the original one. Recalling that

$$s(t) = s_{x_2}(t) = x_e(t) + k_e \int_0^t x_e(\tau)\mathrm{d}\tau, \tag{92}$$

with $e_{x_2}(t) = x_e(t) = \hat{L}_0\left(x_{2\mathrm{d}}(t) - \hat{x}_2(t)\right)$. If (92) is differentiated, then the following expression is obtained

$$\dot{s}_{x_2}(t) = L_0 \dot{e}_{x_2}(t) + k_e L_0 e_{x_2}(t). \tag{93}$$

From (88), the current error is obtained as follows:

$$\dot{e}_{x_2}(t) = -(k_e + \lambda)e_{x_2}(t) - \lambda e_i(t). \tag{94}$$

If (92)–(94) are combined, the following obvious relation is obtained:

$$\dot{s}_{x_2}(t) = -\lambda s_{x_2}(t). \tag{95}$$

In fact, the defined *equivalent control* (21) realises the sliding surface as an *attractive one* and the eigenvalue $\rho_2 = -\lambda$, as already mentioned, states the velocity of attraction of the sliding surface $s(t)$. From (92) it follows that

$$e_{x_2}(t) = \frac{s(t)}{L_0} - k_e e_i(t). \tag{96}$$

Thus, combining (92)–(94), the following relation is obtained:

$$\dot{e}_i(t) = \frac{1}{k_e L_0}(k_e + \lambda)s(t) - \lambda\frac{1}{k_e L_0}s(t) - (k_e + \lambda)e_i(t) + \lambda e_i(t), \tag{97}$$

and it follows that

$$\dot{e}_i(t) = \frac{1}{L_0}s(t) - k_e e_i(t). \tag{98}$$

Being $\dot{e}_{x_1}(t) = \frac{1}{C_0}e_{x_2}(t)$ and combining (92)–(94), with similar calculations the following relation is obtained:

$$\dot{e}_{x_1}(t) = \frac{1}{C_0 L_0}s(t) - \frac{k_e}{C_0}e_i(t). \tag{99}$$

If the following system error is considered:

$$\underbrace{\begin{bmatrix} \dot{e}_{x_1}(t) \\ \dot{s}_{x_2}(t) \\ \dot{e}_i(t) \\ \dot{e}_{\hat{x}_1}(t) \\ \dot{e}_{\hat{x}_2}(t) \end{bmatrix}}_{\dot{\mathbf{e}}(t)} = \underbrace{\begin{bmatrix} \mathbf{A}_{\mathrm{CS}} & \mathbf{0}_{3\times 2} \\ \mathbf{0}_{2\times 3} & (\mathbf{A}_0 - \mathbf{GH}) \end{bmatrix}}_{\mathbf{A}_{\mathrm{COS}}} \underbrace{\begin{bmatrix} e_{x_1}(t) \\ e_{x_2}(t) \\ e_i(t) \\ e_{\hat{x}_1}(t) \\ e_{\hat{x}_2}(t) \end{bmatrix}}_{\mathbf{e}(t)} + \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \Delta q_1(t) \\ \Delta q_2(t) \\ \Delta q_3(t) \end{bmatrix}$$

$$- \begin{bmatrix} 0 \\ 0 \\ 0 \\ K_1 \\ K_2 \end{bmatrix} \mathrm{sign}\big(e_{\hat{x}_2}(t)\big) + \begin{bmatrix} 0 \\ \frac{1}{L_0} \\ 0 \\ 0 \\ \frac{1}{L_0} \end{bmatrix} d(t) - \begin{bmatrix} 0 \\ \beta \\ 0 \\ 0 \\ 0 \end{bmatrix} \mathrm{sign}\big(s_{x_2}(t)\big), \tag{100}$$

with

$$\mathbf{A}_{\mathrm{CS}} = \begin{bmatrix} 0 & \frac{1}{C_0 L_0} & -\frac{1}{C_0} k_e \\ 0 & -\lambda & 0 \\ 0 & \frac{1}{L_0} & -k_e \end{bmatrix}. \tag{101}$$

As already mentioned, matrix (89) and (101) are *similar matrices*, so they have the same eigenvalues ($\rho_1 = 0$, $\rho_2 = -\lambda$ and $\rho_3 = -k_e$).

Theorem 2 *Let us consider the closed-loop control system which consists of the system described in (34), the bang–bang observer defined in (35) and of the controller in (21). There are two real and finite constants $\hat{\beta} > 0$ and $\hat{K}_2 > 0$ such that, if $\beta > \hat{\beta}$ and $K_2 > \hat{K}_2$, then the closed-loop control system is stable in the sense of Lyapunov and*

$$\lim_{t \to +\infty} e_{x_2}(t) = 0, \quad \forall K_1, \text{where} \quad e_{x_2}(t) = x_{2d}(t) - \hat{x}_2(t).$$

Proof Let us define the following Lyapunov function:

$$V_{\mathrm{COS}}(t) = \frac{e_{x_1}^2(t) + s_{x_2}^2(t) + e_i^2(t) + e_{\hat{x}_1}^2(t) + e_{\hat{x}_2}^2(t)}{2}, \tag{102}$$

then it should be shown that

$$\begin{aligned} \dot{V}_{\mathrm{COS}}(t) = {} & e_{x_1}(t)\, \dot{e}_{x_1}(t) + s_{x_2}(t)\, \dot{s}_{x_2}(t) + e_i(t) \dot{e}_i(t) + e_i(t) \dot{e}(t) \\ & + e_{\hat{x}_1}(t)\, \dot{e}_{\hat{x}_1}(t) + e_{\hat{x}_2}(t)\, \dot{e}_{\hat{x}_2}(t) \leq 0, \end{aligned} \tag{103}$$

and that

$$s_{x_2}(t)\, \dot{s}_{x_2}(t) < 0. \tag{104}$$

Multiplying (90) by $\mathbf{e}^T(t)$ it follows that

$$\begin{aligned} \mathbf{e}^T(t) \dot{\mathbf{e}}(t) = {} & \mathbf{e}^T(t) \mathbf{A}_{\mathrm{COS}} \mathbf{e}(t) - \mathbf{e}^T(t) \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \Delta q_1(t) \\ \Delta q_2(t) \\ \Delta q_3(t) \end{bmatrix} - \mathbf{e}^T(t) \begin{bmatrix} 0 \\ 0 \\ 0 \\ K_1 \\ K_2 \end{bmatrix} \\ & \times \operatorname{sign}\left(e_{\hat{x}_2}(t)\right) + \mathbf{e}^T(t) \begin{bmatrix} 0 \\ \frac{1}{L_0} \\ 0 \\ 0 \\ \frac{1}{L} 0 \end{bmatrix} d(t) - \mathbf{e}^T(t) \begin{bmatrix} 0 \\ \beta \\ 0 \\ 0 \\ 0 \end{bmatrix} \operatorname{sign}\left(s_{x_2}(t)\right). \end{aligned} \tag{105}$$

According to the well-known sufficient Lyapunov condition

$$\begin{aligned}\mathbf{e}^T(t)\dot{\mathbf{e}}(t) &= \begin{bmatrix}\mathbf{e}_x(t) & \mathbf{e}_i(t) & \mathbf{e}_{\hat{x}}(t)\end{bmatrix}\begin{bmatrix}\mathbf{e}_x(t)\\ \mathbf{e}_i(t)\\ \mathbf{e}_{\hat{x}}(t)\end{bmatrix}\\ &= \mathbf{e}_x(t)\dot{\mathbf{e}}_x(t) + e_i(t)\dot{e}_i(t) + \mathbf{e}_{\hat{x}}(t)\dot{\mathbf{e}}_{\hat{x}}(t) < 0,\end{aligned} \tag{106}$$

where $\mathbf{e}_x(t) = \begin{bmatrix}e_{x_1}(t)\\ s_{x_2}(t)\end{bmatrix}$ represents the dynamic error and $\mathbf{e}_{\hat{x}}(t) = \begin{bmatrix}e_{\hat{x}_1}(t)\\ e_{\hat{x}_2}(t)\end{bmatrix}$ the estimation error of the voltage and current of the capacitor. Proceeding in a similar way as for the demonstration of Sect. 4, let us define $e_{2_{q1}}(t) = q_1(t) - \hat{q}_1(t)$, $e_{2_{q2}}(t) = q_2(t) - \hat{q}_2(t)$ and $e_{2_{q3}}(t) = q_3(t) - \hat{q}_3(t)$. It remains to be shown that

$$\begin{aligned}0 > {} & s_{x_2}(t)\Big(e_{2_{q_1}}(t) + e_{2_{q_2}}(t) + e_{2_{q_3}}(t)\Big) - \underbrace{s_{x_2}(t)\ \beta\ \mathrm{sign}\Big(s_{x_2}(t)\Big)}_{\beta|e_{x_2}(t)|}\\ & + s_{x_2}(t)\frac{1}{L_0}d(t) + e_{\hat{x}_2}(t)\Big(e_{2_{q_1}}(t) + e_{2_{q_2}}(t) + e_{2_{q_3}}(t)\Big)\\ & - \underbrace{e_{\hat{x}_2}(t)\ K_2\ \mathrm{sign}\Big(e_{\hat{x}_2}(t)\Big)}_{K_2|e_{\hat{x}_2}(t)|} + e_{\hat{x}_2}(t)\frac{1}{L_0}d(t).\end{aligned} \tag{107}$$

From (107), if

$$\beta > \hat{\beta} = \max\left(\|e_{2_{q1}}(t) + e_{2_{q2}}(t) + e_{2_{q3}}(t) + \frac{d(t)}{L_0}\|\right), \tag{108}$$

and if

$$K_2 > K_2^{\star} = \max\left(\|e_{2_{q1}}(t) + e_{2_{q2}}(t) + e_{2_{q3}}(t) + \frac{1}{L_0}d(t)\|\right) + \hat{\beta}, \tag{109}$$

then (41) has been proven and the Lyapunov condition stated in (37) is satisfied. □

The estimation of the norm is performed offline and the selected values are based on experimental observations. Relations (108) and (109) state two sufficient conditions which guarantee the asymptotical convergence of the tracking of current, inductance voltage and of the estimated variables using the bang–bang observer for the closed-loop control scheme. As already explained, an eigenvalue of the error dynamics is not assignable by the controller and remains equal to zero. This eigenvalue states the dynamics of the capacitor voltage which remains limited. The structure of the *equivalent control* of (21) and the bang–bang observer of (35) remain the same also in their discrete forms already discussed.

Figure 6 shows the tracking control obtained for the inductance voltage $u_L(t)$ with $K_2 = 0$ in presence of the disturbance $d(t)$, which is depicted in Fig. 5. The

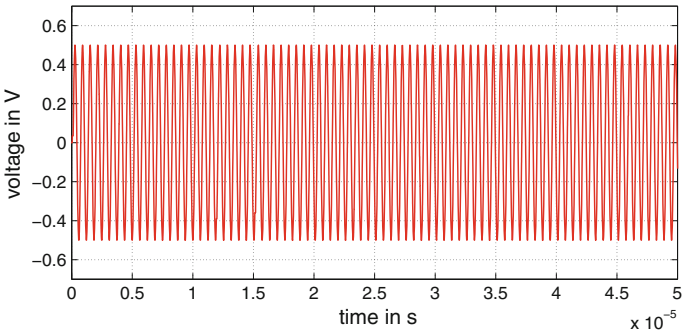


Fig. 5 Graphical representation of the disturbance $d(t)$

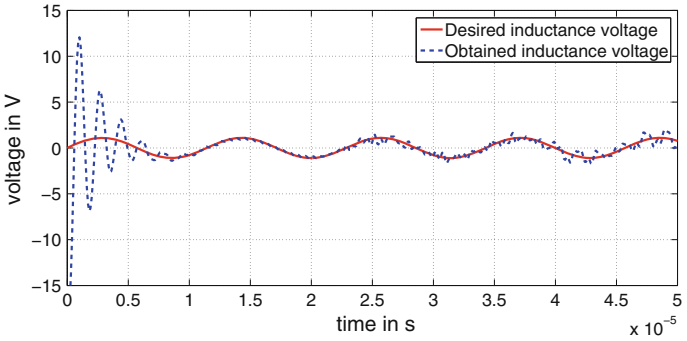


Fig. 6 Graphical representation of the obtained tracking control on variable $u_L(t)$ in case of $K_2 = 0$

tracking control obtained in the same case for variable $u_L(t)$ is presented in Fig. 7 in detail, in which the convergence of the tracking is not guaranteed. In fact, with the configuration of $K_2 = 0$ the observer is a Luenberger observer and, because of the presence of the disturbance and the uncertainties, the observer does not guarantee tracking convergence in terms of estimation. Figure 8 shows the desired and observed variable $\hat{u}_L(t)$. On the contrary, Fig. 9 shows a graphical representation of the tracking control obtained for the inductance voltage $u_L(t)$ in the case of $K_2 > 0$, also in presence of disturbance $d(t)$. Figure 10 presents the observed variable $\hat{u}_L(t)$ in detail. Figure 11 indicates the controlling variable $u_{\text{in}}(t)$ in the sliding strategy. It is possible to see that an initial tracking error occurred. This is due to the phase delay in the relative observed voltage $\hat{u}_L(t)$, which is depicted in Fig. 12 together with Figs. 13 and 14 in detail. The tracking of the desired signals, observed and controlled, is quite accurate. This is due to the known capability of the bang–bang observer and sliding mode techniques to make the whole control loop robust. Figure 15 shows the switching signal $\text{sign}\big(x_2(k-1) - \hat{x}_2(k-1)\big)$ of the bang–bang observer in which a filter is used to limit the chattering problem. The second reason for the phase delay is connected with this filtered switching signal of the bang–bang observer, because of

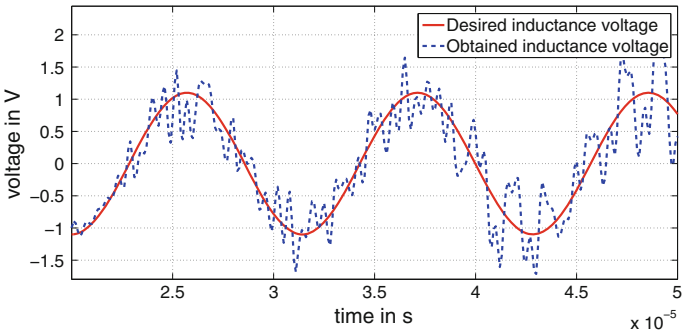


Fig. 7 Graphical representation of the obtained tracking control on variable $u_L(t)$ in case of $K_2 = 0$ in detail

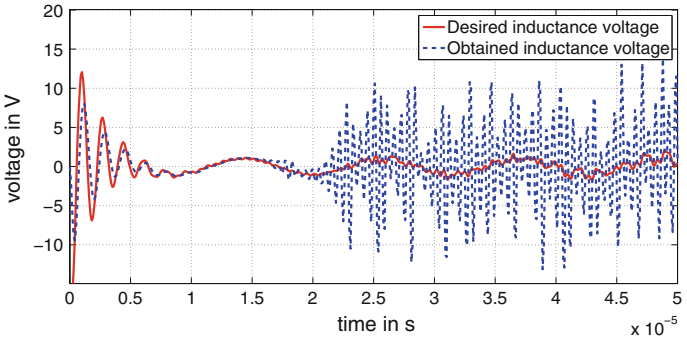


Fig. 8 Graphical representation of the desired and observed variable $\hat{u}_L(t)$ in case of $K_2 = 0$

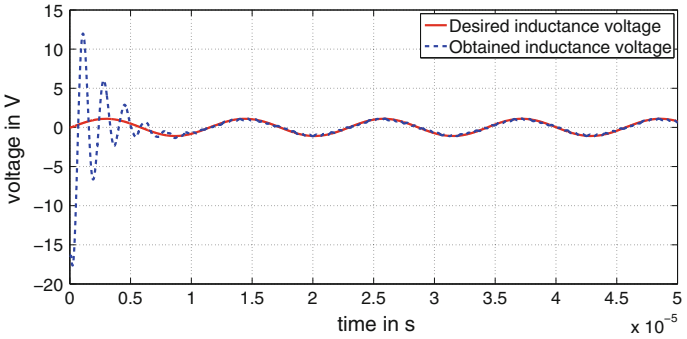


Fig. 9 Graphical representation of the obtained tracking control on variable $u_L(t)$ in case of $K_2 > 0$

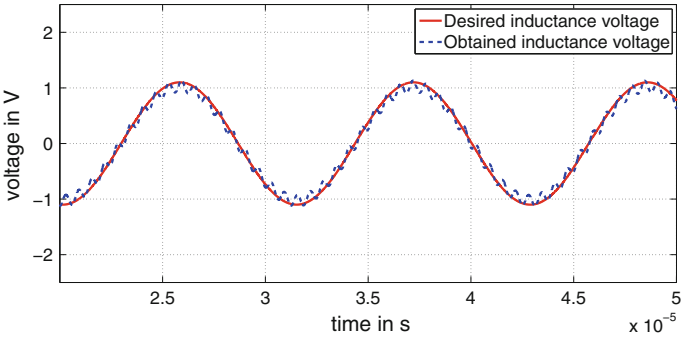


Fig. 10 Graphical representation of the obtained tracking control on variable $u_L(t)$ in case of $K_2 > 0$ in detail

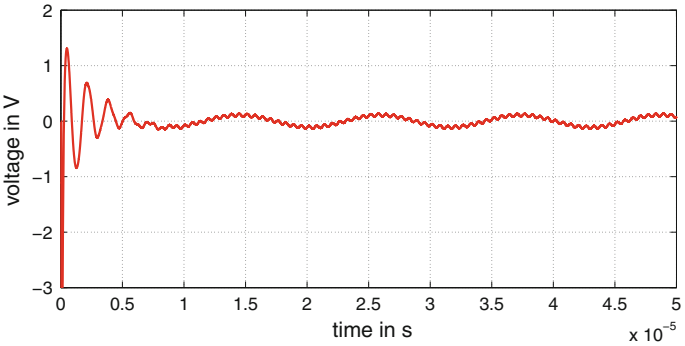


Fig. 11 Graphical representation of the obtained controlling variable $u_{\text{in}}(t)$ from the sliding strategy

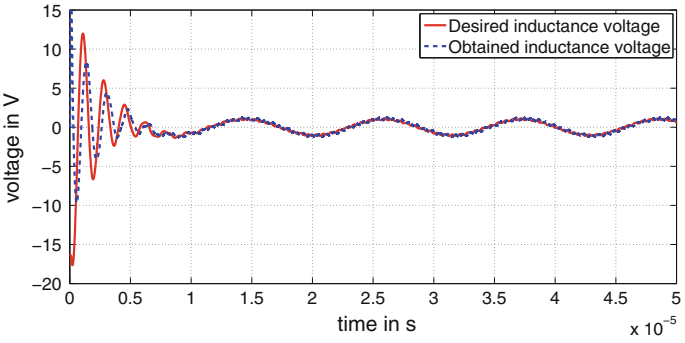


Fig. 12 Graphical representation of the desired and observed variable $\hat{u}_L(t)$

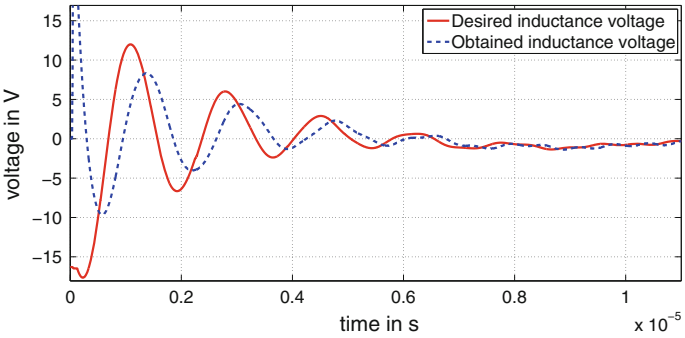


Fig. 13 Graphical representation of the desired and observed variable $\hat{u}_L(t)$ in detail with visible phase delay

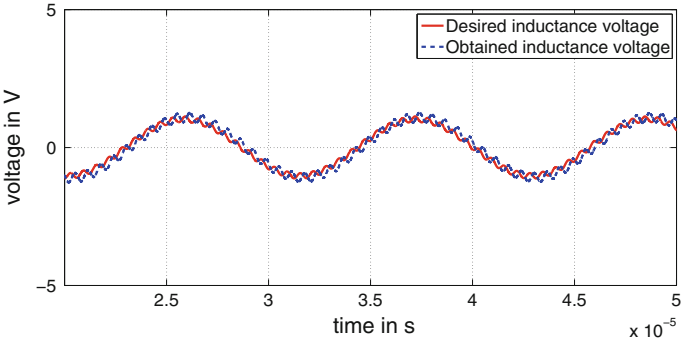


Fig. 14 Result of the desired and observed variable $\hat{u}_L(t)$ after a time sequence

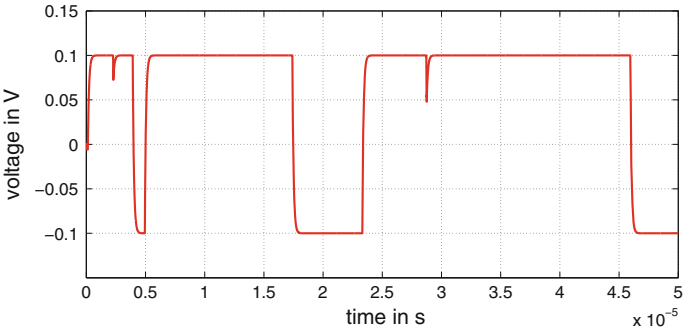


Fig. 15 Resulting filtered switching variable of the bang–bang observer

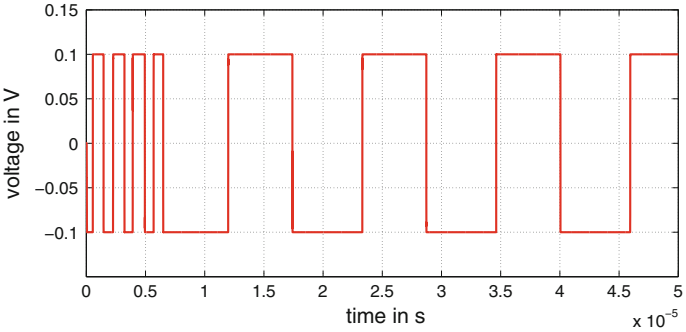


Fig. 16 Graphical representation filtered switching variable of the sliding mode control

the reduction of the switching frequency, due to the filter action, which manipulates the controlling signal which becomes slower. Therefore, the controlled output signal shows a delay. Figure 16 shows a graphical representation of the filtered switching sliding mode controller $\text{sign}\big(s(k-1)\big)$.

7 Conclusions

This chapter presented a SISO adaptive sliding mode control, combined with an adaptive bang–bang observer to improve a metal–polymer composite sensor system. The proposed techniques improved the disturbance rejection of the sensor system and thus their reliability in a harsh, industrial environment for the detection of particle pollution. The proposed application is innovative, as it endeavours to improve the performance of sensor robustness with respect to parametric uncertainties, as well as insensibility to disturbances. The whole control scheme was designed using the well-known Lyapunov approach and a particular sliding surface was defined in order to obtain the inductive voltage as a form of controlled output. A generic identification technique was obtained by use of scaling data. From this proposed technique, a broader sampling rate and low-frequency input signal was adopted to pinpoint the small-scale parameters. A further consideration might be based upon the reduction of the measuring system's transmitters themselves. The results of this study attest to the practicality and robustness of the algorithm set forth.

References

1. Corradini ML, Jetto L, Parlangeli G (2004) Robust stabilization of multivariable uncertain plants via switching control. IEEE Trans Autom Control 49(1):107–114
2. Jian-Xin X, Abidi K (2008) Discrete-time output integral sliding-mode control for a piezomotor-driven linear motion stage. IEEE Trans Ind Electron 55(11):3917–3926

3. Xinkai C, Hisayama T (2008) Adaptive sliding-mode position control for piezo-actuated stage. IEEE Trans Ind Electron 55(11):3927–3934
4. Pan Y, Ozgiiner O, Dagci OH (2008) Variable-structure control of electronic throttle valve. IEEE Trans Ind Electron 55(11):3899–3907
5. She JH, Xin X, Pan Y (2011) Equivalent-input-disturbance approach: analysis and application to disturbance rejection in dual-stage feed drive control system. IEEE/ASME Trans Mechatron 16(2):330340
6. Lee J-D, Duan R-Y (2011) Cascade modeling and intelligent control design for an electromagnetic guiding system. IEEE/ASME Trans Mechatron 16(3):470–479
7. Yang Y-P, Liu J-J, Ye D-H, Chen Y-R, Lu P-H (2013) Multiobjective optimal design and soft landing control of an electromagnetic valve actuator for a camless engine. IEEE/ASME Trans Mechatron 18(3):963–972
8. Levant A (2010) Chattering analysis. IEEE Trans Autom Control 55(6):1380–1389
9. Mercorelli P (2012) A two-stage augmented extended Kalman filter as an observer for sensorless valve control in camless internal combustion engines. IEEE Trans Ind Electron 59(11):4236–4247
10. Mercorelli P (2014) An adaptive and optimized switching observer for sensorless control of an electromagnetic valve actuator in camless internal combustion engines. Asian J Control (Wiley) 4(16):959–973
11. Rauh A, Aschemann H (2012) Interval-based sliding mode control and state estimation for uncertain systems. In: IEEE-17th international conference on methods and models in automation and robotics (MMAR), Miedzyzdrojie, pp 595–600
12. Senkel L, Rauh A, Aschemann H (2013) Optimal input design for online state and parameter estimation using interval sliding mode observers. In: IEEE-52nd annual conference on decision and control (CDC), Firenze, pp 502–507
13. Zhang J, Swain AK, Nguang SK (2012) Detection and isolation of incipient sensor faults for a class of uncertain non-linear systems. IET Control Theory Appl 6(12):1870–1880
14. Zhang J, Swain AK, Nguang SK (2014) Simultaneous robust actuator and sensor fault estimation for uncertain non-linear Lipschitz systems. IET Control Theory Appl 8(14):1364–1374
15. de Loza AF, Cieslak J, Henry D, Dávila J (2015) Sensor fault diagnosis using a non-homogeneous high-order sliding mode observer with application to a transport aircraft. IET Control Theory Appl 9(4):598–607
16. Schimmack M, Mercorelli P (2014) Contemporary sinusoidal disturbance detection and nano parameters identification using data scaling based on recursive least squares algorithms. In: IEEE CoDIT—international conference on control, decision and information technologies, France, pp 1528–1531
17. Gu D.-W, Petkov P, Konstantinov MM (2013) Modelling of uncertain systems. In: Robust control design with MATLAB®. Springer-Verlag, London. ISBN 978-1-84628-091-7
18. Ljung L (1999) System identification: theory for the user. Prentice-Hall, Upper Saddle River
19. Ljung L, Söderström T (1983) Theory and practice of recursive identification. MIT Press, Cambridge
20. Kailath T, Sayed AH, Hassibi B (2000) Linear estimation. Prentice Hall, Upper Saddle River

Sliding Mode Control for a Hydrostatic Transmission in Combination with a Sliding Mode Observer

Hao Sun and Harald Aschemann

Abstract Hydrostatic transmissions are continuously variable hydraulic power converters, which provide lots of advantages and represent a characteristic drive train component in, e.g. all types of working machines, city vehicles and renewable energy plants. In high-performance motion control systems, however, hydrostatic transmissions are less frequently used than electrical and mechanical drives due to their nonlinear behaviour, the impact of unknown disturbances like leakage volume flows as well as disturbance torques, and model uncertainty. In this contribution, a sliding mode approach is applied to the tracking control of a hydrostatic transmission. Moreover—in order to robustly reconstruct the immeasurable system states and the unknown disturbances—a gain-scheduled modified Utkin sliding mode observer is proposed that is based on extended linearisation techniques. This observer-based control structure is compared with an alternative approach, where a flatness-based tracking control is combined with a nonlinear reduced-order observer. The efficiency and the performance of the proposed control structure are highlighted by both simulations and meaningful experimental results.

1 Introduction

A hydrostatic transmission as depicted in Fig. 1 uses hydraulic oil to transmit power from the power source to the drive mechanism. A basic hydrostatic transmission consists of a hydraulic pump and a hydraulic motor, of which at least one must have a variable displacement, operating together in a closed circuit. On the pump side, the mechanical torque provided from the engine is transformed by the hydraulic pump into a pressurised fluid flow that is transformed back to the mechanical torque by the hydraulic motor on the motor side. By varying the tilt angle (changing the

H. Sun (✉) · H. Aschemann
Chair of Mechatronics, University of Rostock, Justus-von-Liebig-Weg 6,
18059 Rostock, Germany
e-mail: Hao.Sun@uni-rostock.de

H. Aschemann
e-mail: Harald.Aschemann@uni-rostock.de

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*,
Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_6

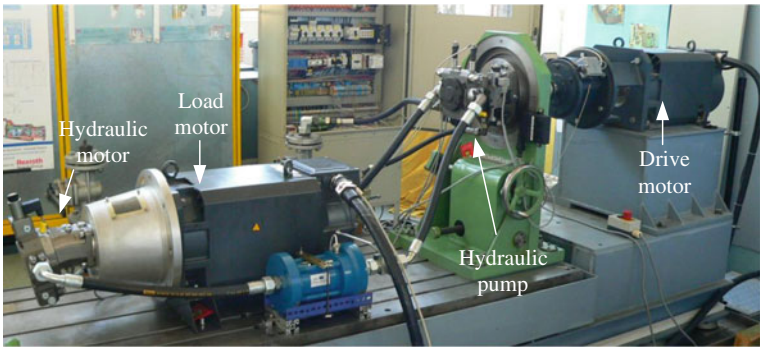


Fig. 1 Test bench of the hydrostatic transmission

displacement) of either pump or motor, any desired transmission ratio can be obtained within the boundaries determined by the design, cf. [12]. Hydrostatic transmissions are widely used as a characteristic component of drive chains practically in all types of working machines like harvesters, wheel loaders, excavators, telehandlers and agricultural tractors. They are typically operated in combination with diesel engines for mobile applications and offer a variety of advantages in comparison to pure mechanical transmissions. Besides the capability of a continuously variable transmission with high power density and the generation of large traction forces at low speeds, hydrostatic transmissions allow for reversing the direction of rotation without changing the gear. Moreover, it is possible to perform wearless breaking manoeuvres. Due to the significant advantages as compared to mechanical gearboxes, hydrostatic transmissions have also attracted the attention of engineers from renewable energy technology, cf. [5, 6].

However, hydrostatic transmissions are subject to several nonlinearities and characterised by uncertain system parameters as well as unknown disturbances. Even so, gain-scheduled PID-controllers are still widely used in current industrial practice for controlling such a system, see [21]. Hence, to improve both the energy efficiency and the control performance in practical applications, nonlinear model-based control approaches must be considered, for which plenty of work have been proposed in the last two decades, see [1–4, 10, 11, 13–15, 18–20, 22–27]. A flatness-based controller is proposed in [2, 20], in which only the unknown disturbance torque was considered as a lumped parameter. A disturbance compensation is realised by employing a nonlinear reduced-order disturbance observer. The simulation results and the experimental evaluation show a good tracking accuracy as well as active damping of pressure oscillations. The simplifying assumptions concerning the actuator dynamics and a constant leakage coefficient, however, restrict the applicability of this approach. In subsequent work [22–24], several advanced nonlinear approaches like adaptive inverse dynamics, robust inverse dynamics and sliding mode control have been investigated for the tracking control of the hydrostatic transmission system, in which the actuator time constants and leakage volume flow are considered as

uncertain parameters and a disturbance input, respectively. These centralised control approaches have been evaluated by simulations only. A decentralised flatness-based controller is presented in [1]. This innovative approach leads to a high tracking accuracy for both controlled outputs, and a singularity due to the vanishing pressure difference, cf. [2], can be avoided. Moreover, the implementation of the decentralised control structure is even simpler than the implementation of the centralised versions. Thereafter, different nonlinear control approaches based on the decentralised control structure with experimental validations can be found in [3, 4, 25, 26].

In [15], Liu proposed an exact linearisation control for a hydrostatic transmission with secondary regulation (SC-HST) and a good tracking performance is guaranteed in simulation by using LQR design based on the linearised model. A nonlinear H_∞ control in frequency domain via Generalised frequency response function (GFRF) for SC-HST is proposed by Jiang, however, only simulation validation can be found in [13]. Hoang, cf. [10], proposed an adaptive fuzzy sliding mode controller for a SC-HST system, where the saturation and dead zone effect existing in the displacement units are introduced into the system model and explicitly considered in the control design. The optimal control input is approximated by using fuzzy logic system. Moreover, a sliding mode control input is utilised to counteract the influence caused by the approximation error, parameter uncertainty as well as the disturbances. The implementation and experimental validation highlight the applicability of the proposed approach. Control design based on neural network has been also applied to hydrostatic transmissions, e.g. neural network based prediction control [19] and radial basis neural network (RBNN) control [11].

The nonlinear model demonstrated in this contribution is characterised by parameter uncertainty and unknown disturbances. Therefore, based on the decentralised control structure, a sliding mode control is proposed. Furthermore, in order to robustly reconstruct the immeasurable system states and estimate the unknown disturbances, the proposed control approach is extended with a modified Utkin sliding mode observer. The complete control algorithm is thoroughly studied by means of simulation and then implemented and validated by experiments for tracking control of the hydrostatic transmission. This chapter is organised as follows: in Sect. 2, the control-oriented nonlinear model of the hydrostatic transmission system is addressed. Based on the derived system model, the decentralised control structure is introduced. Hence, the control design procedure is divided into two sub-control loops according to the controlled variables. The details of the control design for each sub-control loop can be found in Sect. 3. In Sect. 4, two alternative state and disturbance observers: a nonlinear reduced-order observer and a modified Utkin sliding mode observer are introduced. The estimated immeasurable system states as well as unknown disturbances can then be used within the control structure. In Sect. 5, simulation results considering both measurement noise and quantization error as well as the experimental results form a dedicated test bench are presented. Finally, conclusions are given in Sect. 6.

2 Control-Oriented Model of a Hydrostatic Transmission

Dynamic system modelling plays a dominant role in modern control theory. An accurate system model is the key to improve the overall system performance. The model derivation of the hydrostatic transmission is based on the physical fundamentals of hydraulic systems, cf. [12]. A hydraulic scheme of the considered test bench is depicted in Fig. 2. This system can be split into a hydraulic subsystem and a mechanical subsystem, which are coupled by the torque τ_M generated by the hydraulic motor.

2.1 Hydraulic Subsystem

2.1.1 Pressure Dynamics

The hydraulic pump and motor are considered as key components in the hydrostatic transmission system, which are characterised by the volumetric displacement V_P and V_M in m^3 per revolution describing the fluid volume displaced within one complete rotation of the connected shaft. The pump and motor flow rates are determined by nonlinear functions

$$q_P = V_P(\alpha_P)\, n_P, \tag{1}$$

$$q_M = V_M(\alpha_M)\, n_M, \tag{2}$$

where n_P and n_M represent the rotational speed of the corresponding shaft in rpm. The calculations of the volumetric displacement $V_P(\alpha_P)$ and $V_M(\alpha_M)$ depend on the specific mechanical design. With the assumptions of a small swashplate angle $|\alpha_P| \leq 18°$ of the hydraulic pump and a small bent-axis angle $|\alpha_M| \leq 20°$ of the hydraulic motor, cf. [22], the volume flow (1) and (2) can be simplified as follows

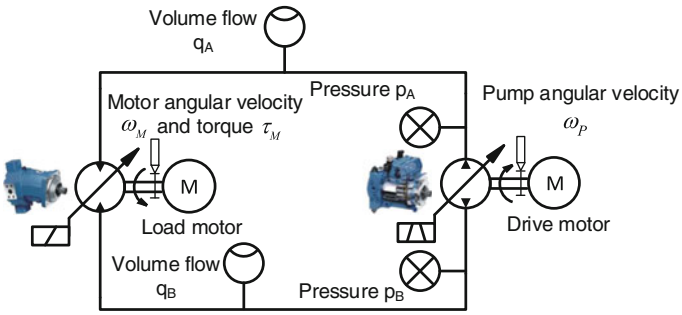


Fig. 2 Structure of the dedicated test bench

$$q_P = \tilde{V}_P \, \tilde{\alpha}_P \, \omega_P, \tag{3}$$

$$q_M = \tilde{V}_M \, \tilde{\alpha}_M \, \omega_M. \tag{4}$$

Here, $\tilde{V}_P$ and $\tilde{V}_M$ are constant parameters resulting from the geometric structure of the hydraulic pump and motor. The variable $\tilde{\alpha}_P \in (-1, 1)$ is the normalised swashplate angle of the pump, whereas $\tilde{\alpha}_M \in (\varepsilon_M, 1), \varepsilon_M > 0$, denotes the normalised bent-axis angle of the motor. The angular velocities of the pump and motor are represented by ω_P and ω_M, respectively.

Neglecting pressure losses in the hydraulic hoses, the high and low pressure sides of the system can be described by the corresponding pressure value p_A and p_B, respectively. In fact, the applications of hydrostatic transmissions have a working pressure in the range of hundreds of bar, hence, the hydraulic fluid cannot be simply considered as incompressible fluid. Using a mass balance equation in combination with an oil model, the pressure dynamics can be obtained as

$$\dot{p}_A = \frac{\beta_A}{V_A}(\tilde{V}_P \tilde{\alpha}_P \omega_P - \tilde{V}_M \tilde{\alpha}_M \omega_M - q_I - q_{E,A}), \tag{5}$$

$$\dot{p}_B = \frac{\beta_B}{V_B}(-\tilde{V}_P \tilde{\alpha}_P \omega_P + \tilde{V}_M \tilde{\alpha}_M \omega_M + q_I - q_{E,B}). \tag{6}$$

Here, β_A and β_B denote the effective bulk modulus of the fluid on high and low pressure sides, respectively. V_A and V_B represent the total volume (hydraulic hose and chamber) of the corresponding pressure side. Moreover, the internal leakage volume flow due to the pressure difference between the high and low pressure side is denoted by q_I. The variables $q_{E,A}$ and $q_{E,B}$ represent the external leakage volume flow of each pressure side, respectively, e.g. leakage volume flow due to imperfect sealing in the pump and motor. Introducing a reasonable symmetry assumption, an order reduction can be achieved regarding the pressure dynamics. This results in a first-order differential equation for the pressure difference Δp

$$\Delta \dot{p} = \dot{p}_A - \dot{p}_B = \frac{2}{C_H}\left(\tilde{V}_P \, \tilde{\alpha}_P \omega_P - \tilde{V}_M \, \tilde{\alpha}_M \omega_M\right) - \frac{q_U}{C_H}. \tag{7}$$

Here, the hydraulic capacitance is given by

$$C_H = \frac{V_A}{\beta_A} = \frac{V_B}{\beta_B}.$$

The corresponding leakage volume flow q_U results from internal and external leakage volume flows in the system

$$q_U = 2\, q_I + q_{E,A} - q_{E,B}.$$

2.1.2 Actuator Dynamics

A detailed description of a hydraulic pump and motor can be found in [16, 17]. However, the complicated system model is not suitable in nonlinear control design. For the pump or motor with electro-hydraulic control unit, the specific displacement depends linearly on the input current of the solenoid-servo valves. In general, the actuator dynamics of the displacement can be neglected in comparison with the dynamics of pressure difference. Accordingly, the dynamics of the displacement units for both pump and motor of the test bench are modelled by first-order lag systems, respectively. The differential equation for the corresponding normalised tilt angle becomes

$$T_{uP}\,\dot{\tilde{\alpha}}_P + \tilde{\alpha}_P = k_P\,u_P, \tag{8}$$

$$T_{uM}\,\dot{\tilde{\alpha}}_M + \tilde{\alpha}_M = k_M\,u_M. \tag{9}$$

The actuator time constants are denoted by T_{uP} and T_{uM}, and the input voltages u_P and u_M of the corresponding proportional valves for the displacement units act as physical control inputs. The parameters k_P and k_M represent the proportional gains of the first-order lag systems. Saturation functions account for the limited outputs of the actuators

$$\mathrm{sat}_b^a(\tilde{\alpha}_i) = \begin{cases} a & \tilde{\alpha}_i \geq a \\ \tilde{\alpha}_i & \text{for } b < \tilde{\alpha}_i < a, \\ b & \tilde{\alpha}_i \leq b \end{cases} \qquad i \in \{P, M\}, \tag{10}$$

where $a = \tilde{\alpha}_{i_{\max}}$ and $b = \tilde{\alpha}_{i_{\min}}$ represent the upper and lower output limits determined by the mechanical design. In the simulation model, Eq. (10) is implemented with limited integrators for $\tilde{\alpha}_P$ and $\tilde{\alpha}_M$, respectively. Note that the tilt angles $\tilde{\alpha}_P$ and $\tilde{\alpha}_M$ as well as the control inputs u_P and u_M represent normalised values, which are dimensionless.

2.2 Mechanical Subsystem

The dynamics of the output side of the hydrostatic transmission, see Fig. 2, is governed by the following equation of motion

$$J_V\,\dot{\omega}_M + d_V\,\omega_M = \underbrace{\tilde{V}_M\,\Delta p\,\tilde{\alpha}_M}_{\tau_M} - \tau_U, \tag{11}$$

with the abbreviation $J_V = J_M + J_E$. Here, J_M represents the mass moment of inertia of the hydraulic motor and J_E the one of the electric motor on the load side. The

parameter d_V is the damping coefficient at the drive shaft. The driving torque of the hydraulic motor is denoted by τ_M, an unknown disturbance torque by τ_U.

2.3 Models for the Decentralised Control Design

The overall system model comprises four first-order differential equations: the state equations (7) for the pressure difference Δp, (8) and (9) for the normalised displacement angles of the pump $\tilde{\alpha}_P$ and the motor $\tilde{\alpha}_M$ as well as (11) for the motor angular velocity ω_M. For the decentralised control design, they are partitioned as follows: the first design model is linear and given by the differential equation for the normalised motor bent-axis angle $\tilde{\alpha}_M$

$$\dot{\tilde{\alpha}}_M = -\frac{1}{T_{uM}}\tilde{\alpha}_M + \frac{k_M}{T_{uM}}u_M, \tag{12}$$

where u_M serves as the control input. Introducing the normalised pump tilt angle $\tilde{\alpha}_P$, the pressure difference Δp, and the motor angular velocity ω_M as state variables, the state vector of the second design model results in $\boldsymbol{x} = [\tilde{\alpha}_P \ \Delta p \ \omega_M]^T$. The corresponding nonlinear state-space representation $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, u_P, \boldsymbol{\tau}_{dis})$ becomes

$$\begin{bmatrix} \dot{\tilde{\alpha}}_P \\ \Delta\dot{p} \\ \dot{\omega}_M \end{bmatrix} = \begin{bmatrix} -\frac{1}{T_{uP}}\tilde{\alpha}_P + \frac{k_P}{T_{uP}}u_P \\ \frac{2\tilde{V}_P\omega_P}{C_H}\operatorname{sat}^1_{-1}(\tilde{\alpha}_P) - \frac{2\tilde{V}_M\omega_M}{C_H}\operatorname{sat}^1_{\varepsilon_M}(\tilde{\alpha}_M) - \frac{q_U}{C_H} \\ -\frac{d_V}{J_V}\omega_M + \frac{\tilde{V}_M}{J_V}\Delta p \operatorname{sat}^1_{\varepsilon_M}(\tilde{\alpha}_M) - \frac{\tau_U}{J_V} \end{bmatrix}, \tag{13}$$

$$\boldsymbol{y}_m = \begin{bmatrix} \Delta p \ \omega_M \end{bmatrix}^T. \tag{14}$$

The control input is given by u_P, the disturbance vector by $\boldsymbol{\tau}_{dis} = [q_U \ \tau_U]^T$. The output vector $\boldsymbol{y}_m$ contains the pressure difference Δp as well as the motor angular velocity ω_M. As a consequence, the multi-input and multi-output (MIMO) dynamic system is considered as two single-input single-output (SISO) subsystems, where couplings are taken into account by gain-scheduling techniques. Moreover, by exploiting the flatness property of the decentralised system model, the saturation functions are not explicitly considered in the control design in the following sequel.

3 Model-Based Decentralised Control Design

It is straightforward to prove that each SISO subsystem is differentially flat with the corresponding flat output $y_{f1} = \tilde{\alpha}_M$ and $y_{f2} = \omega_M$, respectively. The subsystem of the normalised bent-axis angle $\tilde{\alpha}_M$ is presented by a first-order differential equation (12). It is obvious to obtain

$$u_M = \Omega_1(\dot{y}_{f1}, y_{f1}) = T_{uM}\dot{\tilde{\alpha}}_M + k_M\,\tilde{\alpha}_M, \tag{15}$$

and $\dim(y_{f1}) = \dim(u_M)$. Therefore, $y_{f1} = \tilde{\alpha}_M$ represents the flat output of this subsystem.

The subsystem of the motor angular velocity ω_M is governed by Eq. (13). The system states Δp and $\tilde{\alpha}_P$ can be expressed as a function of ω_M as well as its time derivatives, i.e.

$$\Delta p = \Omega_2(\dot{y}_{f2}, y_{f2}) = \frac{J_V\,\dot{\omega}_M + d_V\,\omega_M + \tau_U}{\tilde{V}_M\,\tilde{\alpha}_M}, \tag{16}$$

$$\begin{aligned}\tilde{\alpha}_P &= \Omega_3(\ddot{y}_{f2}, \dot{y}_{f2}, y_{f2}) \\ &= \frac{J_V C_H}{2\tilde{V}_P\omega_P\tilde{V}_M\tilde{\alpha}_M}\,\ddot{\omega}_M + \frac{d_V C_H}{2\tilde{V}_P\omega_P\tilde{V}_M\tilde{\alpha}_M}\,\dot{\omega}_M + \frac{\tilde{V}_M\tilde{\alpha}_M}{\tilde{V}_P\omega_P}\,\omega_M \\ &\quad + \frac{C_H\dot{\tau}_U}{2\tilde{V}_P\omega_P\tilde{V}_M\tilde{\alpha}_M} + \frac{q_U}{2\tilde{V}_P\omega_P}.\end{aligned} \tag{17}$$

Here, again, $\tilde{\alpha}_M$ is considered as a gain-scheduling parameter. Moreover, the control input $u_P = \Omega_4(\dddot{y}_{f2}, \ddot{y}_{f2}, \dot{y}_{f2}, y_{f2})$ can be determined by solving the third time derivative of ω_M for the input u_P. Together with the condition $\dim(y_{f2}) = \dim(u_P)$, it can be concluded that the subsystem for the control of the motor angular velocity ω_M is differentially flat with the flat output $y_{f2} = \omega_M$.

The proposed decentralised control scheme is depicted in Fig. 3. A flatness-based control with an integrator is employed to guarantee an excellent tracking performance for the control of the normalised bent-axis angle $\tilde{\alpha}_M$ of the motor. For the control of the motor angular velocity ω_M, a sliding mode control approach is applied to obtain a robust tracking performance regarding the model uncertainty. The immeasurable system states—the normalised tilt angles $\tilde{\alpha}_P$ and $\tilde{\alpha}_M$—and the unknown disturbances—

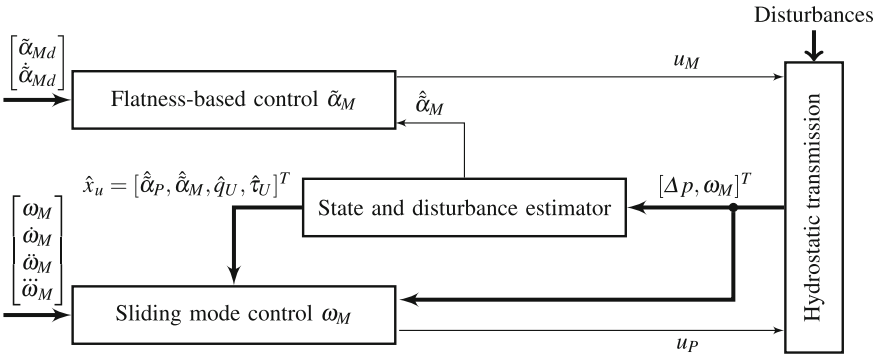


Fig. 3 Scheme of the decentralised control structure in combination with a state and disturbance observer

the leakage volume flow q_U as well as the unknown disturbance torque τ_U—are estimated online by means of state and disturbance observers proposed in Sect. 4.

The proposed decentralised control structure outperforms an alternative centralised one published in [2], where the flat outputs given by the pressure difference Δp and the motor angular velocity ω_M are considered as controlled variables. It avoids a tracking control of the pressure difference Δp, which

- leads to a stiff control system due to the small oil compressibility,
- causes a singularity in the control design, cf. [2].

The proposed decentralised control structure simplifies the control design procedure significantly and, moreover, reduces the effort of the control implementation. In the following, the control design for each subsystem is presented and discussed in detail.

3.1 Control Design of the Normalised Bent-Axis Angle

A flatness-based control (FBC) approach is employed for the tracking control of the normalised bent-axis angle $\tilde{\alpha}_M$ of the motor. Therefore, the inverse dynamics of the corresponding design model (12) is considered

$$u_M = \frac{\tilde{\alpha}_M + \upsilon_M \, T_{uM}}{k_M}, \tag{18}$$

where the first time derivative $\dot{\tilde{\alpha}}_M = \upsilon_M$ is introduced as a stabilising control input. The stabilising control law represents a combination of a feedforward and feedback control as follows

$$\upsilon_M = \dot{\tilde{\alpha}}_{Md} + k_{\alpha 0} \cdot \underbrace{(\tilde{\alpha}_{Md} - \tilde{\alpha}_M)}_{e_{\tilde{\alpha}_M}} + k_{\alpha I} \cdot \int_0^t \underbrace{(\tilde{\alpha}_{Md} - \tilde{\alpha}_M)}_{e_{\tilde{\alpha}_M}} \, d\tau. \tag{19}$$

Therefore, a linear error dynamics is obtained and positive coefficients $k_{\alpha 0} > 0$ and $k_{\alpha I} > 0$ can be chosen according to the desired characteristic polynomial

$$p_{\tilde{\alpha}_M}(s) = (s + s_{\tilde{\alpha}_M,1})(s + s_{\tilde{\alpha}_M,2}), \tag{20}$$

to achieve asymptotic stability. Moreover, the steady-state accuracy is guaranteed by the integral part despite the uncertainty in the system dynamics, e.g. uncertain time constant T_{uM}. Without loss of generality, desired trajectories $\tilde{\alpha}_{Md}$ are specified only within the admissible displacement range $(\varepsilon_M, 1)$.

3.2 Control Design of the Motor Angular Velocity

The motor angular velocity ω_M represents a flat output for the second design model (13) with a reasonable assumption that the normalised bent-axis angle $\tilde{\alpha}_M$, serving as a gain-scheduling parameter, has a negligible time derivative. Moreover, the disturbance models $\dot{\tau}_U = 0$ and $\dot{q}_U = 0$ are considered accordingly in the subsequent derivation. As a consequence, the third time differentiation of ω_M results in

$$\dddot{\omega}_M = f(\boldsymbol{x}, \tilde{\alpha}_M, \boldsymbol{\tau}_{dis}) + g(\tilde{\alpha}_M)\, u_P, \tag{21}$$

where $f(\boldsymbol{x}, \tilde{\alpha}_M, \boldsymbol{\tau_{dis}})$ and $g(\tilde{\alpha}_M)$ represent state- and disturbance-dependent continuous nonlinear functions which can be expressed in the linear form of model parameters as

$$f(\boldsymbol{x}, \tilde{\alpha}_M, \boldsymbol{\tau_{dis}}) = \boldsymbol{\theta}_1^T(\tilde{\alpha}_M)\, \boldsymbol{x} + \boldsymbol{\theta}_2^T(\tilde{\alpha}_M)\, \boldsymbol{\tau}_{dis} = \boldsymbol{\theta}^T(\tilde{\alpha}_M)\, \boldsymbol{x}_E, \tag{22}$$

$$g(\tilde{\alpha}_M) = \theta_3(\tilde{\alpha}_M). \tag{23}$$

Here, $\boldsymbol{\theta}^T(\tilde{\alpha}_M) = [\boldsymbol{\theta}_1^T(\tilde{\alpha}_M)_{1\times 4}\, \boldsymbol{\theta}_2^T(\tilde{\alpha}_M)_{1\times 2}]$ denotes a state-dependent coefficient vector, and $\theta_3(\tilde{\alpha}_M)$ is a state-dependent scalar function, i.e. the elements in $\boldsymbol{\theta}$ as well as the scalar θ_3 are functions of physical parameters and the normalised bent-axis angle $\tilde{\alpha}_M$. The extended state vector $\boldsymbol{x}_E = [\boldsymbol{x}^T\ \boldsymbol{\tau}_{dis}^T]^T$ contains the states of the subsystem as well as the disturbance inputs.

3.2.1 Flatness-Based Control

Equation (21) depends directly on the control input u_P and can be solved for the flatness-based control (FBC) input

$$u_P = g(\tilde{\alpha}_M)^{-1}\big(\upsilon_P - f(\boldsymbol{x}, \tilde{\alpha}_M, \boldsymbol{\tau}_{dis})\big), \tag{24}$$

with an additional stabilising input υ_P. In the case that a perfect system model can be obtained, the control input (24) can lead to a closed-loop system as

$$\dddot{\omega}_M = \upsilon_P. \tag{25}$$

Here, υ_P can be easily designed as

$$\upsilon_P = \dddot{\omega}_{Md} + k_{\omega_M,2}(\ddot{\omega}_{Md} - \ddot{\omega}_M) + k_{\omega_M,1}(\dot{\omega}_{Md} - \dot{\omega}_M) + k_{\omega_M,0}(\omega_{Md} - \omega_M). \tag{26}$$

Therefore, a linear error dynamics is imposed and the asymptotic stability can be achieved by determining the coefficients $k_{\omega_M,i}, i \in \{0, 1, 2\}$ according to the Hurwitz polynomial

$$p_{\omega_M}(s) = (s + s_{\omega_M})^3. \tag{27}$$

In practice, however, it is impossible to achieve a perfect knowledge of all the physical parameters in the system dynamic model. The system dynamics with parameter uncertainty can be expressed as

$$\dddot{\omega}_M = \hat{f}(\boldsymbol{x}, \tilde{\alpha}_M, \boldsymbol{\tau}_{dis}) + \hat{g}(\tilde{\alpha}_M)\, u_P - \tilde{f}(\boldsymbol{x}, \tilde{\alpha}_M, \boldsymbol{\tau}_{dis}) - \tilde{g}(\tilde{\alpha}_M)\, u_P, \tag{28}$$

where $\hat{(\cdot)}$ denotes the estimate value, i.e.

$$\hat{f}(\boldsymbol{x}, \tilde{\alpha}_M, \boldsymbol{\tau}_{dis}) = \hat{\boldsymbol{\theta}}^T(\tilde{\alpha}_M)\, \boldsymbol{x}_E,\ \ \hat{g}(\tilde{\alpha}_M) = \hat{\theta}_3(\tilde{\alpha}_M), \tag{29}$$

and the difference between the estimate and real values is given by $\tilde{(\cdot)} = \hat{(\cdot)} - (\cdot)$. Therefore, the dynamics (28) can be rewritten as

$$\dddot{\omega}_M = \hat{f}(\boldsymbol{x}, \tilde{\alpha}_M, \boldsymbol{\tau}_{dis}) + \hat{g}(\tilde{\alpha}_M)\, u_P - \underbrace{(\tilde{\boldsymbol{\theta}}^T(\tilde{\alpha}_M)\, \boldsymbol{x}_E + \tilde{\theta}_3(\tilde{\alpha}_M)\, u_P)}_{\psi(\tilde{\boldsymbol{\theta}},\ \tilde{\theta}_3,\ \boldsymbol{x}_E,\ u_P)}, \tag{30}$$

where $\psi(\tilde{\boldsymbol{\theta}},\ \tilde{\theta}_3,\ \boldsymbol{x}_E,\ u_P)$ denotes the system uncertainty which is upper bounded by $|\psi| \leq \psi_{\max}$. In this situation, it is obvious to see that Eq. (24) cannot guarantee a perfect compensation of the system nonlinear dynamics. The compensation residue ψ has a significant effect on the closed-loop tracking performance, which is, however, not considered explicitly in the FBC. To achieve more robust performance, a sliding mode control is considered, cf. [7], for the tracking control of the motor angular velocity ω_M represented in (30).

3.2.2 Sliding Mode Control

The design of sliding mode control starts with defining a sliding manifold as

$$s = \ddot{e}_\omega + c_2 \dot{e}_\omega + c_1 e_\omega, \tag{31}$$

with the tracking error $e_\omega = \omega_{Md} - \omega_M$. The candidate for a Lyapunov function can be chosen as

$$V = \frac{1}{2} s^2 \geq 0. \tag{32}$$

Substituting (30) into the first time derivative of (32) results in

$$\begin{aligned} \dot{V} &= s\dot{s} = s(\dddot{\omega}_{Md} + c_2 \ddot{e}_\omega + c_1 \dot{e}_\omega - \dddot{\omega}_M) \\ &= s(\dddot{\omega}_{Md} + c_2 \ddot{e}_\omega + c_1 \dot{e}_\omega - \hat{f}(\boldsymbol{x}, \tilde{\alpha}_M, \boldsymbol{\tau}_{dis}) - \hat{g}(\tilde{\alpha}_M)\, u_P + \psi). \end{aligned} \tag{33}$$

The control input u_P can be designed as

$$u_P = \hat{g}(\tilde{\alpha}_M)^{-1} \left(\dddot{\omega}_{Md} + c_2 \ddot{e}_\omega + c_1 \dot{e}_\omega - \hat{f}(\boldsymbol{x}, \tilde{\alpha}_M, \boldsymbol{\tau}_{dis}) + u_{sm} \right). \tag{34}$$

Here, u_{sm} denotes the sliding mode control input introduced as

$$u_{sm} = \lambda s + k \, \mathrm{sign}(s), \tag{35}$$

where λ is a positive scalar. The parameter k denotes the gain value of the switching term, and it can be either constant or time-varying. Substituting (34) and (35) into (33) leads to

$$\dot{V} = -\lambda s^2 + s(\psi - k \, \mathrm{sign}(s)) = -\lambda s^2 + |s|(\psi \, \mathrm{sign}(s) - k). \tag{36}$$

To guarantee asymptotic stability, the following condition must be satisfied

$$k \geq \psi \, \mathrm{sign}(s). \tag{37}$$

Considering the boundary condition of parameter uncertainty ψ, it is obvious to conclude that the choice $k \geq \psi_{\max}$ leads to $\dot{V} \leq 0$, which guarantees the asymptotic stability of the system. Substituting the proposed sliding mode control inputs (34) and (35) into the dynamic system (30) results in the closed-loop error dynamics

$$\dddot{e}_\omega + (c_2 + \lambda) \, \ddot{e}_\omega + (c_1 + \lambda c_2) \, \dot{e}_\omega + \lambda c_1 \, e_\omega + k \, \mathrm{sign}(s) - \psi = 0. \tag{38}$$

It can be seen that the positive parameters c_1, c_2 and λ can be chosen according to the Hurwitz polynomial (27). The gain k of the switching term in the sliding mode control input is determined by considering the upper bound of the uncertainty ψ. Therefore, substituting the sliding mode control input u_{sm} (35) into (34) and reconsidering the analytical expression ψ in (30) results in

$$\begin{aligned} \psi =& L \, (\dddot{\omega}_{Md} + c_2 \ddot{e}_\omega + c_1 \dot{e}_\omega + \lambda s + k \, \mathrm{sign}(s)) + \underbrace{(\tilde{\boldsymbol{\theta}} - L \, \hat{\boldsymbol{\theta}})^T \, \boldsymbol{x}_E}_{\phi} \\ =& L \, k \, \mathrm{sign}(s) + L \, (c_2 \ddot{e}_\omega + c_1 \dot{e}_\omega + \lambda s) + L \, \dddot{\omega}_{Md} + \phi, \end{aligned} \tag{39}$$

with

$$L = \frac{\tilde{\theta}_3}{\hat{\theta}_3} = 1 - \frac{\theta_3}{\hat{\theta}_3}.$$

Since all the physical parameters of the hydrostatic transmission are positive and bounded, a boundary condition for (39) can be derived as follows

$$|\psi| \leq \gamma_1 |k| + \gamma_2 |e_\omega| + \gamma_3 |e_\omega|^2 + \gamma_4 |\dot{e}_\omega| + \gamma_5, \tag{40}$$

where $\gamma_i,\ i \in \{1, 2, 3, 4, 5\}$, are positive constant scalars. With the fact that k is also bounded by $\psi_{\max}$, the bound of the system uncertainty can be determined as

$$\psi_{\max} = \frac{1}{1-\gamma_1}\left(\gamma_2|e_\omega| + \gamma_3|e_\omega|^2 + \gamma_4|\dot{e}_\omega| + \gamma_5\right). \tag{41}$$

The constant coefficients $\gamma_i,\ i \in \{1, 2, 3, 4\}$ are the weighting parameters in (41) and γ_5 represents the upper bound of a nonlinear function of states and disturbances. Therefore, simulation studies are performed to assess the magnitudes for the corresponding parameters.

The sliding mode control presented so far still suffers from the chattering phenomenon. This undesirable phenomenon is caused by fast switching actions introduced by the sign function and may lead to the excitation of unmodelled high-frequency dynamics. To counteract this effect, a smooth switching function $\tanh(s/\eta)$ with a strictly positive, small constant η—determining a boundary layer thickness—is used. The chattering reduction depends on the value of η at the cost of robustness. The larger the value of η, the more the chattering phenomenon is suppressed.

4 State and Disturbance Observer

For the implementation of the proposed control approaches, two alternative state and disturbance observers are considered in this section—a nonlinear reduced-order observer and a gain-scheduled sliding mode observer. Both provide estimates for the immeasurable system states—the normalised swashplate angle $\tilde{\alpha}_P$ of the pump and the normalised bent-axis angle $\tilde{\alpha}_M$ of the motor—and for the unknown system disturbances—the leakage volume flow q_U and a disturbance torque τ_U acting on the hydraulic motor. For the observer design, the slowly time-varying disturbances are assumed. Hence, integrator models are employed as disturbance models, i.e.

$$\dot{\boldsymbol{\tau}}_{dis} = [\dot{q}_U\ \dot{\tau}_U]^T = \mathbf{0}. \tag{42}$$

4.1 Nonlinear Reduced-Order Observer

The nonlinear state and disturbance observer as proposed in [8] is designed for the hydrostatic transmission. Therefore, the state equations (12) and (13) are extended with disturbance models given by (42). With the additional state variables for the disturbances q_U and τ_U, the vector of immeasurable system states can be stated as

$$\boldsymbol{x}_u = [\tilde{\alpha}_P\ \tilde{\alpha}_M\ q_U\ \tau_U]^T.$$

Therefore, the extended system model for the nonlinear reduced-order observer design becomes

$$\dot{\boldsymbol{y}}_m = \begin{bmatrix} \Delta\dot{p} \\ \dot{\omega}_M \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{2\tilde{V}_P\omega_P}{C_H}\tilde{\alpha}_P - \frac{2\tilde{V}_M\omega_M}{C_H}\tilde{\alpha}_M - \frac{q_U}{C_H} \\ -\frac{d_V}{J_V}\omega_M + \frac{\tilde{V}_M}{J_V}\Delta p\,\tilde{\alpha}_M - \frac{\tau_U}{J_V} \end{bmatrix}}_{\boldsymbol{f}_m(\boldsymbol{y}_m, \boldsymbol{x}_u)}, \tag{43}$$

$$\dot{\boldsymbol{x}}_u = \begin{bmatrix} \dot{\tilde{\alpha}}_P \\ \dot{\tilde{\alpha}}_M \\ \dot{q}_U \\ \dot{\tau}_U \end{bmatrix} = \underbrace{\begin{bmatrix} -\frac{1}{T_{uP}}\tilde{\alpha}_P + \frac{k_P}{T_{uP}}u_P \\ -\frac{1}{T_{uM}}\tilde{\alpha}_M + \frac{k_M}{T_{uM}}u_M \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{f}_u(\boldsymbol{x}_u, \boldsymbol{u})}. \tag{44}$$

Here, $\boldsymbol{u} = [u_P\ u_M]^T$ represents the input vector. The estimated states $\hat{\boldsymbol{x}}_u$ follow from

$$\hat{\boldsymbol{x}}_u = \boldsymbol{H}\,\boldsymbol{y}_m + \boldsymbol{z}, \tag{45}$$

where $\boldsymbol{H}$ denotes the observer gain matrix. The state equations for the observer state vector $\boldsymbol{z}$ are chosen as

$$\dot{\boldsymbol{z}} = \boldsymbol{\Phi}\left(\boldsymbol{y}_m, \hat{\boldsymbol{x}}_u, \boldsymbol{u}\right). \tag{46}$$

The observer gain matrix with four unknowns

$$\boldsymbol{H} = \begin{bmatrix} h_{11} & 0 & h_{31} & 0 \\ 0 & h_{22} & 0 & h_{42} \end{bmatrix}^T, \tag{47}$$

and the vector of nonlinear functions $\boldsymbol{\Phi}$ are determined in such a way that the steady-state observer error $\tilde{\boldsymbol{x}}_u = \boldsymbol{x}_u - \hat{\boldsymbol{x}}_u$ converges to zero. The design proceeds by calculating the first time derivative of the observer error

$$\dot{\tilde{\boldsymbol{x}}}_u = \dot{\boldsymbol{x}}_u - \dot{\hat{\boldsymbol{x}}}_u = \dot{\boldsymbol{x}}_u - \boldsymbol{H}\cdot\dot{\boldsymbol{y}}_m - \boldsymbol{\Phi}\left(\boldsymbol{y}_m, \boldsymbol{x}_u - \tilde{\boldsymbol{x}}_u, \boldsymbol{u}\right). \tag{48}$$

Substituting (43) and (44) into (48) leads to

$$\dot{\tilde{\boldsymbol{x}}}_u = \boldsymbol{f}_u(\boldsymbol{x}_u,\ \boldsymbol{u}) - \boldsymbol{H}\cdot\boldsymbol{f}_m(\boldsymbol{y}_m, \boldsymbol{x}_u) - \boldsymbol{\Phi}\left(\boldsymbol{y}_m, \boldsymbol{x}_u - \tilde{\boldsymbol{x}}_u, \boldsymbol{u}\right). \tag{49}$$

With the stationary condition that the steady-state error becomes $\tilde{\boldsymbol{x}}_u = \boldsymbol{0}$, the vanishing of the right hand side in (49) results in

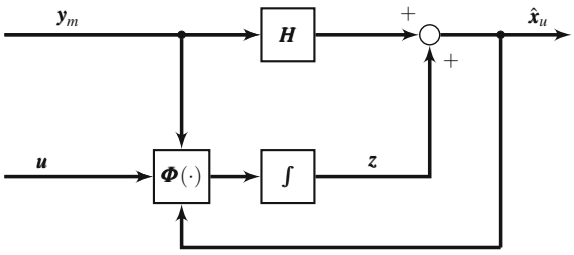


Fig. 4 Block diagram of the nonlinear reduced-order observer

$$\boldsymbol{\Phi}\left(\boldsymbol{y}_m, \hat{\boldsymbol{x}}_u, \boldsymbol{u}\right) = -\boldsymbol{H} \cdot \begin{bmatrix} \frac{2\tilde{V}_P \omega_P}{C_H}\hat{\alpha}_P - \frac{2\tilde{V}_M \omega_M}{C_H}\hat{\alpha}_M - \frac{\hat{q}_U}{C_H} \\ -\frac{d_V}{J_V}\omega_M + \frac{\tilde{V}_M}{J_V}\Delta p\hat{\alpha}_M - \frac{\hat{\tau}_U}{J_V} \end{bmatrix} + \begin{bmatrix} \frac{-\hat{\alpha}_P + k_P u_P}{T_{uP}} \\ \frac{-\hat{\alpha}_M + k_M u_M}{T_{uM}} \\ 0 \\ 0 \end{bmatrix}. \quad (50)$$

To achieve asymptotic stability, the linearised error system

$$\dot{\tilde{\boldsymbol{x}}}_u = \frac{\partial \boldsymbol{\Phi}\left(\boldsymbol{y}_m, \hat{\boldsymbol{x}}_u, \boldsymbol{u}\right)}{\partial \hat{\boldsymbol{x}}_u}\tilde{\boldsymbol{x}}_u \quad (51)$$

is considered. Therefore, all eigenvalues of the Jacobian are placed in the left complex half-plane according to

$$\det\left(s\boldsymbol{I} - \frac{\partial \boldsymbol{\Phi}\left(\boldsymbol{y}_m, \hat{\boldsymbol{x}}_u, \boldsymbol{u}\right)}{\partial \hat{\boldsymbol{x}}_u}\right) \stackrel{!}{=} \prod_{i=1}^{4}(s + s_{Bi}). \quad (52)$$

With positive values $s_{Bi} > 0, i \in \{1, 2, 3, 4\}$, the observer gain matrix $\boldsymbol{H} = \boldsymbol{H}(\Delta p)$ according to (47), follows directly from Eq. (52) and depends on the pressure difference Δp. A block diagram of the implementation can be found in Fig. 4.

4.2 *Sliding Mode Observer Design Based on Extended Linearisation*

A modified Utkin sliding mode observer, cf. [7], in combination with extended linearisation techniques, cf. [8], is proposed in this section to provide a robust state and disturbance construction. For this purpose, the right hand side of the dynamic equations (12) and (13) as well as the disturbance models (42) is rearranged. Following the idea of extended linearisation, they are written in a quasi-linear form with a state-dependent system matrix

$$\underbrace{\begin{bmatrix} \dot{\tilde{\alpha}}_P \\ \dot{\tilde{\alpha}}_M \\ \Delta\dot{p} \\ \dot{\omega}_M \\ \dot{q}_U \\ \dot{\tau}_U \end{bmatrix}}_{\dot{\boldsymbol{x}}_{EL}} = \underbrace{\begin{bmatrix} -\frac{1}{T_{uP}} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{T_{uM}} & 0 & 0 & 0 & 0 \\ \frac{2\tilde{V}_P\omega_P}{C_H} & -\frac{2\tilde{V}_M\omega_M}{C_H} & 0 & 0 & -\frac{1}{C_H} & 0 \\ 0 & 0 & \frac{\tilde{V}_M\tilde{\alpha}_M}{J_V} & -\frac{d_V}{J_V} & 0 & \frac{1}{-J_V} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\boldsymbol{A}(\tilde{\alpha}_M,\omega_M)} \underbrace{\begin{bmatrix} \tilde{\alpha}_P \\ \tilde{\alpha}_M \\ \Delta p \\ \omega_M \\ q_U \\ \tau_U \end{bmatrix}}_{\boldsymbol{x}_{EL}} + \underbrace{\begin{bmatrix} \frac{k_P}{T_{uP}} & 0 \\ 0 & \frac{k_P}{T_{uM}} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\boldsymbol{B}} \underbrace{\begin{bmatrix} u_P \\ u_M \end{bmatrix}}_{\boldsymbol{u}}, \tag{53}$$

$$\underbrace{\begin{bmatrix} \Delta p \\ \omega_M \end{bmatrix}}_{\boldsymbol{y}_m} = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}}_{\boldsymbol{C}} \boldsymbol{x}_{EL}. \tag{54}$$

Here, the system matrix $\mathbf{A}(\tilde{\alpha}_M, \omega_M)$ depends on the two system states—the normalised bent-axis angle $\tilde{\alpha}_M$ and the motor angular velocity ω_M. The input matrix $\mathbf{B}$ and the output matrix $\boldsymbol{C}$ are constant. Moreover, the system represented in (53) is completely observable. As the measured outputs are states of the dynamic system, a coordinate transformation

$$\boldsymbol{x}_c = \boldsymbol{T}_c\, \boldsymbol{x}_{EL}$$

is reasonable to rearrange the order of the states according to

$$\boldsymbol{x}_c = [\tilde{\alpha}_P\ \tilde{\alpha}_M\ q_U\ \tau_U \vdots \Delta p\ \omega_M]^T = [\boldsymbol{x}_1^T \vdots \boldsymbol{y}_m^T]^T.$$

The corresponding system matrix, input matrix as well as output matrix of the transformed representation can be calculated as follows

$$\boldsymbol{A}_c = \boldsymbol{T}_c\, \boldsymbol{A}\, \boldsymbol{T}_c^{-1} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix},\ \boldsymbol{B}_c = \boldsymbol{T}_c\, \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix},\ \boldsymbol{C}_c = \boldsymbol{C}\, \boldsymbol{T}_c^{-1} = \begin{bmatrix} \boldsymbol{0}\ \boldsymbol{I} \end{bmatrix}.$$

After the coordinate transformation, the state-space representation of the system (53) is given by

$$\begin{bmatrix} \dot{\boldsymbol{x}}_1 \\ \dot{\boldsymbol{y}}_m \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{y}_m \end{bmatrix} + \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix} \boldsymbol{u}. \tag{55}$$

Then, the ansatz for the sliding mode observer has the form

$$\underbrace{\begin{bmatrix} \dot{\hat{\boldsymbol{x}}}_1 \\ \dot{\hat{\boldsymbol{y}}}_m \end{bmatrix}}_{\dot{\hat{\boldsymbol{x}}}_c} = \underbrace{\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}}_{\boldsymbol{A}_c} \underbrace{\begin{bmatrix} \hat{\boldsymbol{x}}_1 \\ \hat{\boldsymbol{y}}_m \end{bmatrix}}_{\hat{\boldsymbol{x}}_c} + \underbrace{\begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix}}_{\boldsymbol{B}_c} \boldsymbol{u} - \underbrace{\begin{bmatrix} \boldsymbol{G}_1 \\ \boldsymbol{G}_2 \end{bmatrix}}_{\boldsymbol{G}_c} (\hat{\boldsymbol{y}}_m - \boldsymbol{y}_m) + \underbrace{\begin{bmatrix} \boldsymbol{L} \\ -\boldsymbol{I} \end{bmatrix}}_{\boldsymbol{L}_c} \boldsymbol{v}, \tag{56}$$

where $\hat{\boldsymbol{x}}_1$ and $\hat{\boldsymbol{y}}_m$ represent the state estimates. $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ denote Luenberger-type gain matrices, whereas $\boldsymbol{L} \in \mathbb{R}^{4\times 2}$ denotes a feedback gain matrix, and $\boldsymbol{I} \in \mathbb{R}^{2\times 2}$ the identity matrix. The discontinuous switching part is defined by the vector $\boldsymbol{v}$ according to

$$\boldsymbol{v} = \underbrace{\begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}}_{\boldsymbol{M}_c} \underbrace{\begin{bmatrix} \operatorname{sign}(\Delta\hat{p} - \Delta p) \\ \operatorname{sign}(\hat{\omega}_M - \omega_M) \end{bmatrix}}_{\mathbf{sign}(\cdot)} = \begin{bmatrix} M_1 \operatorname{sign}(\Delta\hat{p} - \Delta p) \\ M_2 \operatorname{sign}(\hat{\omega}_M - \omega_M) \end{bmatrix}, \tag{57}$$

where $M_i, i \in \{1, 2\}$, denote the positive constant gains. Considering the definitions $\boldsymbol{e}_1 = \hat{\boldsymbol{x}}_1 - \boldsymbol{x}$ and $\boldsymbol{e}_{y_m} = \hat{\boldsymbol{y}}_m - \boldsymbol{y}_m$, the estimation error dynamics can be derived from (55) and (56) as

$$\begin{bmatrix} \dot{\boldsymbol{e}}_1 \\ \dot{\boldsymbol{e}}_{y_m} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_{y_m} \end{bmatrix} - \begin{bmatrix} \boldsymbol{G}_1 \\ \boldsymbol{G}_2 \end{bmatrix} \boldsymbol{e}_{y_m} + \begin{bmatrix} \boldsymbol{L} \\ -\boldsymbol{I} \end{bmatrix} \boldsymbol{v}. \tag{58}$$

Introducing a new error variable $\bar{\boldsymbol{e}}_1 = \boldsymbol{e}_1 + \boldsymbol{L}\, \boldsymbol{e}_{y_m}$, the resulting error dynamics w.r.t a new state variable $\bar{\boldsymbol{e}}_1$ and $\boldsymbol{e}_{y_m}$ becomes

$$\begin{bmatrix} \dot{\bar{\boldsymbol{e}}}_1 \\ \dot{\boldsymbol{e}}_{y_m} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{A}}_{11} & \bar{\boldsymbol{A}}_{12} \\ \boldsymbol{A}_{21} & \bar{\boldsymbol{A}}_{22} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{e}}_1 \\ \boldsymbol{e}_{y_m} \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} \\ -\boldsymbol{I} \end{bmatrix} \boldsymbol{v}, \tag{59}$$

where the submatrices are given by

$$\begin{aligned} \bar{\boldsymbol{A}}_{11} &= \boldsymbol{A}_{11} + \boldsymbol{L}\, \boldsymbol{A}_{21}, \\ \bar{\boldsymbol{A}}_{12} &= \boldsymbol{A}_{12} - \bar{\boldsymbol{A}}_{11}\, \boldsymbol{L} - \boldsymbol{G}_1 + \boldsymbol{L}\, (\boldsymbol{A}_{22} - \boldsymbol{G}_2), \\ \bar{\boldsymbol{A}}_{22} &= \boldsymbol{A}_{22} - \boldsymbol{G}_2 - \boldsymbol{A}_{21}\, \boldsymbol{L}. \end{aligned}$$

$\bar{\boldsymbol{A}}_{12} = \boldsymbol{0}$ can be achieved by proper choice of the gain matrix $\boldsymbol{G}_1$. In the case of $\boldsymbol{v} = \boldsymbol{0}$, asymptotic stability of the error dynamic system (59) can be obtained by choosing the gain matrices $\boldsymbol{L}$ and $\boldsymbol{G}_2$ according to

$$\boldsymbol{A}_{11} + \boldsymbol{L}\, \boldsymbol{A}_{21} = \bar{\boldsymbol{A}}^*_{11}, \tag{60}$$

$$\boldsymbol{A}_{22} - \boldsymbol{G}_2 - \boldsymbol{A}_{21}\, \boldsymbol{L} = \bar{\boldsymbol{A}}^*_{22}, \tag{61}$$

where $\bar{\boldsymbol{A}}^*_{11}$ and $\bar{\boldsymbol{A}}^*_{22}$ denote asymptotically stable matrices with the following characteristic polynomials

$$\begin{aligned} \det(s\, \boldsymbol{I} - \bar{\boldsymbol{A}}^*_{11}) &\stackrel{!}{=} (s + s_{B1})(s + s_{B2})(s + s_{B3})(s + s_{B4}), \\ \det(s\, \boldsymbol{I} - \bar{\boldsymbol{A}}^*_{22}) &\stackrel{!}{=} (s + s_{B5})(s + s_{B6}). \end{aligned}$$

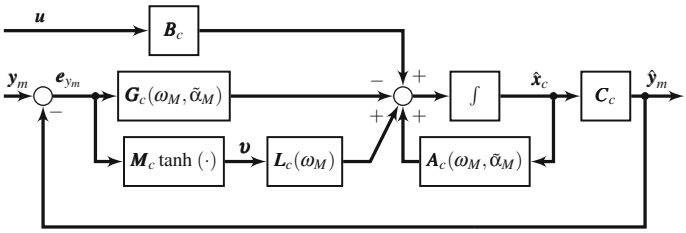


Fig. 5 Block diagram of the gain-scheduled sliding mode observer

In the original Utkin observer, the switching action $\boldsymbol{v}$ guarantees a finite-time convergence on the measurable states, and stabilises the observer error dynamics, cf. [9]. The additional Luenberger gain matrix $\boldsymbol{G}_c$ in the modified Utkin observer yields asymptotic stability of the observer errors and offers the potential to provide robustness against certain classes of model uncertainty by the switching part $\boldsymbol{v}$. Aiming at an avoidance of chattering—caused by the discontinuous switching part, the sign function vector **sign**$(\cdot)$ is replaced in the implementation by the tanh function vector **tanh**$(\cdot)$. As a consequence, the ideal sliding mode becomes a real sliding mode within a resulting boundary layer. The block diagram of the implementation of the sliding mode observer is depicted in Fig. 5.

5 Simulation and Experimental Results

In this section, the proposed sliding mode control in combination with the gain-scheduled sliding mode observer is investigated by both simulations and experimental evaluations. To guarantee realistic simulation results, the system model is extended with measurement noise regarding the pressure sensors and with quantisation errors of the encoders. After thorough simulation studies, the control algorithm can be directly implemented on the corresponding test bench. Nevertheless, the design parameters of the model-based controller have been slightly adjusted at the test bench in order to further improve the tracking performance regarding the controlled variables. To point out the superior tracking properties of the proposed observer-based sliding mode control, an alternative solution given by a flatness-based tracking control in combination with a nonlinear reduced-order observer is considered as a reference for a comparison.

In order to numerically evaluate the performance of the individual control approach, two criteria are introduced—the maximum absolute tracking error $e_{\max}$ as well as the root-mean-square (RMS) error e_{RMS}, which are defined as

$$e_{\max} = \max(|e(k)|), \; e_{RMS} = \sqrt{\frac{1}{N} \sum e^2(k)}, k \in \{1, 2, \ldots, N\}.$$

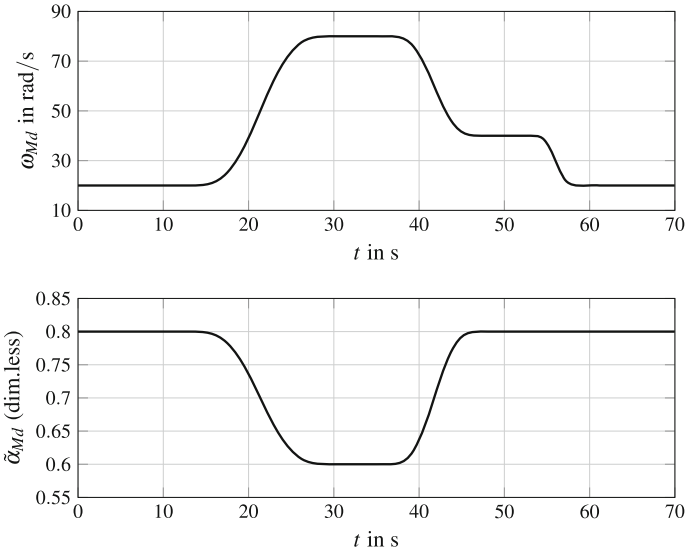


Fig. 6 Desired trajectories for the controlled variables—the motor angular velocity ω_{Md} (*top*) and the normalised motor bent-axis angle $\tilde{\alpha}_{Md}$ (*bottom*)

Figure 6 depicts the desired trajectories for the motor angular velocity ω_{Md} and the normalised bent-axis angle $\tilde{\alpha}_{Md}$, respectively. The synchronised desired trajectories are designed based on smooth, several times continuously differentiable polynomials and comprise a sequence of motions. It is important to note that—by exploiting the flatness property of the system at trajectory planning—any saturation of the tilt angles for the pump and the motor can be avoided.

5.1 Simulation Results

In this section, meaningful simulation results are presented: in the first scenario, the flatness-based controller (FBC) is employed together with either the proposed sliding mode observer (SMO) or the nonlinear reduced-order observer (NROO), see Figs. 7, 8 and 9. In the second scenario—to improve the tracking performance regarding the motor angular velocity ω_M—the sliding mode control (SMC) is used. Again, a comparison between the SMC and the FBC in combination with the gain-scheduled SMO is carried out. It becomes obvious that a superior control performance is achievable with the robust control approach proposed in this chapter, see Figs. 10 and 11. For the simulation studies, disturbance models are introduced: the unknown disturbance torque is modelled as

$$\tau_U = 0.1\, J_V\, \dot{\omega}_{Md} + 7 \tanh\left(\frac{\omega_{Md}}{0.1}\right),$$

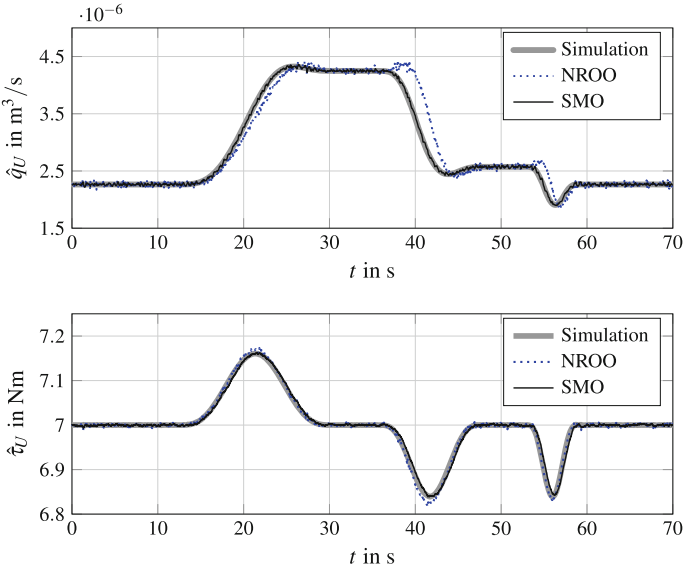


Fig. 7 Simulation results regarding the disturbance estimation: comparison between the nonlinear reduced-order observer (NROO) and the sliding mode observer (SMO)

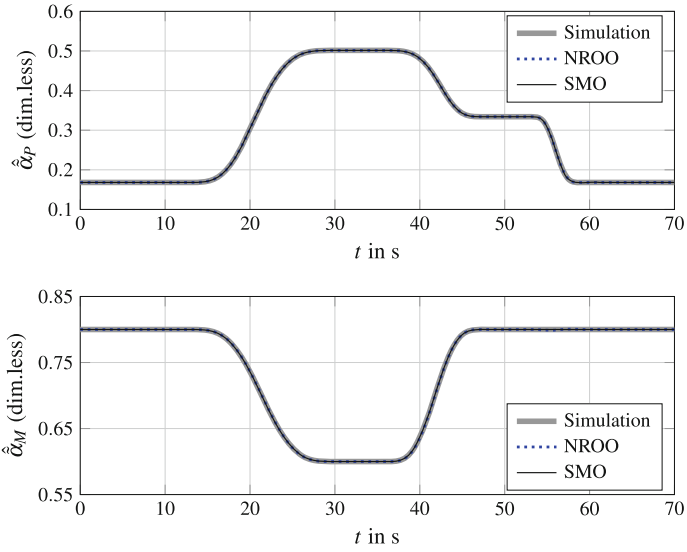


Fig. 8 Simulation results regarding the state estimation: comparison between the nonlinear reduced-order observer (NROO) and the sliding mode observer (SMO)

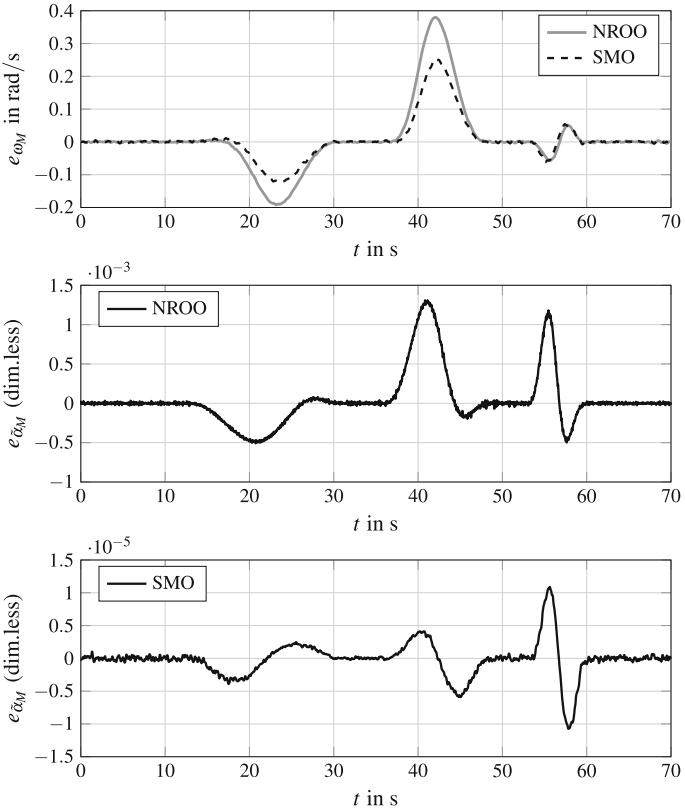


Fig. 9 Simulation results for the tracking errors of the controlled variables using flatness-based control: comparison between the nonlinear reduced-order observer (NROO) and the sliding mode observer (SMO)

which presents a static nonlinear friction torque as well as a load torque caused by 10 % mass of the rotation part at the motor side. The leakage volume flow q_U is assumed to be proportional to the pressure difference Δp which can be stated as

$$q_U = k_l \, \Delta p.$$

Here, $k_l > 0$ denotes the constant leakage coefficient of the system. Moreover, to evaluate the disturbance estimation of the proposed observers, parameter uncertainty is not explicitly introduced in the simulation study. However, the residual of the disturbance compensation—especially in transient phases—can be considered as system uncertainty to be counteracted by the sliding mode control.

Figure 7 depicts a comparison of the simulation results of the estimated disturbances obtained from the two alternative observers. It becomes obvious that the SMO leads to a significantly better estimation of the leakage volume flow q_U, where the

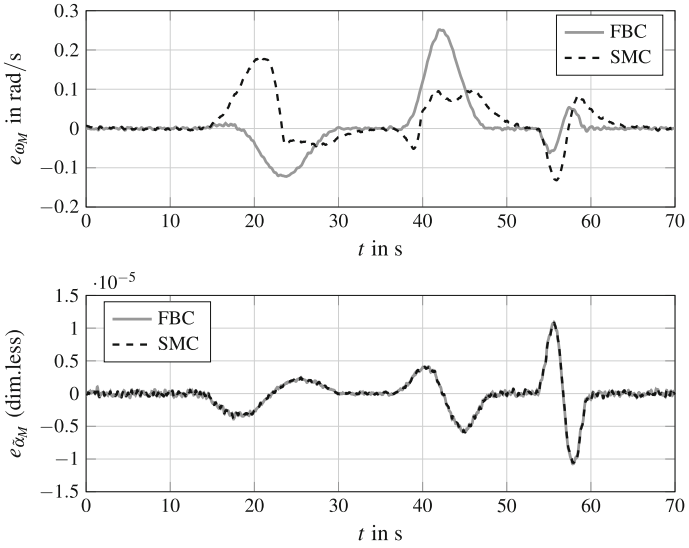


Fig. 10 Simulation results for the tracking errors of the controlled variables using alternative controllers: FBC—flatness-based control; SMC—sliding mode control

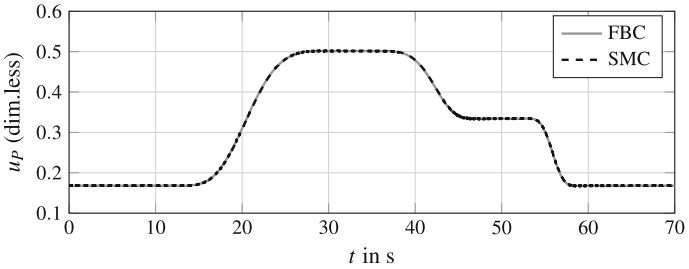


Fig. 11 Simulation results for the control inputs of the alternative controllers: FBC—flatness-based control; SMC—sliding mode control

estimation obtained from NROO has a certain deviation in comparison with the simulated signal. Both observers, however, are able to estimate the disturbance torque τ_U with a high accuracy as shown in Fig. 7. The SMO provides a slightly smaller maximum estimation error of 0.0236 Nm than the NROO with 0.0244 Nm. However, the NROO possesses a relatively smaller RMS value, cf. Table 1. A comparison of the estimation results for the immeasurable states—the normalised tilt angles $\tilde{\alpha}_P$ of the pump and $\tilde{\alpha}_M$ of the motor—can be found in Fig. 8. It can be concluded that both observers offer excellent estimation results. An improvement regarding the state estimation can be obtained by using the SMO, especially for the normalised bent-axis angle $\tilde{\alpha}_M$, cf. Table 1. The states and disturbances provided by the observers can be, hence, directly applied to the controllers. The tracking performance of the FBC with

Table 1 Numerical evaluation: maximum absolute errors and RMS errors for the combined state and disturbance estimation

	ω_M (rad/s)		$\tilde{\alpha}_M$		$\tilde{\alpha}_P$	
	$e_{\max}$	e_{RMS}	$e_{\max}$	e_{RMS}	$e_{\max}$	e_{RMS}
NROO	0.3802	0.0989	1.3×10^{-3}	0.3434×10^{-3}	0.0038	0.9844×10^{-3}
SMO	0.2518	0.0642	0.1090×10^{-4}	0.2637×10^{-5}	0.0022	0.5662×10^{-3}
	q_U (m^3/s)		τ_U (Nm)			
	$e_{\max}$	e_{RMS}	$e_{\max}$	e_{RMS}		
NROO	0.8685×10^{-6}	0.1902×10^{-6}	0.0244	0.0052		
SMO	0.0680×10^{-6}	0.0178×10^{-6}	0.0236	0.0057		

different observers is illustrated in Fig. 9, which depicts the tracking errors of both controlled variables. The simulation results demonstrate that the FBC combined with the SMO outperform the FBC combined with the NROO. By using the SMO, the tracking error of the normalised bent-axis angle $\tilde{\alpha}_M$ is controlled with a scale of 10^{-5}, in comparison to 10^{-3} by using the NROO. The numerical evaluation according to maximum absolute tracking errors $e_{\max}$ as well as root-mean-square errors e_{RMS} of the simulation studies can be found in Table 1.

Figure 10 depicts the tracking errors of the controlled variables—a comparison between FBC and SMC in combination with the SMO. The tracking performance of the motor angular velocity ω_M is further improved by applying SMC. The same conclusion is numerically indicated in Table 2. The control inputs of FBC and SMC can be found in Fig. 11. The chattering of the SMC input has been well dealt with by using the $\tanh(s/\eta)$ function instead of $\text{sign}(s)$.

The thorough simulation study demonstrates that the SMO provides a better estimation of the system disturbances, cf. Fig. 7, and a significant improvement on the estimation of the normalised bent-axis angle $\tilde{\alpha}_M$, cf. Fig. 9. Moreover, a better tracking performance of the motor angular velocity ω_M can be obtained by using SMC instead of FBC. Hence, the proposed control structure—sliding mode control in combination with the sliding mode observer—guarantees a superior result in comparison

Table 2 Numerical evaluation: maximum absolute errors and RMS errors for the controlled variables

	ω_M (rad/s)		$\tilde{\alpha}_M$	
	$e_{\max}$	e_{RMS}	$e_{\max}$	e_{RMS}
FBC + SMO	0.2518	0.0642	0.1090×10^{-4}	0.2637×10^{-5}
SMC + SMO	0.1822	0.0568	0.1083×10^{-4}	0.2623×10^{-5}

with flatness-based control in combination with a nonlinear reduced-order observer. In the following section, an experimental validation and comparison between two alternative control structures will be carried out and discussed.

5.2 Experimental Results

The test bench consists of a hydrostatic transmission system, CANopen devices, a real-time target PC and a host PC with Labview software. The proposed control approaches and observers are implemented using Labview mathscript function in the host PC and complied to the target PC for the real-time test. For the experiments, a sampling time of $T_s = 20\,\text{ms}$ is chosen. The control parameters for each subsystem are listed in Table 3. Both control structures are evaluated in two cases: without a load torque or with a constant 30 Nm load torque generated by the electric motor. For the comparison study, the same characteristic polynomials are chosen to determine the coefficients of the stabilising control inputs for FBC and SMC as well as the eigenvalues to determine the observer gains for the NROO and SMO, respectively.

5.2.1 Evaluation Without a Load Torque

Figure 12 depicts the comparison of the disturbance estimation between different control structures. It can be seen that the SMO yields larger estimations both for the leakage volume flow q_U and the disturbance torque τ_U. The comparison of state estimation can be found in Fig. 13, where both observers give similar estimation results. The tracking performance of the alternative control structures is illustrated in Fig. 14, which indicates that both control structures can guarantee an outstanding tracking performance for both controlled variables. It is obvious that the SMC in combination with the SMO leads to a superior tracking compared with FBC with the NROO. Table 4 summarises the numerical evaluation of two different control

Table 3 Control parameters for the experimental validation

FBC ($\tilde{\alpha}_M$)	$k_{\alpha 0} = 4,\ k_{\alpha I} = 2$
FBC (ω_M)	$k_{\omega_M,0} = 343,\ k_{\omega_M,1} = 147,\ k_{\omega_M,2} = 21$
SMO (ω_M)	$c_1 = 49,\ c_2 = 14,\ \lambda = 7,\ \eta = 0.02$
	$\gamma_1 = 0.05,\ \gamma_2 = 0.01,\ \gamma_3 = 0.001,\ \gamma_4 = 0.001,\ \gamma_5 = 0.01$
NROO	$s_{B1} = 10,\ s_{B2} = 4.5,\ s_{B3} = 10,\ s_{B4} = 4.5$
SMO	$s_{B1} = 10,\ s_{B2} = 4.5,\ s_{B3} = 10,\ s_{B4} = 4.5,\ s_{B5} = 5,\ s_{B6} = 10$
	$M_1 = 0.01,\ M_2 = 0.01,\ \eta = 0.02$

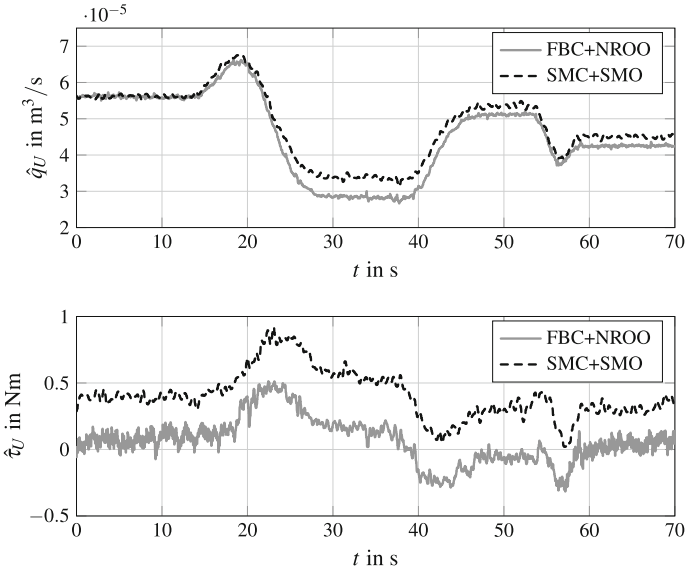


Fig. 12 Comparison of disturbance estimation results without a load torque

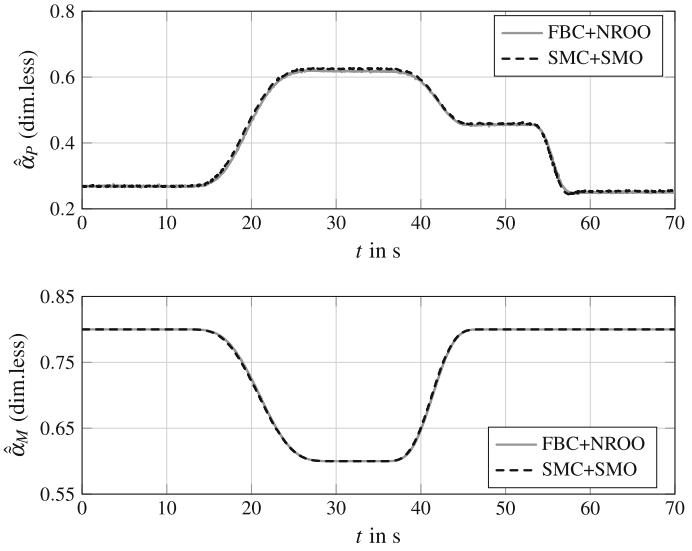


Fig. 13 Comparison of state estimation results without a load torque

structures. Figure 15 depicts the control inputs u_P and u_M. The control inputs show similar time behaviour as the corresponding tilt angles depicted in Fig. 13 where the dynamic relationships are governed by (8) and (9). Moreover, chattering has been well dealt with by using the tanh(·) function.

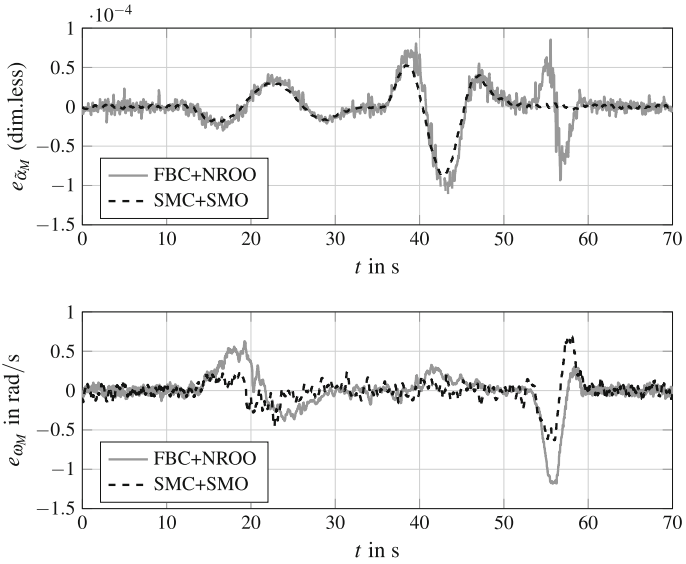


Fig. 14 Comparison of the tracking errors for the controlled variables without a load torque

Table 4 Numerical evaluation: maximum absolute errors and RMS errors for the controlled variables

	ω_M (rad/s)		$\tilde{\alpha}_M$	
	$e_{\max}$	e_{RMS}	$e_{\max}$	e_{RMS}
FBC + NROO	1.1817	0.2408	0.1195×10^{-3}	0.2898×10^{-4}
SMC + SMO	0.6935	0.1577	0.0855×10^{-3}	0.2114×10^{-4}

5.2.2 Evaluation with a Load Torque of 30 Nm

In this case, the electric motor on the load side generates a constant disturbance torque of 30 Nm acting on the hydraulic motor. It becomes obvious in Fig. 16 that the SMO estimates a slightly larger disturbance torque $\hat{\tau}_U$, whereas the leakage volume flow $\hat{q}_U$ only has relatively larger estimates when the motor rotates at a high rotational speed. Figure 17 shows a similar estimation of the immeasurable states between SMO and NROO.

According to the tracking performance illustrated in Fig. 18 as well as the numerical results listed in Table 5, the same conclusion can be obtained as that in the case without a load torque, except that the tracking performance of the normalised motor bent-axis angle is almost identical in case with a load torque of 30 Nm. Moreover, the control inputs of both controllers can be found in Fig. 19.

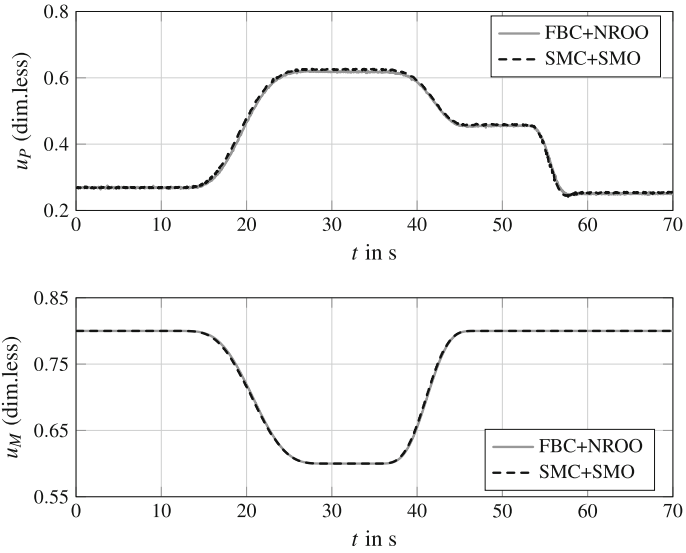


Fig. 15 Comparison regarding the control inputs without a load torque

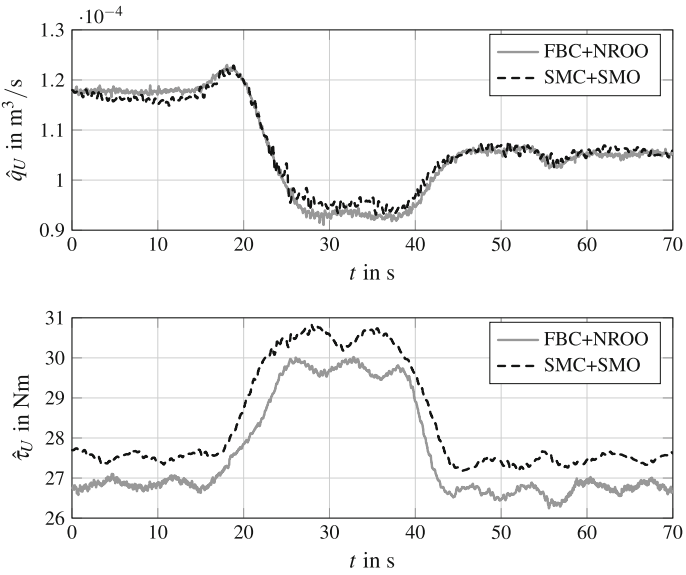


Fig. 16 Comparison of disturbance estimation results with a load torque of 30 Nm

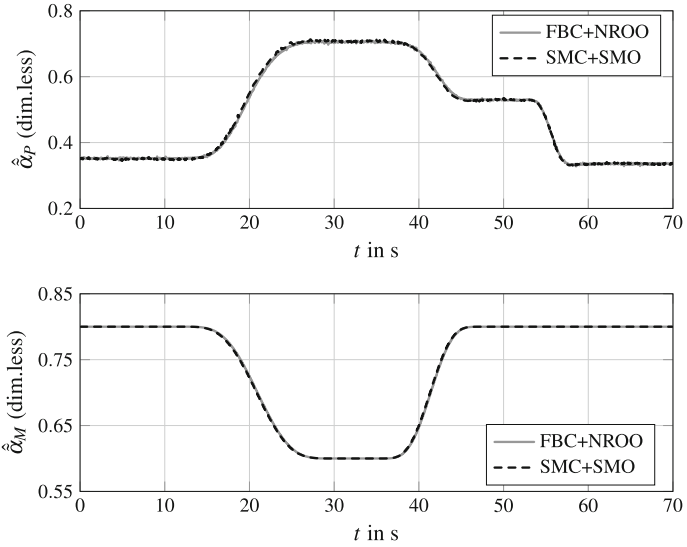


Fig. 17 Comparison of state estimation results with a load torque of 30 Nm

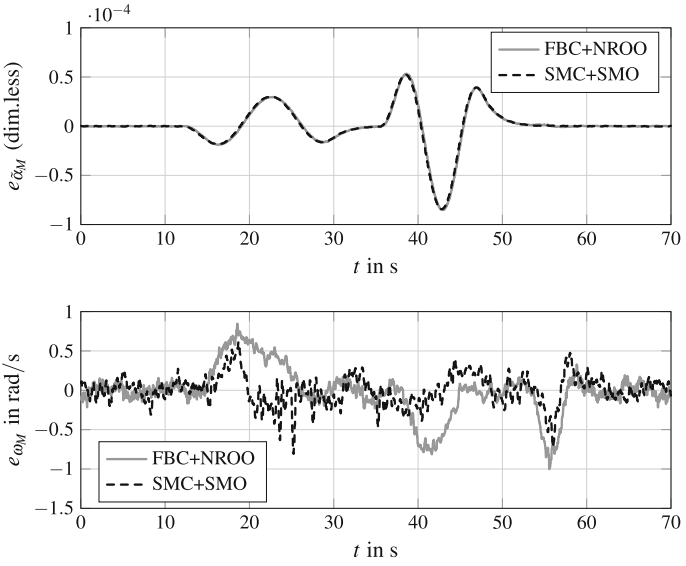


Fig. 18 Comparison of the tracking errors for the controlled variables with a load torque of 30 Nm

5.2.3 Evaluation with Fast Trajectories Without a Load Torque

To further evaluate the robust performance of the control structures experimentally, fast desired trajectories of the controlled variables are introduced as depicted in

Table 5 Numerical evaluation: maximum absolute errors and RMS errors for the controlled variables

	ω_M (rad/s)		$\tilde{\alpha}_M$	
	$e_{\max}$	e_{RMS}	$e_{\max}$	e_{RMS}
FBC + NROO	1.0023	0.2989	0.8518×10^{-4}	0.2106×10^{-4}
SMC + SMO	0.8081	0.1896	0.8436×10^{-4}	0.2096×10^{-4}

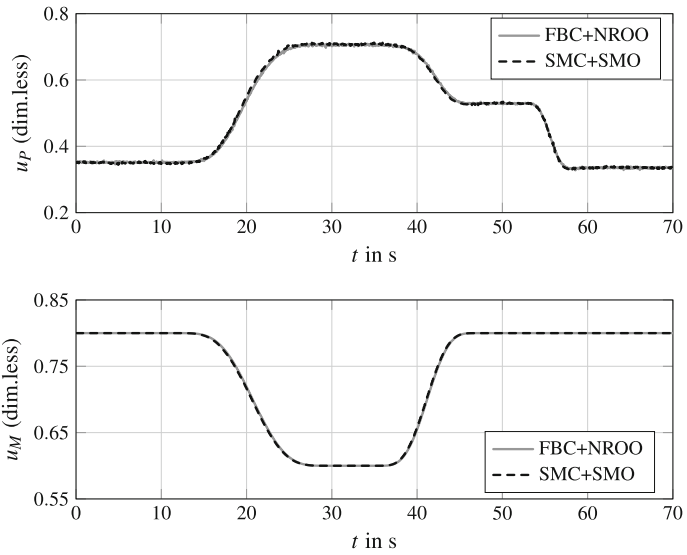


Fig. 19 Comparison regarding the control inputs with a load torque of 30 Nm

Fig. 20. They are characterised by maximum accelerations of $\dot{\omega}_{Md} = 18.9\,\mathrm{rad/s^2}$, which are much higher than those of the slow trajectory with maximum accelerations of $\dot{\omega}_{Md} = 8\,\mathrm{rad/s^2}$.

The estimation results for the system disturbances as well as the immeasurable states are depicted in Figs. 21 and 22, which indicate that both observers provide accurate estimates. The corresponding tracking errors can be found in Fig. 23. Using the fast trajectory, the SMC in combination with the SMO leads to a relatively better tracking of the motor angular velocity ω_M and a similar performance for the normalised bent-axis angle $\tilde{\alpha}_M$ in comparison with the FBC in combination with the NROO.

The numerical values for the error measures are stated in Table 6. Figure 24 depicts the control inputs of both controllers during trajectory tracking.

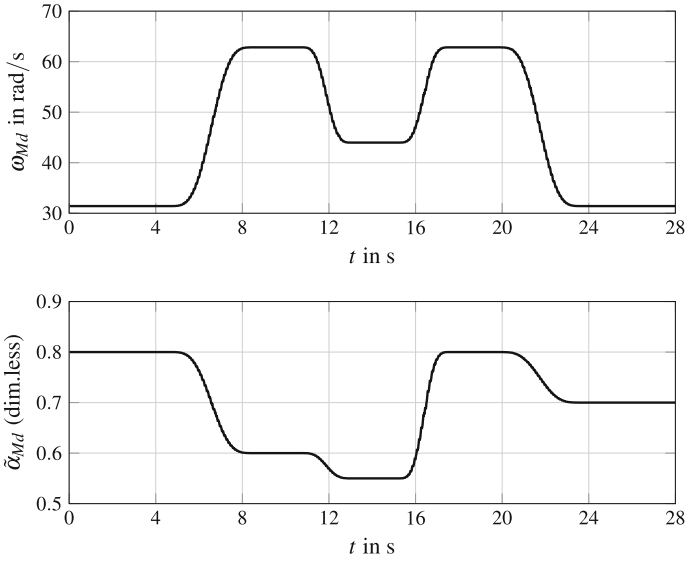


Fig. 20 Fast desired trajectories for the controlled variables—the motor angular velocity ω_{Md} (*top*) and the normalised motor bent-axis angle $\tilde{\alpha}_{Md}$ (*bottom*)

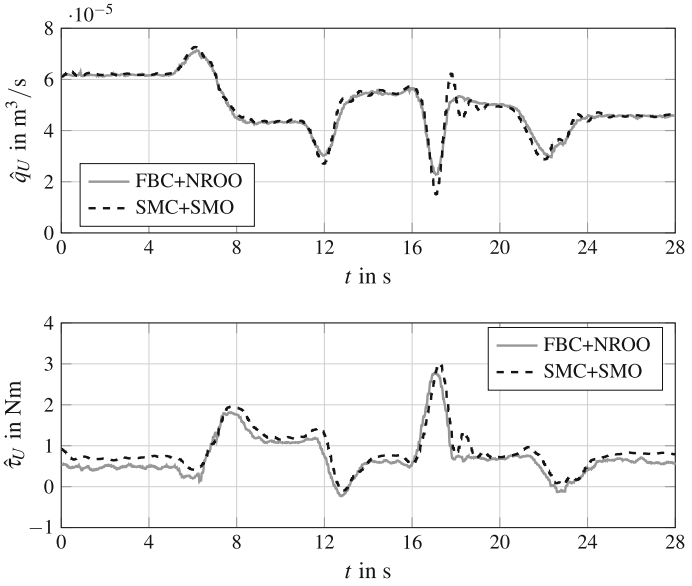


Fig. 21 Comparison of disturbance estimation results with fast trajectories

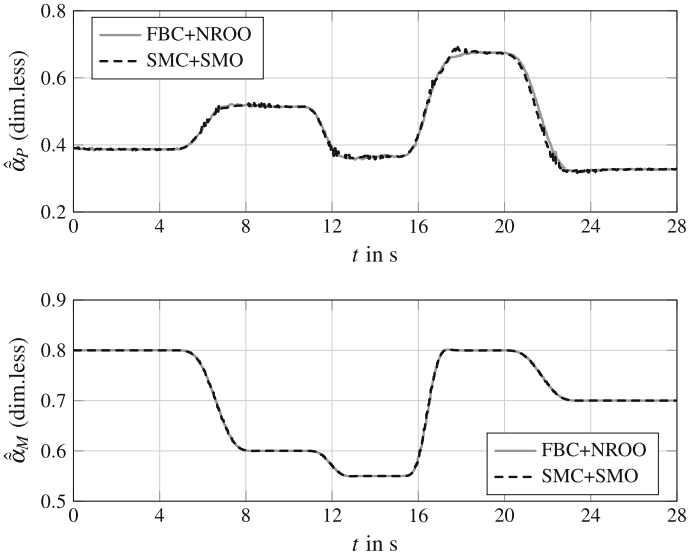


Fig. 22 Comparison of state estimation results with fast trajectories

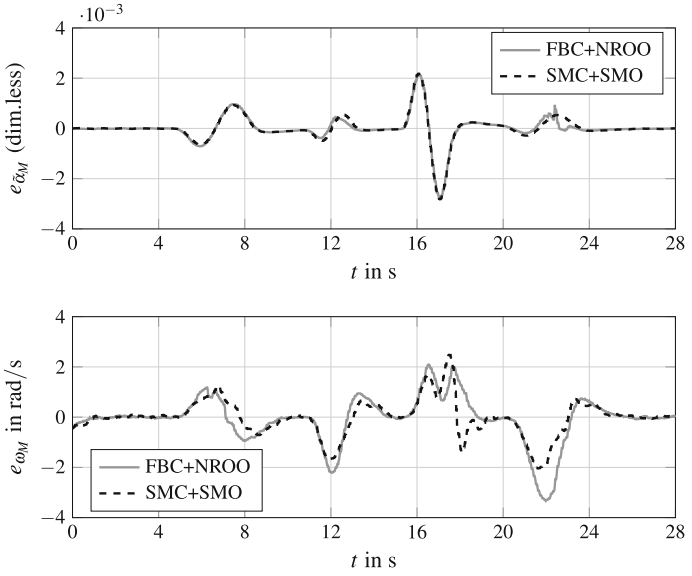


Fig. 23 Comparison of tracking error results for the controlled variables with fast trajectories

Table 6 Numerical evaluation: maximum absolute errors and RMS errors for the controlled variables

	ω_M (rad/s)		$\tilde{\alpha}_M$	
	$e_{\max}$	e_{RMS}	$e_{\max}$	e_{RMS}
FBC + NROO	3.3263	0.9168	0.0027	0.5371×10^{-3}
SMC + SMO	2.4796	0.7104	0.0028	0.5512×10^{-3}

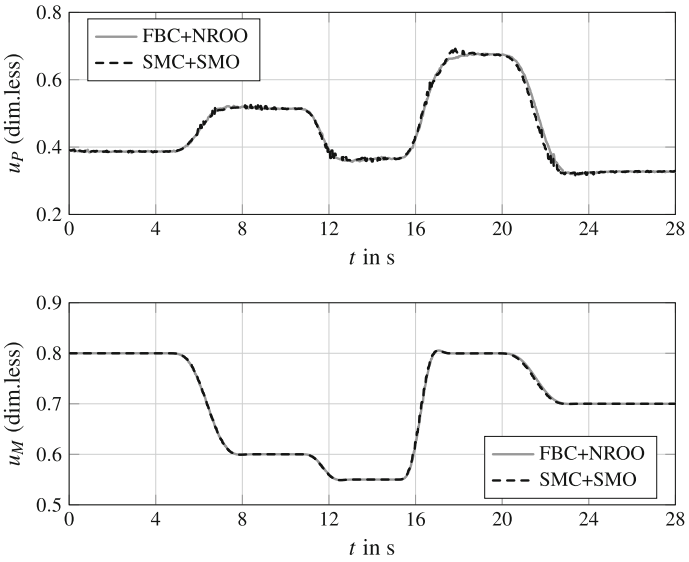


Fig. 24 Comparison regarding the control inputs with fast trajectories

6 Conclusion

In this contribution, a model-based decentralised control structure is proposed for the tracking control of a hydrostatic transmission. A control-oriented modelling of the hydrostatic transmission results in a dynamic system represented by four first-order differential equations, which is structured in two sub-models for the decentralised control design. The control structure consists of a flatness-based control of a first-order model for the normalised bent-axis angle $\tilde{\alpha}_M$ and a sliding mode control of a third-order state-space representation for the motor angular velocity ω_M. To deal with the immeasurable states and unknown disturbances, a gain-scheduled sliding mode observer is designed using extended linearisation techniques. The complete control structure is thoroughly investigated by means of simulations regarding asymptotic stability and the estimation quality w.r.t. simulated disturbances. Furthermore, the control performance is analysed and compared with an alternative reference control structure—a flatness-based tracking control in combination with a nonlinear

reduced-order observer. The experimental validation shows clearly that the proposed sliding mode control combined with a gain-scheduled sliding mode observer outperforms the alternative reference solution. By replacing the nonlinear reduced-order observer with the sliding mode observer, an improved tracking performance is achievable for the flatness-based control as well.

References

1. Aschemann H, Sun H (2013) Decentralised flatness-based control of a hydrostatic drive train subject to actuator uncertainty and disturbances. In: Proceedings of the 18th international conference on methods and models in automation and robotics (MMAR), Miedzyzdroje, Poland
2. Aschemann H, Ritzke J, Schulte H (2009) Model-based nonlinear trajectory control of a drive chain with hydrostatic transmission. In: Proceedings of the 14th international conference on methods and models in automation and robotics (MMAR), Miedzyzdroje, Poland
3. Aschemann H, Meinlschmidt T, Sun H (2014) An experimental study on decentralised backstepping approaches for a hydrostatic drive train with unknown disturbances. In: Proceedings of the American control conference (ACC), Portland, USA
4. Aschemann H, Sun H, Meinlschmidt T (2014) An experimental study of extended linearisation approaches for a hydrostatic transmission with unknown disturbances. In: Proceedings of the European control conference (ECC), Strasbourg, France
5. Diepeveen N, Laguna A (2011) Dynamics modelling of fluid power transmissions for wind turbine. In: Proceedings of the EWEA offshore, Amsterdam, Netherland
6. Dolan B, Aschemann H (2012) Control of a wind turbine with a hydrostatic transmission - an extended linearisation approach. In: Proceedings of the 17th international conference on methods and models in automation and robotics (MMAR), Miedzyzdroje, Poland
7. Edwards C, Spurgeon S (1998) Sliding mode control: theory and application. Taylor & Francis, London
8. Friedland B (1996) Advanced control system design. Prentice-Hall, New Jersey
9. Haskara İ (1996) Sliding mode controllers and observers. Master Thesis, The Ohio State University, USA
10. Hoang T, Kyoung K (2013) Velocity control of a secondary controlled closed-loop hydrostatic transmission system using an adaptive fuzzy sliding mode controller. J Mech Sci Technol 27:875–884
11. Humadi A, Hussein A (2011) Improvement of a hydrostatic transmission control system performance using radial basis neural network. J Eng 17:577–583
12. Jelali M, Kroll A (2003) Hydraulic servo-systems: modelling, identification and control. Springer, London
13. Jiang J, Liu H, Celestine O (2006) Nonlinear H_∞ control in frequency domain of hydrostatic transmission with secondary regulation via GFRF. In: Proceedings of the 6th world congress on intelligent control and automation, vol 2(3), pp 6416–6420
14. Kansala K, Hasemann J (1994) An embedded distributed fuzzy logic traction control system for vehicles with hydrostatic power transmission. In: Proceedings of the 7th Mediterranean electrotechnical conference, Antalya, Turkey
15. Liu H, Jiang J, Celestine O (2006) Nonlinear control via exact linearisation for hydrostatic transmission with secondary regulation. In: Proceedings of the 1st international symposium on system and control in aerospace and astronautics, Harbin, China
16. Manring N, Johnson R (1996) Modelling and designing a variable-displacement open-loop pump. J Dyn Syst Meas Control 118:267–271
17. Manring N, Luecke G (1998) Modelling and designing a hydrostatic transmission with a fixed displacement motor. J Dyn Syst Meas Control 120:45–49

18. Mutschler S (2008) Economic evaluation of hydrostatic drive train concepts for mobile machinery. In: 6th Internationales Fluidtechnisches Kolloquium, Dresden, Germany
19. Nawrocka A, Kwashiewski J (2008) Predictive neural network controller for hydrostatic transmission control. Mechanics 27:62–65
20. Ritzke J, Aschemann H (2011) Design and experimental validation of nonlinear trajectory control of a drive chain with hydrostatic transmission. In: Proceedings of the 12th Scandinavian international conference on fluid power (SICFP), Tampere, Finland
21. Stoll S, Kliffken M, Behm M, Wang X (2007) Regelungskonzepte für hydrostatische Antriebe in mobilen Arbeitsmaschinen (in German). at - Automatisierungstechnik 55:48–57
22. Sun H, Aschemann H (2013) Adaptive inverse dynamics control for a hydrostatic transmission with actuator uncertainties. In: Proceedings of the IEEE international conference on mechatronics (ICM), Tampere, Finland
23. Sun H, Aschemann H (2013) Robust inverse dynamics control for a hydrostatic transmission with actuator uncertainties. In: Proceedings of the 6th IFAC symposium on mechatronic systems, HangZhou, China
24. Sun H, Aschemann H (2013) Sliding-mode control of a hydrostatic drive train with uncertain actuator dynamics. In: Proceedings of the European control conference (ECC), Zurich, Switzerland
25. Sun H, Meinlschmidt T, Aschemann H (2014) Comparison of two nonlinear model predictive control strategies with observer-based disturbance compensation for a hydrostatic transmission. In: Proceedings of the 19th international conference on methods and models in automation and robotics (MMAR), Miedzyzdroje, Poland
26. Sun H, Meinlschmidt T, Aschemann H (2014) Optimal tracking control with observer-based disturbance compensation for a hydrostatic transmission. In: Proceedings of the IEEE multiconference on systems and control (MSC), Antibes, France
27. Zhang H, Li H, Ma B (2012) Control and parameters matching of straight running for high speed tracked vehicle with hydrostatic drive. In: Proceedings of the IEEE international conference on mechatronics and automation, ChengDu, China

Sliding Mode Observation with Iterative Parameter Adaption for Fast-Switching Solenoid Valves

Tristan Braun and Johannes Reuter

Abstract Control of the armature motion of fast-switching solenoid valves is highly desired to reduce noise emission and wear of material. For feedback control, information of the current position and velocity of the armature are necessary. In mass production applications, however, position sensors are unavailable due to cost and fabrication reasons. Thus, position estimation by measuring merely electrical quantities is a key enabler for advanced control, and, hence, for efficient and robust operation of digital valves in advanced hydraulic applications. The work presented here addresses the problem of state estimation, i.e., position and velocity of the armature, by sole use of electrical measurements. The considered devices typically exhibit nonlinear and very fast dynamics, which makes observer design a challenging task. In view of the presence of parameter uncertainty and possible modeling inaccuracy, the robustness properties of sliding mode observation techniques are deployed here. The focus is on error convergence in the presence of several sources for modeling uncertainty and inaccuracy. Furthermore, the cyclic operation of switching solenoids is exploited to iteratively correct a critical parameter by taking into account the norm of the observation error of past switching cycles of the process. A thorough discussion on real-world experimental results highlights the usefulness of the proposed state observation approach.

1 Introduction

Switching solenoid valves or digital valves are used in a variety of automotive and hydraulic applications. In digital hydraulics, where principles from power electronics and digital circuits are transferred to hydraulics in order to increase efficiency and diminish costs, solenoid valves are employed to operate as digital (on–off)

T. Braun (✉) · J. Reuter
Institute of System Dynamics, Constance University of Applied Sciences, Brauneggerstr. 55, 78462 Konstanz, Germany
e-mail: tbraun@htwg-konstanz.de

J. Reuter
e-mail: jreuter@htwg-konstanz.de

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*, Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_7

fast-switching valves [19, 20]. In digital hydraulic applications, such as the hydraulic switching converter, digital valves operate in a pulse-width modulated (PWM) mode to control the volume flow rate of a fluid, comparable with power semiconductors in DC–DC converters [19]. Control of the spool trajectory of those devices is highly desired to reduce noise emission and wear of material, or to avoid impacts of the spool at all via soft-landing control [6, 11]. Motion planning of the spool trajectory regarding shape and travel time requires enhanced feedback-control strategies. Feed-forward control is unsatisfactory due to nonlinear effects and uncertain parameters, and, therefore, infeasible with regard to robustness and precision. For full state feedback, information of the position and velocity of the spool is necessary. In mass production applications, position sensors are normally not desired due to cost and fabrication reasons. Thus, position estimation based on electrical quantities that are easy to measure is a key enabler for advanced control, and, therefore, robust operation of digital valves in advanced applications. Solenoids are typically affected by nonlinear magnetization properties, such as saturation of material and magnetic hysteresis. In particular, modeling the effect of eddy currents is a key to represent the systems dynamic correctly. To cope with these effects in a suitable model is a formidable challenge, particularly if it should be as simple as possible for numerical efficiency and accessible for methodical observer design. Therefore, robust state estimation must be achieved in the presence of uncertain parameters such as eddy current resistance, modeling inaccuracy in electromagnetic characteristics as well as the influence of possibly unmodeled force terms.

Several approaches are addressed in the literature to solve the state estimation problem for solenoid valves, which can be divided into parameter estimation and observer design. In the case of parameter estimation the valve position is typically retrieved by the estimated position-dependent differential inductance, whereby the solenoid is excited by a PWM signal [13, 27]. This approach, however, is more favorable for relatively slow-moving actuators, where the convergence time is less restrictive. Observer-based position estimation for different kinds of actuators is an ongoing research activity, see [1], where an electro-pneumatic actuator is investigated. However, observer design for fast-switching solenoid valves, where no additional sensors are used, instead of voltage and current measurements, can rarely be found in the literature. In [21], a nonlinear observer based on extended linearization for an idealized dynamic model is considered; however, merely simulation results are shown. A sliding mode-based sensorless control scheme is investigated in [9] on a special-type electromagnetic actuator, based on a sophisticated model. Although experimental results are presented, poor information regarding the underlying estimation approach is given.

In this work, the problem of robust position estimation for fast-switching valves is tackled by the design of a sliding mode observer (SMO) for the nonlinear model, based on an equivalent output injection method (cf. [8, 18]). In this regard, the work presented in [4, 5] is improved and extended in various ways regarding performance and validation aspects as well as analysis of error convergence in the presence of modeling uncertainty and time-varying eddy current resistance.

The type of digital spool valve considered here consists of two symmetric coils, which are designed to control the spool by alternately energizing the coils for desired acceleration and deceleration of the spool. A cut of the considered dual-coil solenoid valve is depicted in Fig. 1. Since coils are alternately energized, it is beneficial to exploit the information of the spool velocity and position from the induced voltage in the non-energized coil.[1]

This contribution is structured as follows. The dynamic model of the solenoid valve is addressed in Sect. 2. In this respect, in Sect. 2.1, the electromagnetic subsystem is examined, and it is shown how the look-up tables for the magnetic characteristics can be obtained by a nonlinear reluctance network. Furthermore, a detailed time-varying model of the eddy current resistance is discussed in Sect. 2.2, which yields better estimation results compared to a constant parameter. The mechanical subsystem is briefly derived in Sect. 2.3. The corresponding observer-oriented model approach is pointed out in Sect. 2.4, where the system equations are constituted and system variables together with the observer output are defined. Some useful remarks conclude this section. The nonlinear SMO approach is discussed in Sect. 3. The error convergence in the presence of several sources of modeling uncertainty and inaccuracy in the system equations as well as in the output equation are presented in Sect. 3.1. It is shown in practical lab-bench experiments in Sect. 4 that the nonlinear observer works robustly and achieves satisfying estimation results for both velocity and position. Furthermore, in addition to the nonlinear SMO, the case of long-term variation of the eddy current resistance is examined, which might occur due to temperature variations in the material of the solenoid. To this end, in Sect. 5, an iterative parameter correction scheme is proposed, whereby the periodic operation of the digital valve is exploited.

2 Dynamic Model of a Fast-Switching Dual-Coil Solenoid Valve

The dynamics of a solenoid valve are typically nonlinear. On the one hand, this is due to saturation effects of the material, which is described by the magnetization curve, and on the other hand, due to the Maxwell force being approximately inversely proportional to the square of the air gap. Another effect that plays a crucial role in modeling involves eddy currents that depend on the variation rate of the energizing current, and the velocity of the spool, crossing magnetic field lines. Magnetic hysteresis can approximately be neglected since switching valves are in on–off operation and the spool passes through the whole air gap. The model is divided into two dependent subsystems, the electromagnetic system and the mechanical system, which account for the spool dynamics. The electromagnetic subsystem is described by the induced voltage [16]

[1] It should be noticed that standard solenoid valves with a single coil can be considered, too.

$$V_\mathrm{i} = V_\mathrm{drv} - R_\mathrm{cu}\, i = \frac{\mathrm{d}\Psi(i,z)}{\mathrm{d}t} \tag{1}$$

$$= \frac{\partial \Psi(i,z)}{\partial i}\dot{i} + \frac{\partial \Psi(i,z)}{\partial z}\,\dot{z}, \tag{2}$$

with the current i, the driving voltage V_drv, the copper resistance R_cu, the position z, and the flux linkage $\Psi(i,z)$. The partial derivatives are written as $L_\mathrm{d}(i,z) := \partial\Psi(i,z)/\partial i$ and $L_\mathrm{i}(i,z) := \partial\Psi(i,z)/\partial z$, where $L_\mathrm{d}(\cdot)$ is the differential inductance and $L_\mathrm{i}(\cdot)$ is the incremental inductance. The latter is of particular interest in fast-switching applications. The mechanical model can be derived by Newton's second law of motion, i.e.,

$$m\,\dot{v} = \sum_k f_k, \tag{3}$$

with the lumped mass m of the spool, the velocity v, and the force components f_k that consist of a magnetic force term, in the following denoted as $F(i,z)$, plus friction and contact force terms. Both subsystems are elaborated in the following in view of an observer-oriented model.

2.1 Magnetic Characteristics

The characteristics $L_\mathrm{d}(i,z)$, $L_\mathrm{i}(i,z)$ and the magnetic force $F(i,z)$ can be obtained by the magnetic flux $\Phi(i,z) = \Psi(i,z)/N$, with N being the number of turns of the coil. A common way is to create the look-up tables by rather tedious measurements. In contrast, a model-based approach using just the magnetization curve of the material is proposed here. The characteristic of the magnetic flux $\Phi(i,z)$ is determined by a magnetic reluctance network (cf. [12, 25]). In Fig. 1, a sketch with dimensions of the

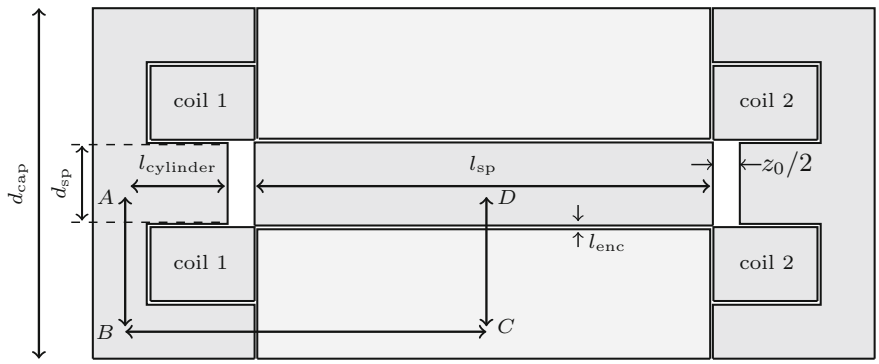


Fig. 1 Sketch of the dual-coil solenoid valve

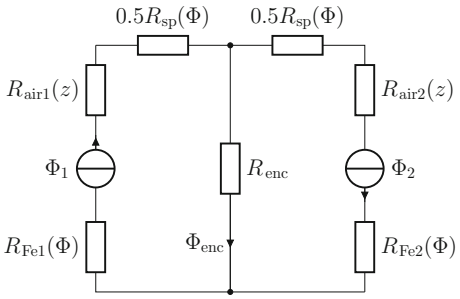


Fig. 2 Magnetic circuit of the dual-coil solenoid valve

considered dual-coil solenoid valve is depicted. The corresponding magnetic circuit is shown in Fig. 2 (cf. [25]). Here, R denotes the magnetic reluctance that can be calculated using the geometry of the actuator and the permeability $\mu(\Phi)$ by

$$R(\Phi) = \frac{l_{\mathrm{m}}}{A_{\mathrm{m}}\,\mu(\Phi)}, \tag{4}$$

with the length l_{m} and the cross-sectional area A_{m} of the considered geometry [16]. Each reluctance R corresponds to a certain area with associated dimensions respective to the magnetic flux path (A-B-C-D-A), as sketched in Fig. 1. Therefore, $R_{\mathrm{sp}}(\mu(\Phi))$, $R_{\mathrm{air}}(z)$, and $R_{\mathrm{Fe}}(\mu(\Phi))$ correspond to the spool, the air gap, and the ferromagnetic material, respectively. To obtain an equation for the flux $\Phi(\cdot)$, the magnetic network is set up by applying Hopkinson's law

$$\theta = R\,\Phi = Ni, \tag{5}$$

with the magnetomotive force θ. For nonlinear networks, the superposition law does not hold. Thus, the network equations have to be solved numerically, which allows for a consideration of magnetic properties like the nonlinear magnetization curve and the flux-dependent permeability. Deploying a balance equation, the fluxes Φ_1 and Φ_2 can be obtained by

$$\begin{aligned}\Phi_1(i,z) \;= \arg\min_{\Phi_1} \Bigg[&(\Phi_1 - \Phi_{\mathrm{enc}})\,R_{\mathrm{Fe2}}(\Phi_2) + (\Phi_1 - \Phi_{\mathrm{enc}})\,R_{\mathrm{air2}}(z) \\ &+ (\Phi_1 - \Phi_{\mathrm{enc}})\,\frac{1}{2}R_{\mathrm{sp}}(\Phi_2)\Bigg]^2 \\ &+ \Bigg[N\,I_1 - \Phi_1\left(R_{\mathrm{Fe1}}(\Phi_1) + R_{\mathrm{air1}}(z) + \frac{1}{2}R_{\mathrm{sp}}(\Phi_1)\right)\Bigg]^2, \end{aligned} \tag{6}$$

with $N\,I_1 = \theta_2$. The corresponding look-up tables for the characteristics $L_{\mathrm{d}}(i,z)$ and $L_{\mathrm{i}}(i,z)$ are obtained with $\Psi(i,z) = N\,\Phi(i,z)$ by computing the partial derivatives as defined in (2), and the magnetic force by

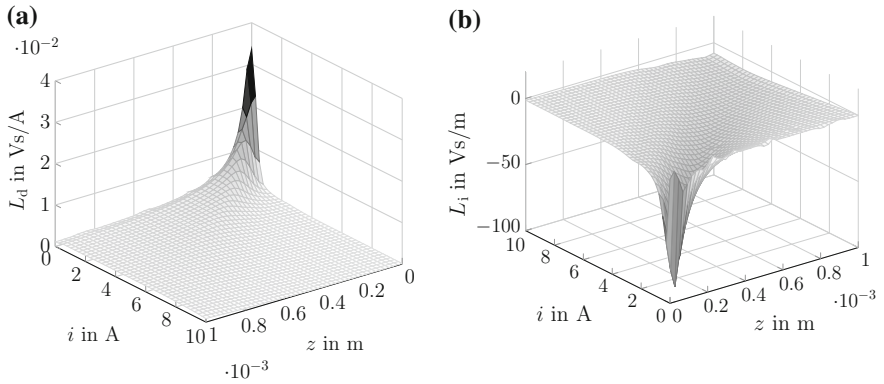


Fig. 3 Differential inductivity $L_\mathrm{d}(i, z)$ (**a**) and derivative of the flux linkage w.r.t. position $L_\mathrm{i}(i, z)$ (**b**) dependent on the current i and the position z

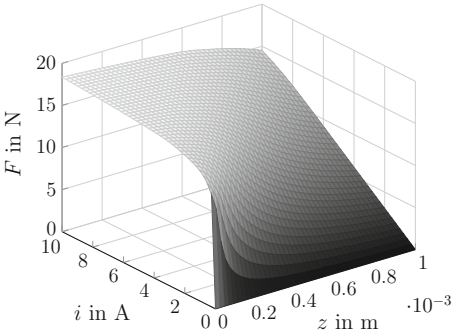


Fig. 4 Magnetic force characteristic $F(i, z)$ dependent on the current i and the position z

$$F(i, z) = \frac{\partial}{\partial z} \int_0^i \Psi(\iota, z) \mathrm{d}\iota. \tag{7}$$

The results for $L_\mathrm{d}(i, z)$, $L_\mathrm{i}(i, z)$, and $F(i, z)$ are shown in Figs. 3a, b, and 4, respectively.

2.2 Eddy Current Resistance

The effect of eddy currents i_ed can be described by including a lumped eddy current resistance R_ed in parallel to the induced voltage V_i that yields the simple dependency [16, 24]

$$i_\mathrm{ed} = \frac{1}{R_\mathrm{ed}} \frac{\mathrm{d}}{\mathrm{dt}} \Psi(i, z) = \frac{V_\mathrm{i}}{R_\mathrm{ed}}. \tag{8}$$

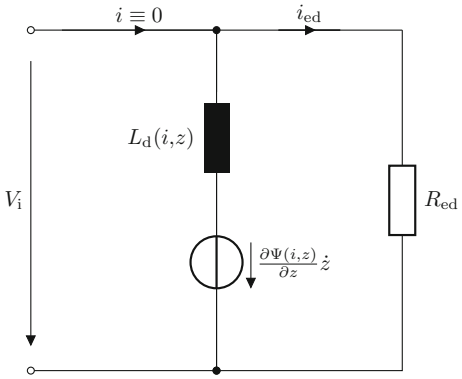


Fig. 5 Electrical equivalent circuit of a solenoid with eddy current resistance for the non-energized coil

Figure 5 shows the equivalent circuit of the solenoid with eddy current resistance for the non-energized case, i.e., merely the induced voltage V_{i} is present, and the copper resistance R_{cu} can be neglected.

A relationship of R_{ed} for radially symmetric electromagnets is given by [16] as

$$R_{\mathrm{ed}} = \frac{N^2\, l_{\mathrm{ed}}}{\sigma\, A_{\mathrm{ed}}}, \tag{9}$$

where σ is the electrical conductivity of iron, A_{ed} is the effective eddy current area, and l_{ed} is the effective path of angular eddy current density $N\, i_{\mathrm{ed}}/A_{\mathrm{ed}}$. Instead of assuming a constant R_{ed}, it might be more realistic to express the effective eddy current area A_{ed} dependent on the spool position. In Fig. 6, the position-dependent eddy current area A_{ed} is depicted. Therefore, a possible expression for A_{ed} is

$$A_{\mathrm{ed}}(z) = r\,(h + z_0 - z), \tag{10}$$

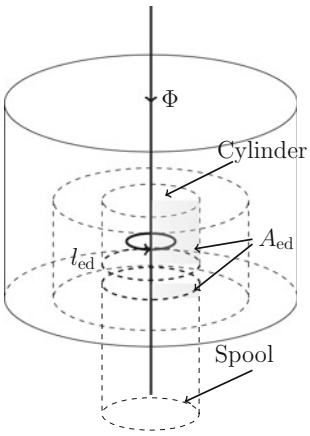


Fig. 6 Effective cross-sectional area dependent on the air gap

where r is the radius of the spool and cylinder, h is the length of the cylinder, and z_0 is the maximum air gap. With $l_{\mathrm{ed}} = 2\,\pi\,r^2$, it follows that

$$R_{\mathrm{ed}}(z) = \frac{N^2\,2\,\pi}{\sigma\,(h + z_0 - z)}. \tag{11}$$

2.3 Mechanical Subsystem

Here, a simplified model of the spool motion is considered. The magnetic force $F(i, z)$ attracts the spool, which is stopped by the end-caps if the motion is uncontrolled. Therefore, the limit stops might be modeled by an elastic impact model with the contact force

$$F_{\mathrm{c}}(z, v) = \frac{1}{2}(1 - \mathrm{sign}(z))(c_{\mathrm{s}}\,z + c_{\mathrm{d}}\,v), \tag{12}$$

with the position z, the velocity v, and constants c_{s} and c_{d}.[2] Friction is taken into account by a viscous friction term $c_\mu\,v$, with constant coefficient c_μ. Taking into account (3), the state-space model which describes the spool motion yields

$$\dot{z} = v, \qquad z(0) = z_0 \tag{13}$$

$$\dot{v} = -\frac{1}{m}\,F(z, i) - \frac{c_\mu}{m}\,v - \frac{1}{m}\,F_{\mathrm{c}}(z, v), \quad v(0) = 0. \tag{14}$$

The force $F_{\mathrm{c}}(z, v)$ solely affects the dynamics at the end of a trajectory, i.e., if the spool motion starts at $z = z_0$, then, at $z = 0$, bounces of the spool occur until it returns to the idle state.

2.4 Observer-Oriented Model

The observer-oriented model results from the previous derivations, whereby the key point is to define a suitable system output. Since the induced voltage V_{i} of the non-energized coil contains information regarding the spool dynamics, it is exploited for the observer model. This is advantageous since it is not perturbed by the energizing current i and, therefore, the typically temperature-dependent copper resistance is not needed to be taken into account when measuring the induced voltage. Therefore, in the following, V_{i} is defined as the system output $y := V_{\mathrm{i}}$. Considering the equivalent circuit in Fig. 5, the induced voltage can be described by

$$R_{\mathrm{ed}}(z)\,i_{\mathrm{ed}} = -L_{\mathrm{d}}(i_{\mathrm{ed}}, z)\,\dot{i}_{\mathrm{ed}} - L_{\mathrm{i}}(i_{\mathrm{ed}}, z)\dot{z}, \tag{15}$$

[2] Compare with [11], where a more sophisticated contact force model is used in this context.

and the resulting nonlinear state equation for the electromagnetic subsystem, with the state $i_{\mathrm{ed}} =: x_1$, the position $z =: x_2$, and the velocity $v =: x_3$ yields

$$\dot{x}_1 = -\frac{R_{\mathrm{ed}}(x_2)}{L_{\mathrm{d}}(x_1, x_2)}\, x_1 - \frac{L_{\mathrm{i}}(x_1, x_2)}{L_{\mathrm{d}}(x_1, x_2)}\, x_3. \tag{16}$$

Then, with (13)–(14), it can be obtained the dynamic system of the dual-coil solenoid valve, with the input $u := i$

$$\dot{x}_1 = -\frac{R_{\mathrm{ed}}(x_2)}{L_{\mathrm{d}}(x_1, x_2)}\, x_1 - \frac{L_{\mathrm{i}}(x_1, x_2)}{L_{\mathrm{d}}(x_1, x_2)}\, x_3, \qquad x_1(0) = 0 \tag{17}$$

$$\dot{x}_2 = x_3, \qquad x_2(0) = z_0 \tag{18}$$

$$\dot{x}_3 = -\frac{1}{m} F(x_2, u) - \frac{c_\mu}{m}\, x_3 - \frac{1}{m} F_{\mathrm{c}}(x_2, x_3), \quad x_3(0) = 0 \tag{19}$$

$$y = [R_{\mathrm{ed}}(x_2), 0, 0]\, \mathbf{x},$$

or abbreviated as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, u), \quad \mathbf{x}(0) = \mathbf{x}_0 \tag{20}$$

$$y = \mathbf{c}^T(x_2)\, \mathbf{x}. \tag{21}$$

In this model, since the square of the current is proportional to the magnetic force, the energizing current is used as the input, directly governing the attracting force $F(x_2, u)$ without considering the electrodynamics. Therefore, the system order of the attracting coil is reduced by one.

Before moving on to the observer design and convergence analysis, this section concludes with some remarks that would be useful for the following.

Remark 1 All initial states of the real system are always known, which is derived from practical considerations of digital valve applications. That is, where initially the valve is either opened or closed, the armature is at rest and, thus, no eddy current is flowing. Therefore, the major task of the state observer is to be robust against modeling and parameter uncertainty. The observer might be considered as a (robust) simulator [30].

Remark 2 Obviously, all states are bounded, and, therefore, $\|\mathbf{x}\|_\infty < d^+$ holds. The considered bounded domain of the state space is denoted as $D \subseteq \mathbb{R}^3$, whereby $D := \{\mathbf{x} \in \mathbb{R}^3 : a_i \leq x_i \leq b_i, i = 1, \ldots, 3\}$, for some constants a_i, b_i.

Remark 3 All parameters and characteristics are bounded and the following properties hold: $L_{\mathrm{d}}(x_1, x_2) > 0$, $L_{\mathrm{i}}(x_1, x_2) < 0$, $F(i, x_2) \geq 0$, $R_{\mathrm{ed}}(x_2) > 0 \; \forall \, \mathbf{x} \in D$, and $c_\mu > 0$.

Remark 4 The system is locally observable, as discussed in [4].

3 Sliding Mode Observer Design and Analysis of Error Convergence

Due to an uncertainty in the parameter $R_{\mathrm{ed}}(\cdot)$ and expectable inaccuracy in the magnetic characteristics $L_{\mathrm{d}}(\cdot)$, $L_{\mathrm{i}}(\cdot)$, and $F(\cdot)$, a robust observer design is demanded. The switching time of digital valves is normally smaller than 5 ms; therefore, a fast and finite time convergence of the estimated states to the real states has to be assured. An SMO design seems promising due to its robustness against matched uncertainty and model inaccuracy, as well as its fast convergence time [26, 28]. Other possible observer approaches in this context, which are based on extended linearization, can be found in [5]. There, it is discussed that these observers fail in practical experiments due to poor numerical efficiency and robustness. Also, the extended Luenberger observer in [21] fails in real-life experiments due to a lack of robustness.

In the sequel, first, the principle of the SMO approach for the system under consideration is pointed out, assuming well-determined system dynamics. Then, error convergence in the presence of modeling uncertainty is studied. The nonlinear SMO is formulated as

$$\dot{\hat{\mathbf{x}}} = \mathbf{f}(\hat{\mathbf{x}}, u) + \mathbf{K}\,\mathrm{sign}(\mathbf{e}), \quad \hat{\mathbf{x}}(0) = \mathbf{x}_0, \quad \mathbf{e}(0) = \mathbf{0} \tag{22}$$

$$\hat{y} = \mathbf{c}^T(\hat{x}_2)\,\hat{\mathbf{x}}, \tag{23}$$

with the observer error $\mathbf{e} := \mathbf{x} - \hat{\mathbf{x}}$, the observer gain matrix $\mathbf{K} := \mathrm{diag}\,([k_1, 0, k_3])$, the discontinuous observer injections $\mathrm{sign}(\mathbf{e}) := \left[\mathrm{sign}(e_1),\ 0,\ \mathrm{sign}(e_{3\mathrm{eq}})\right]^T$, $\mathbf{c}(\hat{x}_2) = [R_{\mathrm{ed}}(\hat{x}_2), 0, 0]$, and the sign function

$$\mathrm{sign}(e_i) = \begin{cases} 1, & e_i > 0, \\ -1, & e_i < 0, \end{cases} \quad i = 1, 3. \tag{24}$$

The value $e_{3\mathrm{eq}}$ is the equivalent value [8, 15] for the error e_3 that can be obtained by the collapsed dynamics [26] of the electrodynamic equation that is demonstrated below. An equivalent value for the error e_2 cannot be obtained, which is why the corresponding injection is dropped. Indeed, an observer injection for the second system equation is not needed, as it will be shown in the following.

Remark 5 Due to the purpose of trajectory observation, solely the states before reaching the limit stops, i.e., the states of the set $\{\mathbf{x} \in D : 0 < x_2 \leq z_0\}$, are of interest, where $F_{\mathrm{c}}(x_2, x_3)$ can assumed to be zero. Hence, for the sake of simplicity, it is reasonable to neglect $F_{\mathrm{c}}(x_2, x_3)$ in (20) as well as in (22) for the observer design and error analysis.

Then, the error dynamics of (22) yield

$$\dot{e}_1 = -\frac{R_{\mathrm{ed}}(\hat{x}_2)}{L_{\mathrm{d}}(\hat{x}_1,\hat{x}_2)}\, e_1 - \frac{L_{\mathrm{i}}(\hat{x}_1,\hat{x}_2)}{L_{\mathrm{d}}(\hat{x}_1,\hat{x}_2)}\, e_3 + \Delta f_1(e_1,e_2) - k_1\,\mathrm{sign}(e_1) \tag{25}$$

$$\dot{e}_2 = e_3 \tag{26}$$

$$\dot{e}_3 = -\frac{\Delta F(e_2)}{m} - \frac{c_\mu}{m}\, e_3 - k_3\,\mathrm{sign}(e_3), \tag{27}$$

where, for the moment, fully determined dynamics are assumed, i.e., disturbances $\Delta f_1(\cdot)$ and $\Delta F(\cdot)$, arising due to observation errors, can be considered as vanishing perturbations. Then, $\mathbf{e} = \mathbf{0}$ is an equilibrium point of (25)–(27). Furthermore, at this step, possible additive disturbances in the arguments of the observer injections are neglected, i.e., $\mathrm{sign}(y - \hat{y}) = \mathrm{sign}(R_{\mathrm{ed}}(\cdot)\, e_1) = \mathrm{sign}(e_1)$.

Remark 6 Since the error-dynamic system (25)–(27) is discontinuous at points where $e_1 = 0$ or $e_3 = 0$, all solutions $\mathbf{e}$ should be understood in the sense of Filippov [10]. In the considerations here, it is not necessary to go into detail with the properties of Filippov regularization and associated differential inclusions; refer to [7, 10, 22] for greater details and basic definitions.

The sliding variable for the electromagnetic subsystem is defined as e_1. Hence, the reaching condition $e_1\dot{e}_1 \leq 0$ can be proved by the Lyapunov function $V_1(e_1) = 1/2\, e_1^2$. It follows for the derivative w.r.t. time

$$\dot{V}_1(e_1) = e_1\dot{e}_1 = -\frac{R_{\mathrm{ed}}}{L_{\mathrm{d}}(\hat{x}_1,\hat{x}_2)}\, e_1^2 - \frac{L_{\mathrm{i}}(\hat{e}_1,\hat{x}_2)}{L_{\mathrm{d}}(\hat{x}_1,\hat{x}_2)}\, e_1 e_3 + \Delta f_1(e_1,e_2) e_1 - k_1\,|e_1| \tag{28}$$

$$\leq \left(\left\| \frac{L_{\mathrm{i}}(\hat{x}_1,\hat{x}_2)}{L_{\mathrm{d}}(\hat{x}_1,\hat{x}_2)} \right\|_\infty \|e_3\|_\infty + \|\Delta f_1(e_1,e_2)\|_\infty - k_1 \right) |e_1| \tag{29}$$

$$\leq -\nu_1\,|e_1|, \tag{30}$$

for $k_1 = \left\| L_{\mathrm{i}}(\cdot)/L_{\mathrm{d}}(\cdot) \right\|_\infty \|e_3\|_\infty + \|\Delta f_1(\cdot)\|_\infty + \nu_1$, $\nu_1 > 0$. Taking into account the comparison principle [17], a finite reaching time

$$t_{\mathrm{r1}} \leq \frac{2|e_1(0)|}{\nu_1} \tag{31}$$

can be obtained. Since $e_1(0) = 0$, convergence of e_1, and, therefore, sliding occurs immediately. Now, consider Eq. (25) with regard to the sliding condition $\dot{e}_1 = e_1 = 0$. Then, the collapsed dynamics of the electromagnetic subsystem yields

$$0 = -\frac{L_{\mathrm{i}}(\hat{x}_1,\hat{x}_2)}{L_{\mathrm{d}}(\hat{x}_1,\hat{x}_2)}\, e_3 + \Delta f_1(0,e_2) - k_1\,\mathrm{sign}(e_1). \tag{32}$$

Due to the matching condition [26], the equivalent output injection,[3] for small errors e_2, is given by

[3] Regarding equivalent output injection, compare [8], and see [18] for an application-driven design.

$$\left(k_1 \operatorname{sign}(y - \hat{y})\right)_{\text{eq}} \approx -\frac{L_{\text{i}}(\hat{x}_1, \hat{x}_2)}{L_{\text{d}}(\hat{x}_1, \hat{x}_2)} e_3, \tag{33}$$

and one can obtain the equivalent value for the error e_3 as

$$e_{3\text{eq}} \approx -\frac{L_{\text{d}}(\hat{x}_1, \hat{x}_2)}{L_{\text{i}}(\hat{x}_1, \hat{x}_2)} \left(k_1 \operatorname{sign}(e_1)\right)_{\text{eq}}. \tag{34}$$

This value can be used for the observer injection of the last system equation. Hence, it is possible to define e_3 as the sliding variable for (27). Taking $V_3(e_3) = 1/2\, e_3^2$ as a Lyapunov function yields for the gradient along the trajectories e_3

$$\dot{V}_3(e_3) = e_3 \dot{e}_3 = -\frac{\Delta F(e_2)}{m} e_3 - \frac{c_\mu}{m} e_3^2 - k_3 |e_3| \tag{35}$$

$$\leq \left(\left\| \frac{\Delta F(e_2)}{m} \right\|_\infty - k_3 \right) |e_3| \tag{36}$$

$$\leq -\nu_3 |e_3|, \tag{37}$$

with $k_3 = \|\Delta F(e_2)/m\|_\infty + \nu_3$, $\nu_3 > 0$. Therefore, a sliding motion on $e_3 = 0$ occurs immediately since $e_3(0) = 0$, and since the finite reaching time can be calculated by

$$t_{\text{r}3} \leq \frac{2|e_3(0)|}{\nu_3}. \tag{38}$$

Remark 7 It is worth noting that the sliding conditions are also fulfilled for bounded disturbances of the dynamics since merely gains k_1, k_3 must be chosen large enough [26]. However, disturbances that are too high principally corrupt the equivalent value of e_3; therefore, they will influence the position error e_2 as well. The effect of disturbances is examined in the next section separately.

While e_1 and e_3 are in a sliding mode, the position error e_2 is stable inside a positive-invariant set $\{e_2 : |\Delta F(e_2)/m| < k_3, e_3 = 0\}$, containing the zero-equilibrium point of the error system. This can be obtained with a generalized Krasovskii–LaSalle Theorem [3] by considering the discontinuity surfaces of the mechanical subsystem (27) when the error system in e_1 (25) is collapsed. The reader is directed to [3, 22] for further details.

3.1 Analysis of Error Convergence in the Presence of Modeling Uncertainty

The robustness of the observation error of the SMO (22) regarding the system dynamics (20) in the presence of both uncertainty of the modeled characteristics

$L_{\mathrm{d}}(\cdot)$, $L_{\mathrm{i}}(\cdot)$, $F(\cdot)$ and of the observer injections is now examined. Errors in the argument of the observer injections may arise, on the one hand, due to the eddy current resistance $R_{\mathrm{ed}}(x_2)$, which is contained in the output equation $y = \mathbf{c}^T(x_2)\,\mathbf{x}$ and is not exactly known. On the other hand, this leads to an inaccuracy in the argument of the equivalent output injection term $\mathrm{sign}(e_{3\mathrm{eq}})$. It will be shown that the reaching condition in the case of bounded disturbances of the output equation and the system dynamics still holds for a uniform bound of the estimation error, which can be diminished with increasing observer gains. Principally, the error of SMOs in the presence of output noise effects is still bounded; however, it depends on the bound of the output noise or disturbance. This is shown, e.g., for stochastic output noise effects in [2], and for both stochastic and deterministic output disturbances in [23, 29]. Here, solely deterministic output disturbances will be considered. The error in the argument of the first switching term $\mathrm{sign}(y - \hat{y})$, i.e., the error of the measurement vector, will be written as

$$y - \hat{y} = \mathbf{c}^T\,\mathbf{x} - \hat{\mathbf{c}}^T\,\hat{\mathbf{x}} = \hat{R}_{\mathrm{ed}}(\hat{x}_2)\,e_1 + \Delta R_{\mathrm{ed}}(x_2, \hat{x}_2)\,x_1, \tag{39}$$

where the disturbance term $\Delta R_{\mathrm{ed}}(x_2, \hat{x}_2)\,x_1$ is nonvanishing. The equivalent value $e_{3\mathrm{eq}}$ in the case of modeling uncertainty will be defined as

$$e_{3\mathrm{eq}} = e_3 + \Delta e_{3\mathrm{eq}}, \tag{40}$$

which is also nonvanishing since the value is affected by modeling uncertainty in the electromagnetic subsystem due to (33). Consequently, the output injection error vector yields

$$\xi = [\xi_1,\ 0,\ \xi_3]^T := \left[\Delta R_{\mathrm{ed}}(\cdot)\,x_1,\ 0,\ \Delta e_{3\mathrm{eq}}\right]^T. \tag{41}$$

Then, the error dynamics with modeling errors $\Delta\mathbf{f}(\mathbf{x}, \hat{\mathbf{x}}) = \mathbf{f}(\mathbf{x}) - \hat{\mathbf{f}}(\hat{\mathbf{x}})$ is rewritten in detail as

$$\begin{aligned}\dot{e}_1 = &-\frac{\hat{R}_{\mathrm{ed}}(\hat{x}_2)}{\hat{L}_{\mathrm{d}}(\hat{x}_1, \hat{x}_2)}\,e_1 - \frac{\hat{L}_{\mathrm{i}}(\hat{x}_1, \hat{x}_2)}{\hat{L}_{\mathrm{d}}(\hat{x}_1, \hat{x}_2)}\,e_3 \\ &+ \Delta f_1(\mathbf{x}, \hat{\mathbf{x}}) - k_1\,\mathrm{sign}\left(\hat{R}_{\mathrm{ed}}(\hat{x}_2)\,e_1 + \xi_1\right)\end{aligned} \tag{42}$$

$$\dot{e}_2 = e_3 \tag{43}$$

$$\dot{e}_3 = \Delta f_3(\mathbf{x}, \hat{\mathbf{x}}) - \frac{\hat{c}_\mu}{m}\,e_3 - k_3\,\mathrm{sign}(e_3 + \xi_3), \tag{44}$$

where the $\hat{}$ above the modeled characteristics and the parameters indicates that they are not exactly known.

Now, the following assumptions are made:

Assumption 1 Modeling errors, respectively disturbances due to structured uncertainty in the system dynamics, are bounded, i.e.,

$$\|\Delta f_i(\mathbf{x}, \hat{\mathbf{x}})\|_\infty < \Delta f_i^+ < \infty, \quad i = 1, 3. \tag{45}$$

Assumption 2 Disturbances due to structured uncertainty in the arguments of the observer injections are bounded, i.e.,

$$\|\xi_i\|_\infty < \xi_i^+ < \infty, \quad i = 1, 3. \tag{46}$$

Remark 8 These assumptions are underpinned since the states and characteristics are bounded.

Theorem 1 *Provided that Assumptions 1–2 hold, the errors e_1 and e_3 of the uncertain system (42)–(44) are uniformly bounded by*

$$|e_i| \leq |e_i|^+ := \sqrt{\left(\frac{\nu_i}{\sqrt{2}\,\alpha_i}\right)^2 + \frac{2\delta_i(\nu_i)}{\alpha_i}} - \frac{\nu_i}{\sqrt{2}\alpha_i}, \quad i = 1, 3, \tag{47}$$

with constants $\alpha_i > 0$ and $\nu_i, \delta_i \geq 0$. The bounds $|e_i|^+$ are decreasing for increasing values ν_i, representing observer gains, whereby

$$\lim_{\nu_1 \to \infty} |e_1|^+ = 2\,\xi_1^+ \big/ \hat{R}_{\mathrm{ed\,min}}, \tag{48}$$

$$\lim_{\nu_3 \to \infty} |e_3|^+ = 2\,\xi_3^+. \tag{49}$$

Remark 9 Due to (9) and (10), the smallest value of $\hat{R}_{\mathrm{ed}}(\hat{x}_2)$ can be indicated for $\hat{x}_2 = 0$, i.e., $\hat{R}_{\mathrm{ed\,min}} := \hat{R}_{\mathrm{ed}}(0)$, and here $\hat{R}_{\mathrm{ed}}(0) > 1\,\Omega$.

Proof The reaching conditions are examined as before with the Lyapunov function $V_1(e_1) = 1/2\,e_1^2$ for e_1 and $V_3(e_3) = 1/2\,e_3^2$ for e_3. For the error e_1, it follows with (42)

$$\dot{V}_1(e_1) = -\frac{\hat{R}_{\mathrm{ed}}(\hat{x}_2)}{\hat{L}_{\mathrm{d}}(\hat{x}_1, \hat{x}_2)}\, e_1^2 - \frac{\hat{L}_{\mathrm{i}}(\hat{x}_1, \hat{x}_2)}{\hat{L}_{\mathrm{d}}(\hat{x}_1, \hat{x}_2)}\, e_1 e_3 + \Delta f_1(\mathbf{x}, \hat{\mathbf{x}})\, e_1 \tag{50}$$

$$-\tilde{k}_1 \hat{R}_{\mathrm{ed}}(\hat{x}_2)\, e_1 \operatorname{sign}\left(\hat{R}_{\mathrm{ed}}(\hat{x}_2)\, e_1 + \xi_1\right), \tag{51}$$

where $k_1 = \tilde{k}_1 \hat{R}_{\mathrm{ed}}(\hat{x}_2)$. Using the identity

$$x \operatorname{sign}(x + z) = (x + z) \operatorname{sign}(x + z) - z \operatorname{sign}(x + z) \tag{52}$$

$$= |x + z| - z \operatorname{sign}(x + z), \tag{53}$$

and the fact that $|x + z| \geq |x| - |z|$, the following inequality can be obtained[4]

$$x \operatorname{sign}(x + z) \geq |x| - |z| - z \operatorname{sign}(x + z) \geq |x| - 2\,|z|. \tag{54}$$

[4]This is a scalar representation of the vector-valued identity in [23].

This leads to

$$\dot{V}_1(e_1) \leq -\frac{\hat{R}_{\mathrm{ed}}(\hat{x}_2)}{\hat{L}_{\mathrm{d}}(\hat{x}_1,\hat{x}_2)}\, e_1^2 + \left|\frac{\hat{L}_{\mathrm{i}}(\hat{x}_1,\hat{x}_2)}{\hat{L}_{\mathrm{d}}(\hat{x}_1,\hat{x}_2)}\right| |e_1|\,|e_3| + \left|\Delta f_1(\mathbf{x},\hat{\mathbf{x}})\right|\, |e_1| \tag{55}$$

$$-\tilde{k}_1 \hat{R}_{\mathrm{ed\,min}}\, |e_1| + 2\,\tilde{k}_1\, \xi_1 \tag{56}$$

$$\leq -\alpha_1\, V_1 - \nu_1 \sqrt{V_1}\, + \delta_1(\nu_1), \tag{57}$$

where

$$\tilde{k}_1 \;=\; \left(\left\|\hat{L}_{\mathrm{i}}(\cdot)/\hat{L}_{\mathrm{d}}(\cdot)\right\|_\infty \|e_3\|_\infty + \Delta f_1^+ + \nu_1/\sqrt{2}\right) \Big/ \hat{R}_{\mathrm{ed\,min}} \tag{58}$$

$$= \left(k_{1\,\mathrm{min}} + \nu_1/\sqrt{2}\right) \Big/ \hat{R}_{\mathrm{ed\,min}}, \quad \nu_1 \geq 0, \tag{59}$$

$$\delta_1(\nu_1) := 2\,\tilde{k}_1(\nu_1)\,\xi_1^+ = 2\,\xi_1^+\, k_{1\,\mathrm{min}} \big/ \hat{R}_{\mathrm{ed\,min}} + \sqrt{2}\,\xi_1^+\,\nu_1 \big/ \hat{R}_{\mathrm{ed\,min}} = const, \tag{60}$$

$$\alpha_1 := \hat{R}_{\mathrm{ed\,min}} / \|\hat{L}_{\mathrm{d}}(\cdot)\|_\infty > 0. \tag{61}$$

Remark 10 The value $k_{1\,\mathrm{min}}$ denotes the minimal needed observer gain to bound the observation error e_1 in the case of modeling uncertainty $\Delta f_1(\cdot)$ in the system equations.

It follows for the equilibrium $\tilde{V}_1^*$ of

$$\dot{\tilde{V}}_1(e_1) = -\alpha_1\, \tilde{V}_1 - \nu_1 \sqrt{\tilde{V}_1}\, + \delta_1(\nu_1), \tag{62}$$

that

$$\tilde{V}_1^* = \left(\sqrt{\left(\frac{\nu_1}{2\alpha_1}\right)^2 + \frac{\delta_1(\nu_1)}{\alpha_1}} - \frac{\nu_1}{2\alpha_1}\right)^2, \tag{63}$$

which involves that $\dot{V} = e_1\,\dot{e}_1$ is negative for

$$|e_1| > \sqrt{\left(\frac{\nu_1}{\sqrt{2}\,\alpha_1}\right)^2 + \frac{2\delta_1(\nu_1)}{\alpha_1}} - \frac{\nu_1}{\sqrt{2}\alpha_1}. \tag{64}$$

Hence, it can be concluded that e_1 is uniformly bounded [17, 29]. Similar assertions can be made for the error e_3 with $V_3(e_3) = 1/2\, e_3^2$. It follows with (44)

$$\dot{V}_3(e_3) = -\frac{\hat{c}_\mu}{m}\, e_3^2 + \Delta f_3(\mathbf{x},\hat{\mathbf{x}})\, e_3 - k_3\, e_3\, \mathrm{sign}\,(e_3 + \xi_3) \tag{65}$$

$$\leq -\frac{\hat{c}_\mu}{m}\, e_3^2 + \Delta f_3^+\, |e_3| - k_3\;\, |e_3| + 2\, k_3\, \xi_3 \tag{66}$$

$$\leq -\alpha_3\, V_3 - \nu_3 \sqrt{V_3}\, + \delta_3(\nu_3), \tag{67}$$

where

$$k_3 = \Delta f_3^+ + \nu_3/\sqrt{2} = k_{3\,\min} + \nu_3/\sqrt{2}, \quad \nu_3 \geq 0, \tag{68}$$

$$\delta_3(\nu_3) := 2\,k_3(\nu_1)\xi_3^+ = 2\,\xi_3^+ k_{3\,\min} + \sqrt{2}\,\xi_3^+\,\nu_3 = const, \tag{69}$$

$$\alpha_3 := \hat{c}_\mu/m > 0. \tag{70}$$

As above, $\nabla V_3\,\dot{e}_3$ is found to be negative for

$$|e_3| > \sqrt{\left(\frac{\nu_3}{\sqrt{2}\,\alpha_3}\right)^2 + \frac{2\delta_3(\nu_3)}{\alpha_3}} - \frac{\nu_3}{\sqrt{2}\alpha_3}, \tag{71}$$

which explains the uniform boundedness of e_3. Now, the limit of $|e_i|^+$ for increasing values ν_i is considered. Rewriting (47) as

$$|e_i|^+ = \frac{1}{\sqrt{2}\alpha_i \frac{1}{\nu_i}} \left(\sqrt{1 + \left(\frac{\sqrt{2}\,\alpha_i}{\nu_i}\right)^2 \frac{2\delta_i(\nu_i)}{\alpha_i}} - 1 \right), \tag{72}$$

and inserting (69) for $i = 3$ yields[5]

$$|e_3|^+ = \frac{1}{\sqrt{2}\,\alpha_3\,\frac{1}{\nu_3}} \left(\sqrt{1 + \frac{8\,k_{3\,\min}\xi_3^+\,\alpha_3}{\nu_3^2} + \frac{4\sqrt{2}\xi_3^+\,\alpha_3}{\nu_3}} - 1 \right). \tag{73}$$

Since the evaluation of (73) for $\nu_3 \to \infty$ leads to an undefined expression of the form $0/0$, l'Hôpital's rule is applied. Independent differentiation of both the numerator denoted by $N(\nu_3)$ and the denominator denoted by $D(\nu_3)$ w.r.t. ν_3 leads to

$$\frac{N(\nu_3)'}{D(\nu_3)'} := \frac{\left(\frac{8\,k_{3\,\min}\,\xi_3^+}{\sqrt{2}\,\nu_3} + 2\,\xi_3^+\right)}{\sqrt{1 + \frac{8\,k_{3\,\min}\xi_3^+\,\alpha_3}{\nu_3^2} + \frac{4\sqrt{2}\xi_3^+\,\alpha_3}{\nu_3}}}, \tag{74}$$

whereby, $\lim_{\nu_3\to\infty} N(\nu_3)'/D(\nu_3)' = 2\xi_3^+$. Similar calculations can be done for $i = 1$ with (60). Therefore, it can be stated that the limits of $|e_1|$ and $|e_3|$ yield

$$\lim_{\nu_1\to\infty} |e_1|^+ = 2\,\xi_1^+/\hat{R}_{\text{ed}\,\min}, \tag{75}$$

$$\lim_{\nu_3\to\infty} |e_3|^+ = 2\,\xi_3^+. \tag{76}$$

This completes the proof. □

[5]Calculation steps are shown for $i = 3$ due to a more compact notation.

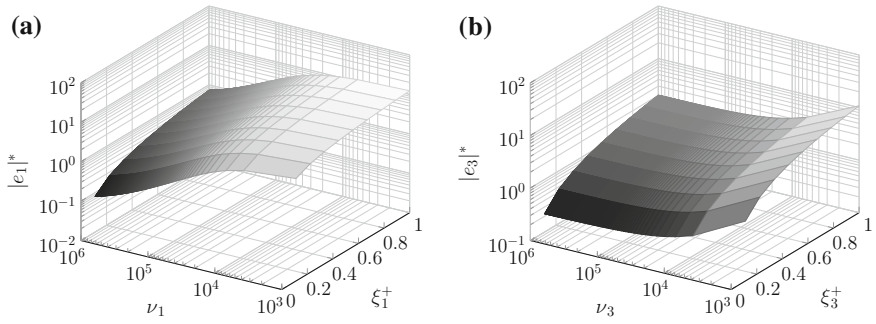


Fig. 7 Error bounds in the presence of modeling uncertainty dependent on the observer gains ν_i and the bounds of the output disturbances ξ_i^+

Remark 11 If the modeling errors $\Delta f_i(\cdot)$ are small, then smaller values of ν_i lead to smaller bounds $|e_i|^+ < 2\,\xi_i^+$. This would be the case for a parameter relation $\alpha_i \geq 2\xi_i^+ \, k_{i\min}$.

Remark 12 The errors e_1 and e_3 are also (uniformly) ultimately bounded (cf. [14, 17]), which can already be concluded by the steps above. This is not shown in detail, since the initial errors are zero and, thus, (uniform) boundedness is sufficient here.

Remark 13 Since the states are solely observed in a small finite time interval $T \ll 1$ the maximal error e_2 can be estimated by

$$\|e_2\|_\infty = \left\| \int_0^T e_3 \mathrm{d}t \right\|_\infty \leq \int_0^T \|e_3\|_\infty \, \mathrm{d}t = T \|e_3\|_\infty. \tag{77}$$

In Fig. 7, the error bounds $|e_i|$ dependent on the observer gains ν_i and the output disturbances ξ_i^+ are depicted. The bounds $|e_i|$ clearly diminish for increasing observer gains ν_i. It should be noted that the worst-case bounds of disturbances of the system equations are considered, i.e., $\Delta f_i^+ = 2\|f_i(\cdot)\|_\infty$. The stability analysis exhibits that the errors are still bounded in the presence of inaccuracy in the system equations and the output equation, and a selection of higher observer gains will lead to better estimation results. Since the limit of $|e_i|^+$ depends on errors in the output equations, the use of a precise model of $R_{\mathrm{ed}}(x_2)$ is mandatory, and an adaption of this parameter is motivated.

4 Experimental Results

Experiments have been performed on a rapid prototyping environment. For validation of the state estimates, a high-precision optical position sensor is used. Figure 8 shows the experimental setup with the dual-coil solenoid valve and an optical position sensor.

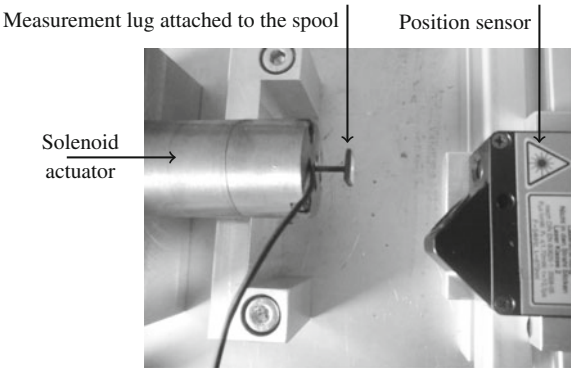


Fig. 8 Dual-coil solenoid valve mock-up with optical position sensor

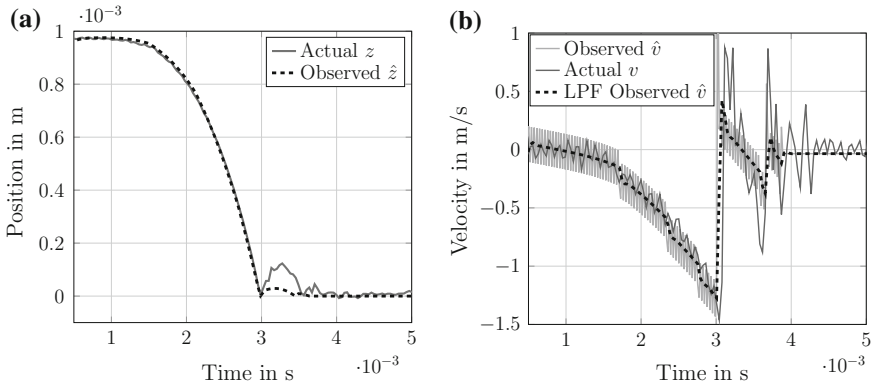


Fig. 9 Actual and observed position (**a**) and velocity (**b**)

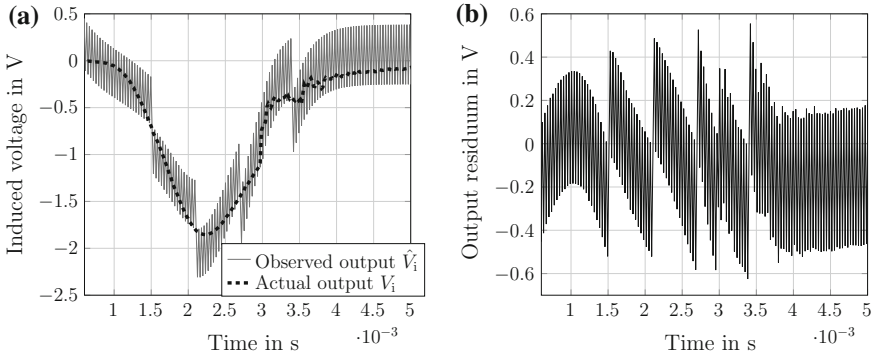


Fig. 10 Actual and observed output (**a**) and corresponding error (**b**)

4.1 Position and Velocity Observation Results

Observation results are presented for the uncontrolled case. That is, one coil is energized so that the spool is attracted, and at the same time the induced voltage is measured at the non-energized coil, which provides the injection for the observer system. The sample time of the SMO model is 20 µs, and a fixed-step size solver with third-order Runge–Kutta method is used. In Fig. 9a, experimental results for the position observation are shown. The observed trajectory (black-dashed) matches accurately the measured trajectory (gray) in spite of model and parameter uncertainty. The moment of impact, at approximately $t = 3$ ms, is also estimated correctly. The whole flight time is less than 2 ms. It should be mentioned that the estimation error after the first hit, i.e., during the bouncing phase, is not of interest. The observer gains k_1 and k_3 are chosen as 16×10^3 and 17×10^3, respectively. Experimental results for the velocity observation are shown in Fig. 9b. The observed signal $\hat{v}$ (light gray) is obviously in a sliding motion; however, its chattering is rather strong, due to the high values of the observer gains. This is compared to its low-pass filtered (LPF) value (black-dashed) as well as to the differential quotient of the low-pass filtered measured trajectory z (gray). Although strong chattering is present, the filtered observed velocity signal is rather smooth up to the point where the spool hits the end-cap. The step-like gradations in the velocity signal are typical in the case of modeling inaccuracy. They become notable due to the large gains. The same can be observed in Fig. 10a, where the actual output, i.e., induced voltage V_i (dashed), and the corresponding observer signal $\hat{V}_\mathrm{i} = \hat{R}_\mathrm{ed}\,\hat{i}_\mathrm{ed}$ are shown. In Fig. 10b, the according error is depicted. The observed signal is in a sliding motion, while the step-like chattering shape can be ascribed to modeling inaccuracy. In the sequel, repeatability is demonstrated and the effect of different models of the eddy current resistance $R_\mathrm{ed}(\cdot)$ is illustrated. The constant observer gains have the same values as used for the experiments above. In Fig. 11a, a histogram of absolute maximum position estimation errors of 75

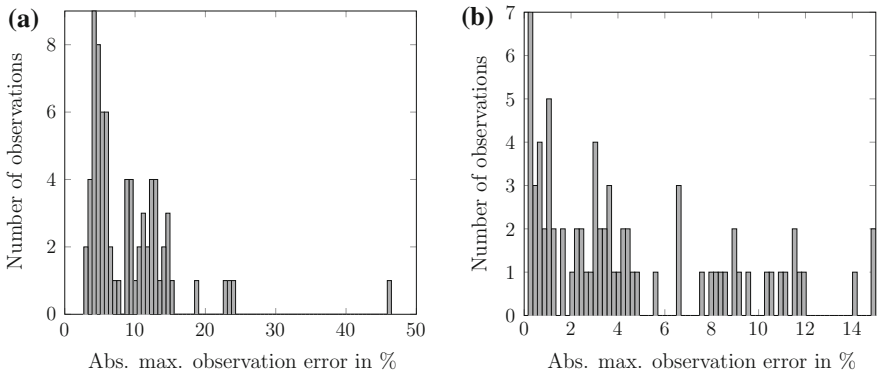


Fig. 11 Histogram of absolute maximum errors of consecutive switching cycles with $R_\mathrm{ed} \equiv const$ (**a**), and with position-dependent eddy current resistance $R_\mathrm{ed}(z)$ (**b**)

consecutive switching cycles is shown, with a constant modeled eddy current resistance R_{ed}. The average observation error is approximately 5 %. Figure 11b depicts another histogram of absolute maximum position estimation errors of 75 consecutive switching cycles, but with a non-constant modeled eddy current resistance $R_{\mathrm{ed}}(z)$. The position observations are considerably improved and the average observation error is smaller than 2 %.

It should be mentioned that the few outliers in Fig. 11a, b are ascribed to changing magnetization levels, occurring after each switching cycle, which are not regarded in the model. The fact that, in each switching operation with the position-dependent eddy current resistance, the position is estimated with an acceptable magnitude of estimation errors underlines the robustness of the observer.

5 Iterative Parameter Correction—An Experimental Approach

The previous results have shown robust and repeatable behavior of the SMO in real-life experiments. Now, the case of disturbances caused by a variation of the parameter $R_{\mathrm{ed}}(x_2)$ during long-term operation is examined. It is recalled that the eddy current resistance depends on the conductivity σ of iron, since $R_{\mathrm{ed}}(x_2) = N^2\, l_{\mathrm{ed}}/(\sigma\, A_{\mathrm{ed}}(x_2))$. For the long-term case, a temperature dependency for the conductivity of iron is considered. This is expressed as $R_{\mathrm{ed}}(x_2, \sigma(T)) = \theta(T)\, R_{\mathrm{ed}}(x_2)$, with an unknown parameter θ that reflects a change in temperature. Due to the cyclic operation of the solenoid, the parameter θ might be adapted dependent on the past measurement errors $y - \hat{y}$. At first glance, extracting the information of the low-pass filtered value of the switching term for parameter identification seems promising; however, this fails when both parameter uncertainty and modeling inaccuracy are present [26]. Therefore, another approach is addressed that takes into account the behavior of the switching term in the presence of time-varying uncertain arguments, i.e.,

$$\operatorname{sign}(y - \hat{y}) = \operatorname{sign}(y - \theta \hat{R}_{\mathrm{ed}}(\hat{x}_2)\, \hat{x}_1). \tag{78}$$

Since $\hat{R}_{\mathrm{ed}}(\hat{x}_2)$ weights the chattering state $\hat{x}_1$, chattering of the residuum $e_y = y - \theta \hat{R}_{\mathrm{ed}}(\hat{x}_2)\, \hat{x}_1$ is also considerably influenced. In the following, K error-signal samples over an entire switching cycle are collected in a vector $\mathbf{e}_y$, with the vector norm

$$\|\mathbf{e}_y\| = \sqrt{\sum_{k=1}^{K} \mathbf{e}_y(k)^2}. \tag{79}$$

Ensuing from a proper operation, a greater weight θ will cause an increased error norm $\|\mathbf{e}_y\|$, whereas a smaller gain will cause a decreased error norm $\|\mathbf{e}_y\|$ due to increased or decreased chattering amplitudes, respectively. This is motivated by the

observations of practical experiments and can be verified with numerical simulations. Thus, it can be defined the dead-zone function

$$\mathrm{DZ}(\|\mathbf{e}_y\|) = \begin{cases} -1, & \|\mathbf{e}_y\| > \mathrm{UB}, \\ 0, & \mathrm{LB} \leq \|\mathbf{e}_y\| \leq \mathrm{UB}, \\ 1, & \|\mathbf{e}_y\| < \mathrm{LB}, \end{cases} \tag{80}$$

where LB and UB are the lower and upper bounds of $\|\mathbf{e}_y\|$, respectively. Hence, the empirical update law

$$\theta_k = \theta_{k-1} + \gamma \cdot \mathrm{DZ}(\|\mathbf{e}_y\|), \tag{81}$$

where $\gamma > 0$ indicates the size of the correction step, can be determined. In the following, practical experiments are shown for a correction of a deviated eddy current resistance. The experiments were conducted online together with the SMO in the same test environment as described above. In a nominal operation, the value of θ is equal to one. In Fig. 12a, consecutive and converging position estimates are shown where a greater value of R_{ed} is pretended, i.e., weighted by $\theta > 1$. Due to a smaller value $\hat{i}_{\mathrm{ed}}$, a slower velocity $\hat{v}$ is assumed by the observer; hence, it leads to a lag in the position estimate $\hat{z}$. The first and last actual position trajectories are plotted for better correspondence (black-dotted). The differences occur due to different magnetization levels of the solenoid. Figure 12b depicts the convergence of the parameter θ to its nominal value with the update law (81), where a $\gamma = 0.15$ is selected. Appropriate upper and lower bounds are obtained empirically. The position estimation error tends proportionally to its minimum values, and remains there.

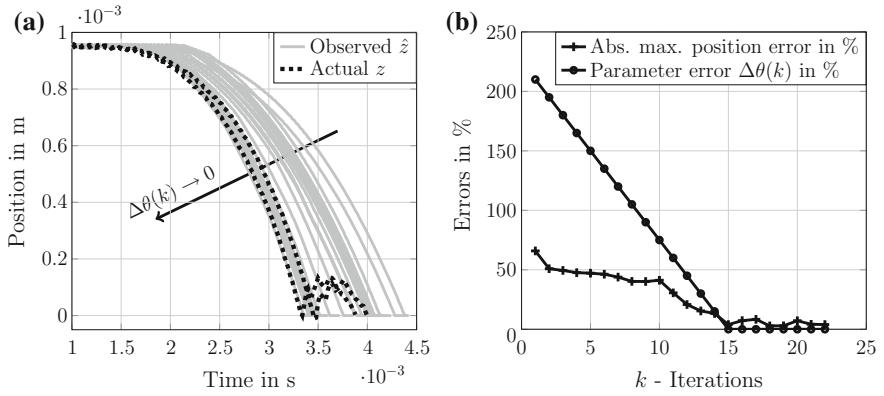


Fig. 12 In **a**, estimated trajectories from the right with a too large $R_{\mathrm{ed}}(\hat{x}_2)$ (i.e. $\theta > 1$) converging to the actual trajectory, where the first and last actual trajectories are plotted (*black-dotted*). In **b**, the converging parameter error $\Delta\theta$ as well as the converging position (trajectory) error e_2 are presented

6 Conclusion and Future Work

A nonlinear SMO approach for state observation of fast-switching solenoid valves has been presented. Solely the measured induced voltage of the non-energized coil is necessary to provide robust estimates of both position and velocity. It has been shown by a Lyapunov stability analysis that the observation errors are uniformly bounded dependent on the bounds of deterministic disturbances, arising due to an uncertainty of the dynamic observer model and an uncertainty in the measurement output. The robust performance of the proposed SMO has been demonstrated in real-world experiments, where the estimates of position trajectories are validated by a high-precision optical position sensor. This is, on the one hand, shown by the fact that the proposed observer provides repeatedly satisfying position estimates for consecutive switching operations. On the other hand, the case of robust long-term performance has also been investigated. Since the eddy current resistance might change dependent on the temperature of the iron, it is reasonable in this case to correct its value dependent on the observer residuum. Therefore, the cyclic operation of the solenoid valve has been exploited and an empirical iterative update law was found that is able to correct slow or long-term variations of that critical parameter. Experimental results were provided to verify that approach. In further research, the nonlinear SMO will be combined with a feedback-control scheme to accomplish sensorless soft landing of the moving spool. Therefore, the SMO will be combined with a flatness-based approach [6]. Moreover, the application of the proposed approach for other types of solenoid valves, particularly for standard (single coil) valves, will be investigated further. Furthermore, the influence of combined disturbance observers regarding the performance of the equivalent output injection method will be investigated as well.

Acknowledgments The support of this work by the German Federal Ministry of Education and Research under grant 03FH047PX2 is gratefully acknowledged. Furthermore, the authors would like to thank the anonymous reviewers for their invaluable comments and suggestions, as well as F. Straußberger for helpful discussions.

References

1. Abry F, Brun X, Di Loreto M, Sesmat S, Bideaux É (2015) Piston position estimation for an electro-pneumatic actuator at standstill. Control Eng Pract 41:176–185
2. Alessandri A, Cuneo M, Punta E (2010) State observers with first-/second-order sliding-mode for nonlinear systems with bounded noises. In: Proceedings of the 11th international workshop on variable structure systems (VSS), Mexico, pp 493–497
3. Alvarez J, Orlov YV, Acho L (2000) An invariance principle for discontinuous dynamic systems with application to a coulomb friction oscillator. ASME J Dyn Syst Meas Control 122(4):687–690
4. Braun T, Schwab M, Straußberger F, Reuter J (2014) State estimation for fast-switching solenoid valves—a nonlinear sliding-mode-observer approach. In: 19th IEEE international conference on methods and models in automation and robotics (MMAR14), Miedzyzdroje, pp 282–287

5. Braun T, Straußberger F, Reuter J (2015) State estimation for fast-switching solenoid valves: a study on practical nonlinear observers and new experimental results. In: 20th IEEE international conference on methods and models in automation and robotics (MMAR15), Miedzyzdroje, pp 862–867
6. Braun T, Straußberger F, Reuter J, Preissler G (2015) A semilinear distributed parameter approach for solenoid valve control including saturation effects. In: American control conference (ACC15), Chicago, pp 2600–2605
7. Danca M-F, Codreanu S (2002) On a possible approximation of discontinuous dynamical systems. Chaos Solitons Fractals 13(4):681–691
8. Drakunov S (1992) Sliding-mode observers based on equivalent control method. In: Proceedings of the 31st IEEE conference on decision and control vol 2. Tucson, pp 2368–2369
9. Eyabi P, Washington G (2006) Modeling and sensorless control of an electromagnetic valve actuator. Mechatronics 16(3–4):159–175
10. Filippov A (1988) Differential equations with discontinuous righthand sides. Mathematics and its applications. Kluwer Academic Publishers, Dordrecht
11. Glück T (2013) Soft landing and self-sensing strategies for electromagnetic actuators. Modellierung und Regelung komplexer dynamischer Systeme. Shaker, Aachen
12. Glück T, Kemmetmüller W, Kugi A (2011) Trajectory optimization for soft landing of fast-switching electromagnetic valves. In: Proceedings of the 18th IFAC world congress, Milano, pp 11532–11537
13. Glück T, Kemmetmüller W, Tump C, Kugi A (2011) A novel robust position estimator for self-sensing magnetic levitation systems based on least squares identification. Control Eng Pract 19(2):146–157
14. Haddad W, Chellaboina V (2011) Nonlinear dynamical systems and control: a Lyapunov-based approach. Princeton University Press, Princeton
15. Haskara I, Özgüner Ü (1999) Equivalent value filters in disturbance estimation and state observation. In: Young K, Özgüner Ü (eds) Variable structure systems. Sliding mode and nonlinear control. Volume 247 of lecture notes in control and information sciences. Springer, London, pp 167–179
16. Kallenbach E (2008) Elektromagnets: basics, dimensioning, design and application (in German). ViewegTeubner [GWV Fachverlage GmbH], Wiesbaden, 3rd edn
17. Khalil H (2002) Nonlinear systems. Prentice Hall PTR, Upper Saddle River
18. Kim I-S (2010) A technique for estimating the state of health of Lithium batteries through a dual-sliding-mode observer. IEEE Trans Power Electron 25(4):1013–1022
19. Kogler H, Scheidl R (2008) Two basic concepts of hydraulic switching converters. In: Proceedings of the first workshop on digital fluid power DFP 2008, Tampere, pp 113–128
20. Linjama M (2011) Digital fluid power—state of the art. In: The Twelfth Scandinavian International conference on fluid power (SICFP), Tampere, pp 331–353
21. Lynch AF, Koch CB, Chladny R (2003) Nonlinear observer design for sensorless electromagnetic actuators. Dynamics of continuous discrete and impulsive systems-Series B-applications and algorithms, special issue, pp 317–322
22. Orlov Y (2009) Discontinuous systems: Lyapunov analysis and robust synthesis under uncertainty conditions. Communications and control engineering. Springer, London
23. Poznyak AS (2004) Deterministic output noise effects in sliding mode observation. In: Sabanovic A, Fridman LM, Spurgeon S (eds) Variable structure systems: from principles to implementation. IEE control engineering series vol 66. The Institution of Engineering and Technology (IET), London, pp 45–80
24. Reinertz O, Murrenhoff H (2009) Dynamic modeling of switching valves: an approach for one-dimensional simulation (in German). *O+P Ölhydraulik und Pneumatik* 16(4):152–155
25. Reuter J (2006) Flatness based control of a dual coil solenoid valve. In: 4th IFAC symposium on mechatronic systems. Ruprecht-Karls-University, Germany, pp 48–54
26. Shtessel Y, Edwards C, Fridman L, Levant A (2014) Sliding mode control and observation. Control engineering. Birkhäuser, Basel

27. Straußberger F, Schwab M, Braun T, Reuter J (2014) Position estimation in electro-magnetic actuators using a modified discrete time class a/b model reference approach. In: American control conference (ACC14), Portland, pp 3686–3691
28. Utkin V (1992) Sliding modes in control and optimization. Communications and control engineering. Springer, Berlin
29. Yaz E, Azemi A (1993) Variable structure observer with a boundary-layer for correlated noise/disturbance models and disturbance minimization. Int J Control 57(5):1191–1206
30. Zeitz M (1977) Nonlinear observers for chemical reactors (in German). Fortschrittberichte der VDI-Zeitschriften; Nr. 27. VDI-Verlag, Düsseldorf

Sliding Mode Observer for Fault Diagnosis: LPV and Takagi–Sugeno Model Approaches

Horst Schulte and Florian Pöschke

Abstract This chapter investigates recently proposed fault reconstruction methods by sliding mode observers defined by two different model classes: linear parameter varying and Takagi–Sugeno models. Both model classes are used to design the sliding mode observers. They may be considered as a polytopic extension of the canonical form restricted to uncertain linear time-invariant systems originally introduced by Edwards and Spurgeon. This approach is best suited for plants which can be thought of as predominantly linear in the characteristics or for nonlinear plants which can be modelled well (at least locally) by linear approximations. For highly nonlinear plants which are operated in a large operating range, a structure restricted to uncertain linear time-invariant systems is not ideal, as the sliding term would then have to capture both: the nonlinear plant dynamics and the influence of the faults. The chapter describes the observer design for linear parameter varying and Takagi–Sugeno models, which are illustrated by the means of the inverted pendulum and the wind turbine benchmark from the literature. Simulation results are shown to demonstrate the capability of the designed observers.

1 Introduction

In the last decades there has been an explosion of interest in sliding mode observers (SMOs) for fault detection and isolation (FDI), reconstruction and fault tolerant control (FTC). The sliding mode concept based on variable structure control algorithms can be used for controllers and observers. It includes a nonlinear switching term, which establishes and maintains a motion on a so-called sliding surface, where reduced-order dynamics compared to the normal system appears [25, 26].

For observers the sliding motion on the error between the output of the observer and the measured plant output ensures that a SMO produces a set of estimated states,

H. Schulte (✉) · F. Pöschke
Department of Engineering I, Control Engineering, HTW Berlin, Berlin, Germany
e-mail: horst.schulte@htw-berlin.de

F. Pöschke
e-mail: florian.poeschke@htw-berlin.de

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*,
Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_8

which are precisely corresponding with the current output of the plant. To achieve and maintain the sliding mode an injection signal has to be applied. The average value of this signal, the so-called equivalent injection signal, can be evaluated to yield direct fault estimates, because it contains useful information about the mismatch between the control-oriented model used to define the observer and the actual plant. The fault estimates or rather the reconstruction of the faults can be exploited for fault tolerant control in the sense that sensor and actuator faults are corrected before the measurements are used for the controller or rather the control signal acts on the plant.

Originally, the sliding mode method is based on linear time-invariant (LTI) systems with unknown but bounded terms [4]. An important extension to the sliding mode observer concept is to introduce a convex combination of LTI systems. The combination can be parameter varying and/or state varying. The parameter varying concept was presented in [1] and is based on a class of linear parameter varying (LPV) systems. In contrast, the state varying sliding mode concept was first proposed in [9], where the observer is implemented within a Takagi–Sugeno (TS) model structure to account for system nonlinearities. Both proposed approaches are extensions of the LTI scheme of sliding mode observer design to a convex combination of LTI systems, which is a suitable compromise between a full nonlinear design and the LTI framework.

In this chapter, first, we introduce TS and LPV model structures and describe them by means of an inverted pendulum and a wind turbine benchmark from the literature [14]. Then, the underlying relations between the model classes are exposed and discussed by comparing the necessary conditions and efficiency of the design process. Finally, simulation results of two case studies, the inverted pendulum [3] and a wind turbine benchmark from the literature [17], are presented to illustrate the influence of the underlying design model on the quality of the fault reconstruction and thus the quality of fault compensation.

Notation: Throughout this chapter, the notation $\|\cdot\|$ will be used to represent the Euclidean vector norm or its induced matrix norm. The identity and zero matrix of order n are represented by $\mathbf{I}_n$ and $\mathbf{0}_n$. $\mathbf{P} > 0$, ($\mathbf{P} < 0$) means that $\mathbf{P}$ is a positive (negative) definite matrix.

2 Model Structure

2.1 TS Model Structure

Takagi–Sugeno (TS) models provide a uniform framework for controller and observer design of nonlinear systems. Methods based on linear matrix inequalities (LMIs) using a Lyapunov function allow for a unified design for TS models [24, 27]. Introduced 30 years ago in the context of fuzzy systems [21], TS models are weighted combinations of linear submodels. These can either be derived from measured data using offline system identification techniques [20, 21] or from analytical models of nonlinear systems.

A general TS model structure is given by

$$\begin{aligned} \dot{\mathbf{x}} &= \sum_{i=1}^{N_r} h_i(\mathbf{z}) \left(\mathbf{A}_i \, \mathbf{x} + \mathbf{B}_i \, \mathbf{u} \right), \\ \mathbf{y} &= \sum_{i=1}^{N_r} h_i(\mathbf{z}) \, \mathbf{C}_i \, \mathbf{x}, \end{aligned} \tag{1}$$

where N_r denotes the number of linear state-space submodels, which share the common state vector $\mathbf{x} \in \mathbb{R}^n$, the input vector $\mathbf{u} \in \mathbb{R}^m$, and the output vector $\mathbf{y} \in \mathbb{R}^p$ with the constant matrices $\mathbf{A}_i \in \mathbb{R}^{n \times n}$, $\mathbf{B}_i \in \mathbb{R}^{n \times m}$ and $\mathbf{C}_i \in \mathbb{R}^{p \times n}$. The functions $h_i(\mathbf{z})$, $i \in \{1, \ldots, N_r\}$ are normalised to fulfil the convex sum condition

$$\sum_{i=1}^{N_r} h_i(\mathbf{z}) = 1 \,, \quad 0 \leq h_i(\mathbf{z}) \leq 1 \quad \forall \quad i \in \{1, \ldots, N_r\}. \tag{2}$$

The vector $\mathbf{z} \in \mathbb{R}^{N_l}$ of premise variables may comprise state variables $x_k \in \mathbb{R}$, inputs $u_k \in \mathbb{R}$, and external variables $\chi_k \in \mathbb{R}$: $\mathbf{z} = \mathbf{z}(\mathbf{x}, \mathbf{u}, \boldsymbol{\chi})$. To obtain a Takagi–Sugeno form and a Takagi–Sugeno (TS) structure of a nonlinear model respectively, and assuming the mathematical model is given by

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}), \quad \mathbf{y} = \mathbf{g}(\mathbf{x}), \tag{3}$$

there are two different methods to derive (1). The first method is based on an *approximation* of (3) by local Taylor linearisation of the nonlinear model around N_r stationary points and following fuzzy blending of the linear submodels $\{\mathbf{A}_i, \mathbf{B}_i, \mathbf{C}_i\}$ to a weighted sum according to (1) with the normalisation condition (2). The second method is the so-called sector nonlinearity approach [16, 23], which can be employed to obtain an *exact Takagi–Sugeno model representation* of a given nonlinear model (3).

In the following, we assume that the sector nonlinearity approach is used to yield a nonlinear system in Takagi–Sugeno's form. This has the advantage that the switching term in the sliding mode observer (presented in Sect. 4) must only account for the disturbances and faults and not for the approximation error caused by local linearisations and fuzzy blending.

2.2 LPV Model Structure

Linear parameter varying (LPV) descriptions of dynamic systems have been shown to be a powerful modelling approach for data-dependent systems, where the dependence might be governed by a nonlinear term, cf. [13]. The idea emerged from the

analysis of gain scheduling techniques and was presented by Shamma in [19]. In [13], Shamma describes the LPV modelling as a framework, which results in linear but non-stationary dynamics shown in (4):

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}(\boldsymbol{\theta})\mathbf{x} + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}, \\ \mathbf{y} &= \mathbf{C}(\boldsymbol{\theta})\mathbf{x}, \end{aligned} \tag{4}$$

where $\mathbf{x} \in \mathbb{R}^n$ are the system states, $\mathbf{u} \in \mathbb{R}^m$ the inputs and $\mathbf{y} \in \mathbb{R}^p$ describes the output of the system. $\boldsymbol{\theta}$ denotes the exogenous (i.e. state independent) varying parameters of the system leading to the non-stationary behaviour. However, this model of the system can be exploited for the description of nonlinear dynamic systems, where the nonlinear behaviour can either be induced by an exogenous or an endogenous (i.e. state-dependent) signal. Stability analysis and observer/controller synthesis have been addressed in literature, for example in [1, 2, 13]. To the authors' knowledge, however, the obtainment of a LPV model from a nonlinear system description is poorly described in a large quantity of the literature. For this reason, the next section will provide an approach for the construction of an LPV model out of an analytic nonlinear system description.

The basic idea of the LPV approach is to bring a nonlinear function z or an exogenous signal under the exploitation of the knowledge of the range bounds of functions values $\overline{z} = \sup\{z\}$ and $\underline{z} = \inf\{z\}$ into a certain structure. Consider the nonlinear function z, which is unknown but bounded, described by

$$z = c_1 + \theta\, c_2, \tag{5}$$

where $-1 \leq \theta \leq 1$ holds and c_1, c_2 are constant values depending on the upper $\overline{z}$ and lower $\underline{z}$ bound of z. By the choice of $c_1 = \underline{z} + \frac{\Delta z}{2}$, $c_2 = \frac{\Delta z}{2}$ and $\theta = \frac{z - \underline{z} - \Delta z/2}{\Delta z/2}$ with the definition of $\Delta z = \overline{z} - \underline{z}$ the aforementioned conditions are fulfilled and it can be easily verified that

$$\underbrace{\underline{z} + \frac{\Delta z}{2}}_{=:c_1} + \underbrace{\frac{z - \underline{z} - \Delta z/2}{\Delta z/2}}_{=:\theta} \underbrace{\frac{\Delta z}{2}}_{=:c_2} = z. \tag{6}$$

Since $-1 \leq \theta \leq 1$ holds, a convex form is achieved. Consider a nonlinear system of the form

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}(z)\mathbf{x} + \mathbf{B}\mathbf{u}, \\ \mathbf{y} &= \mathbf{C}\mathbf{x}, \end{aligned} \tag{7}$$

where z can either be an exogenous signal/function or an endogenous nonlinear function depending on a state of the system. Let $\mathbf{A}$ be defined by

$$\mathbf{A}(z) = \begin{bmatrix} a_{11} + z & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Then, according to the scheme described before, the system can be described in the LPV structure

$$\mathbf{A}(z) = \begin{bmatrix} a_{11} + \underbrace{c_1 + \theta\, c_2}_{z} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + c_1 & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \theta \begin{bmatrix} c_2 & 0 \\ 0 & 0 \end{bmatrix}$$
$$= \mathbf{A}_0 + \theta\, \mathbf{A}_1 = \mathbf{A}(\theta). \tag{8}$$

If nonlinear functions or exogenous signals appear in **B** or **C**, the way for achieving a LPV description can be applied analogously. Since the formulation leads to a convex combination of matrices, this can be exploited in the synthesis of controllers or observers accordingly.

3 Case Studies: Modelling

3.1 Case Study I: Inverted Pendulum

3.1.1 Physical Model

The inverted pendulum benchmark, in particular the cart version illustrated in Fig. 1, has been considered in many references to solve the problem of designing controllers around the unstable operating point or as a nonlinear control problem in the full operating range of $\Theta \in [\,0,\ \pi\,)$ [12]. In this chapter, we use the latter case to be able to use advantageously the LPV and TS sliding mode observer approach.

The cart with inverted pendulum consists of a moveable carriage with one degree of freedom on which a pendulum is mounted and freely rotatable in driving direction (Fig. 1). The carriage is driven by a motor that exerts a force F through a belt-drive transmission. The main control objective is to swing up the pendulum from the stable equilibrium to the unstable equilibrium, and then balance the pendulum at the upright position, and further move the cart to a specified position along the track.

The process can be described by the state-space model

$$\dot{\mathbf{x}} = \begin{pmatrix} \dot{\Theta} \\ \dfrac{g\,\sin(\Theta) - m\,l\,a\,\dot{\Theta}^2\cos(\Theta)\,\sin(\Theta) - a\,\cos(\Theta)\left(u - F_f\right) - \frac{d_{Mf}\,\dot{\Theta}}{m\,l}}{2\,l - m\,l\,a\,\cos^2(\Theta)} \\ \dot{x} \\ \dfrac{2a\left(m\,l\,\dot{\Theta}^2\,\sin(\Theta) - \frac{1}{2}\,m\,g\,\cos(\Theta)\,\sin(\Theta) + u - F_f + \frac{1}{2l}\cos(\Theta)\,d_{Mf}\,\dot{\Theta}\right)}{2 - m\,a\,\cos^2(\Theta)} \end{pmatrix} \tag{9}$$

with the state vector $\mathbf{x} = \left[\ \Theta\ \ \dot{\Theta}\ \ x\ \ \dot{x}\ \right]^{\mathrm{T}}$, where $a = 1/(m + M)$, F_{f} denotes the unknown but bounded friction force between the cart and the track and g denotes the

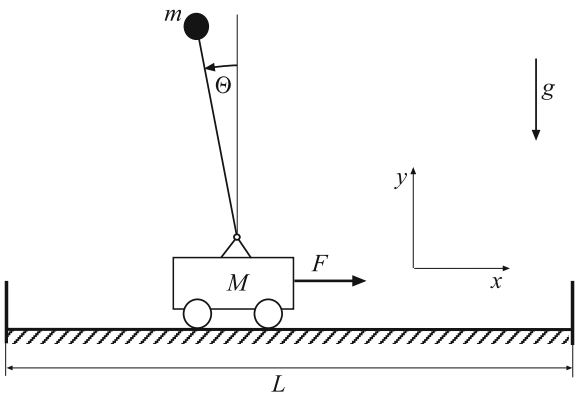


Fig. 1 Cart with inverted pendulum

gravitational acceleration. The external force F is the input $u := F$ and is limited to $|F| \leq F_{\max}$. The angular position of pendulum Θ, the linear displacement x, and the speed of the cart $\dot{x}$ are the outputs of this system. The parameters and variables of the inverted pendulum on the cart are recorded in Table 1.

3.1.2 TS Model

In order to prepare for the observer design, a TS model is derived using the sector nonlinearity approach [22]. First, we rewrite (9) in the form

Table 1 Variables and parameters of cart with inverted pendulum

Symbol	Description	Value	Unit
Θ	Angular position of the pendulum	–	rad
x	Linear displacement of the cart	–	m
u	Input: external force on the cart	–	N
m	Point mass of the pendulum	0.356	kg
M	Mass of the cart	4.8	kg
l	Distance from the joint to the mass point m	0.56	m
d_{Mf}	Viscous friction of the joint	0.035	Nms/rad
F_c	Coulomb friction coefficient	4.9	N
g	Gravitational constant	9.81	m/s^2
L	Total length of rail	2	m
$F_{\max}$	Maximum input value (actuator saturation)	120	N

$$\dot{\mathbf{x}} = \underbrace{\begin{bmatrix} 0 & 1 & 0\,0 \\ f_1\, f_2(\frac{g}{l} - m\,a\, f_3\, f_4) & -\frac{d_{Mf}}{ml^2}\, f_1 & 0\,0 \\ 0 & 0 & 0\,1 \\ m\,a\, f_1\, f_2\,(2\,l\, f_3 - g\, f_4) & \frac{a\,d_{Mf}}{l}\, f_1\, f_4 & 0\,0 \end{bmatrix}}_{\mathbf{A}(\mathbf{x})} \mathbf{x} + \underbrace{\begin{bmatrix} 0 \\ -\frac{a}{l}\, f_1\, f_4 \\ 0 \\ 2\,a\, f_1 \end{bmatrix}}_{\mathbf{B}(\mathbf{x})} \big(u - F_f\big). \quad (10)$$

The scalar-valued nonlinear functions f_j, $j \in \{1, \ldots, N_l\}$ in (10) are given by

$$\begin{aligned} f_1(x_1) &= \frac{1}{2 - m\,a\,\cos^2(x_1)}, & f_2(x_1) &= \frac{\sin(x_1)}{x_1}, \\ f_3(x_2) &= x_2^2, & f_4(x_1) &= \cos(x_1), \end{aligned} \quad (11)$$

where $N_l = 4$ denotes the number of nonlinearities. For the transition to the TS model structure (1), these functions are written in a different form. Let $\underline{f}_j$ and $\overline{f}_j$ denote the minimum and maximum values of each function f_j. The following identities hold:

$$\begin{aligned} f_j\,(x_1) &= w_{j1}\,(x_1)\ \overline{f}_j + w_{j2}\,(x_1)\ \underline{f}_j, \quad j \in \{1, 2, 4\} \\ f_3\,(x_2) &= w_{31}\,(x_2)\ \overline{f}_3 + w_{32}\,(x_2)\ \underline{f}_3, \quad j = 3, \end{aligned} \quad (12)$$

where the weighting functions w_{jk}, $k = 1, 2$, are given by

$$\begin{aligned} w_{j1}\,(x_1) &:= \frac{f_j\,(x_1) - \underline{f}_j}{\overline{f}_j - \underline{f}_j}, & w_{j2}\,(x_1) &:= \frac{\overline{f}_j - f_j\,(x_1)}{\overline{f}_j - \underline{f}_j}, & j &\in \{1, 2, 4\} \\ w_{31}\,(x_2) &:= \frac{f_3\,(x_2) - \underline{f}_3}{\overline{f}_3 - \underline{f}_3}, & w_{32}\,(x_2) &:= \frac{\overline{f}_3 - f_3\,(x_2)}{\overline{f}_3 - \underline{f}_3}, & j &= 3, \end{aligned} \quad (13)$$

which satisfy the property $w_{j1} + w_{j2} = 1$. From this property and by defining the sum of the membership functions as the product of the convex sums of weighting functions w_{jk}, we obtain the definition

$$\sum_{i=1}^{N_r} h_i(\mathbf{z}) := \prod_{j=1}^{N_l=4} \big(w_{j1} + w_{j2}\big), \quad (14)$$

where it directly follows that $\sum_{i=1}^{N_r} h_i(\mathbf{z}) = 1$, i.e. the convex sum condition (2) holds. In order to replace the nonlinear functions f_j in (10) by (12) these are multiplied first by an appropriate choice of the convex sum

$$f_j\,(x_2) = \big(w_{j1}\,\overline{f}_j + w_{j2}\,\underline{f}_j\big) \prod_{i \neq j} \big(\underbrace{w_{i1} + w_{i2}}_{=1}\big), \quad i \in \{1, 2, 3, 4\}. \quad (15)$$

Using this formulation of f_j, the nonlinearities have been shifted to the membership functions, which are described in this case study by

$$\begin{aligned} h_1(x_1, x_2) &= w_{11}(x_1)\; w_{21}(x_1)\; w_{31}(x_2)\; w_{41}(x_1) \\ h_2(x_1, x_2) &= w_{12}(x_1)\; w_{21}(x_1)\; w_{31}(x_2)\; w_{41}(x_1) \\ h_3(x_1, x_2) &= w_{11}(x_1)\; w_{22}(x_1)\; w_{31}(x_2)\; w_{41}(x_1) \\ h_4(x_1, x_2) &= w_{12}(x_1)\; w_{22}(x_1)\; w_{31}(x_2)\; w_{41}(x_1) \\ &\vdots \qquad\qquad\quad \vdots \\ h_{16}(x_1, x_2) &= w_{12}(x_1)\; w_{22}(x_1)\; w_{32}(x_2)\; w_{42}(x_1). \end{aligned} \tag{16}$$

All other constant matrix entries of (10) are multiplied by $\sum_{i=1}^{N_r} h_i(\mathbf{z}) = 1$, such that $\mathbf{A}(\mathbf{x})$ and $\mathbf{B}(\mathbf{x})$ can be written as a weighted sum of $N_r = 16$ matrices with constant coefficients according to

$$\mathbf{A}(\mathbf{x}) = \sum_{i=1}^{N_r} h_i(\mathbf{z})\, \mathbf{A}_i, \qquad \mathbf{B}(\mathbf{x}) = \sum_{i=1}^{N_r} h_i(\mathbf{z})\, \mathbf{B}_i, \tag{17}$$

where $\mathbf{z} = \left[\, x_1,\; x_2 \right]^{\mathrm{T}}$. Thus, the matrices of the first submodel $i = 1$ are

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \overline{f}_1\, \overline{f}_2 (\frac{g}{l} - m\, a\, \overline{f}_3\, \overline{f}_4) & -\frac{d_{Mf}}{m l^2}\, \overline{f}_1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ m\, a\, \overline{f}_1\, \overline{f}_2\, (2\, l\, \overline{f}_3 - g\, \overline{f}_4) & \frac{a\, d_{Mf}}{l}\, \overline{f}_1\, \overline{f}_4 & 0 & 0 \end{bmatrix}, \qquad \mathbf{B}_1 = \begin{bmatrix} 0 \\ -\frac{a}{l}\, \overline{f}_1\, \overline{f}_4 \\ 0 \\ 2\, a\, \overline{f}_1 \end{bmatrix}.$$

The state-space model (9) can thus be transformed into the TS form (1)

$$\begin{aligned} \dot{\mathbf{x}} &= \sum_{i=1}^{N_r = 16} h_i(\mathbf{z}) \left(\mathbf{A}_i\, \mathbf{x} + \mathbf{B}_i\, (u - F_f) \right), \\ \mathbf{y} &= \mathbf{C}\, \mathbf{x}. \end{aligned} \tag{18}$$

Note that the friction force F_f is here considered as an additional unknown input.

3.1.3 LPV Model

Consider the nonlinear model of the inverted pendulum presented in Sect. 3.1.1. The nonlinear model can be described by

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{g}{l} z_1 - m\,a\,z_2 & -\frac{d_{Mf}}{ml^2} z_3 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 2mlaz_5 - magz_6 & \frac{ad_{Mf}}{l} z_4 & 0 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ -\frac{a}{l} z_4 \\ 0 \\ 2az_3 \end{bmatrix} (u - F_f), \tag{19}$$

where the nonlinear functions of the model are substitutions using (11) $z_1 = f_1 f_2$, $z_2 = f_1 f_2 f_3 f_4$, $z_3 = f_1$, $z_4 = f_1 f_4$, $z_5 = f_1 f_2 f_3$, and $z_6 = f_1 f_2 f_4$. Based on the knowledge of the individual bounds of the nonlinear functions f_i, the individual bounds of the nonlinear functions $z_i = c_{1i} + \theta_i c_{2i}$ can be obtained, cf. Table 2. Using the pattern described in Sect. 2.2 and especially the separation of constant matrices and parameter varying components as shown in (8), the LPV model of the inverted pendulum can be described in the form

$$\dot{\mathbf{x}} = \left(\mathbf{A}_0 + \sum_{i=1}^{6} \theta_i \mathbf{A}_i \right) \mathbf{x} + \left(\mathbf{B}_0 + \sum_{i=3}^{4} \theta_i \mathbf{B}_i \right) u, \tag{20}$$

with, for example,

$$\mathbf{A}_0 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{g}{l} c_{11} - mac_{12} & -\frac{d_{Mf}}{ml^2} c_{13} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 2mlac_{15} - magc_{16} & \frac{ad_{Mf}}{l} c_{14} & 0 & 0 \end{bmatrix} \text{ and } \mathbf{A}_1(\theta_1) = \theta_1 \begin{bmatrix} 0 & \\ \frac{g}{l} c_{21} & \mathbf{0}_{4\times 3} \\ 0 & \\ 0 & \end{bmatrix}.$$

The sliding mode observer design for the inverted pendulum based on the LPV problem description—contrary to the TS model—requires a model approximation because of the structure of the fault distribution matrix $\mathbf{F}(\boldsymbol{\theta})$, which is introduced later in Sect. 4. As proposed by Alwi and Edwards in [1], the distribution matrix can be factorised into

Table 2 Lower and upper bounds of the inverted pendulum

Function	Calculation	Upper bound $\overline{f}_i, \overline{z}_i$	Lower bound $\underline{f}_i, \underline{z}_i$
f_1	$1/(2 - ma\cos(\Theta)^2)$	$1/(2 - ma)$	$1/2$
f_2	$\sin(\Theta)/\Theta$	1	-0.22
f_3	$\dot{\Theta}^2$	36π	0
f_4	$\cos(\Theta)$	1	-1
z_1	$f_1 f_2$	$\overline{f}_1 \overline{f}_2$	$\overline{f}_1 \underline{f}_2$
z_2	$f_1 f_2 f_3 f_4$	$\overline{f}_1 \overline{f}_2 \overline{f}_3 \overline{f}_4$	$\overline{f}_1 \overline{f}_2 \overline{f}_3 \underline{f}_4$
z_3	f_1	$\overline{f}_1$	$\underline{f}_1$
z_4	$f_1 f_4$	$\overline{f}_1 \overline{f}_4$	$\overline{f}_1 \underline{f}_4$
z_5	$f_1 f_2 f_3$	$\overline{f}_1 \overline{f}_2 \overline{f}_3$	$\overline{f}_1 \underline{f}_2 \overline{f}_3$
z_6	$f_1 f_2 f_4$	$\overline{f}_1 \overline{f}_2 \overline{f}_4$	$\overline{f}_1 \overline{f}_2 \underline{f}_4$

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbf{H}\mathbf{E}(\boldsymbol{\theta}), \tag{21}$$

where $\mathbf{H} \in \mathbb{R}^{n\times q}$ is fixed and a virtual fault is introduced by $\mathbf{f}_v = \mathbf{E}(\boldsymbol{\theta})\mathbf{f}$. $\mathbf{E} \in \mathbb{R}^{q\times q}$ varies dependent on the nonlinear terms and is assumed to be invertible. Thus, from the knowledge of $\mathbf{f}_v$, $\mathbf{f}$ can be calculated by $\mathbf{f} = \mathbf{E}(\boldsymbol{\theta})^{-1}\mathbf{f}_v$. Anyhow, in case of the inverted pendulum the distribution matrix $\mathbf{F}(z_3,\ z_4) = \left[0\ \frac{a}{l}z_4\ 0\ -2az_3\right]^{\mathrm{T}} = \mathbf{F}(\boldsymbol{\theta})$ cannot be factorised due to the fact that two nonlinear functions are included in one vector. A possible solution to this problem is the allocation of the distribution vector into a matrix

$$\tilde{\mathbf{F}}(z_3,\ z_4) = \begin{bmatrix} 0 & 0 \\ 0 & \frac{a}{l}z_4 \\ 0 & 0 \\ -2az_3 & 0 \end{bmatrix}, \tag{22}$$

where the presented factorisation can be accomplished by the matrix $\mathbf{E}(z_3,\ z_4) = \begin{bmatrix} z_3 & 0 \\ 0 & z_4 \end{bmatrix}$, leading to a new distribution matrix $\tilde{\mathbf{H}}$ for the virtual fault $\mathbf{f}_v = \mathbf{E}(z_3,\ z_4)\begin{bmatrix} f_F \\ f_F \end{bmatrix}$

$$\tilde{\mathbf{H}} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{a}{l} \\ 0 & 0 \\ -2a & 0 \end{bmatrix}. \tag{23}$$

This factorisation violates the necessary assumption from Sect. 4 $\operatorname{rank}(\tilde{\mathbf{H}}) = q = \operatorname{rank}(\mathbf{C}\,\tilde{\mathbf{H}})$ in the design process. For this reason, a model approximation has to be deployed to the LPV model of the inverted pendulum. The new distribution matrix is set to $\mathbf{F}(z_3)_{\text{ap.}} = \left[0\ 0\ 0\ -2az_3\right]^{\mathrm{T}} = \mathbf{F}(\theta_3)$. Then, the factorisation can be applied using $\mathrm{E}(\theta_3) = \mathrm{E}(z_3) = z_3$, the virtual fault $f_v = \mathrm{E}(z_3)f_F$ and the distribution matrix for the virtual fault

$$\mathbf{H}_{\text{ap.}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2a \end{bmatrix}, \tag{24}$$

which is used during the observer design. The observer reconstructs the virtual fault $\hat{f}_v$. From this reconstruction, the estimated occurring fault $\hat{f} = \mathrm{E}(z_3)^{-1}\,\hat{f}_v$ can be calculated.

3.2 *Case Study II: Wind Turbine*

3.2.1 Physical Model

In the second case study, we use a wind turbine benchmark model presented in [14], which describes a generic pitch-controlled three-blade horizontal variable-

speed wind turbine with a rated power of $P = 4.8\,\text{MW}$. The original purpose of this benchmark was to provide a model on which researchers, who are working in the field of fault diagnosis and fault tolerant control, can compare different FDI/FTC methods to a wind turbine.

The model consists of four submodels: The mechanical submodel, which is reduced to the drive train dynamics, the aerodynamics, the pitch actuators and the generator-converter dynamics. The coupling of these submodels is illustrated in Fig. 2. The mechanical submodel with two degrees of freedom (rotor and generator rotation) is described by the motion equation

$$\begin{aligned} J_r\,\dot{\omega}_r &= T_r - K_{dt}\,\theta_s - (B_{dt} + B_r)\omega_r + \frac{B_{dt}}{N_g}\,\omega_g, \\ J_g\,\dot{\omega}_g &= \frac{\eta_{dt}\,K_{dt}}{N_g}\,\theta_s + \frac{\eta_{dt}\,B_{dt}}{N_g}\,\omega_r - \left(\frac{\eta_{dt}\,B_{dt}}{N_g^2} + B_g\right)\omega_g - T_g, \end{aligned} \tag{25}$$

where θ_s denotes the shaft torsion angle with the angular velocity $\dot{\Theta}_s = \omega_r - \frac{1}{N_g}\omega_g$. For sake of clarity, all parameters and system variables are summarised in Table 3. The aerodynamic submodel comprises the expression for the rotor torque T_r. This torque depends on the aero map C_Q for the torque coefficient [5]

$$T_r = \frac{1}{2}\rho\,\pi\,R^3\,v^2\,\frac{1}{3}\sum_{i=1}^{3} C_Q(\lambda, \beta_i), \tag{26}$$

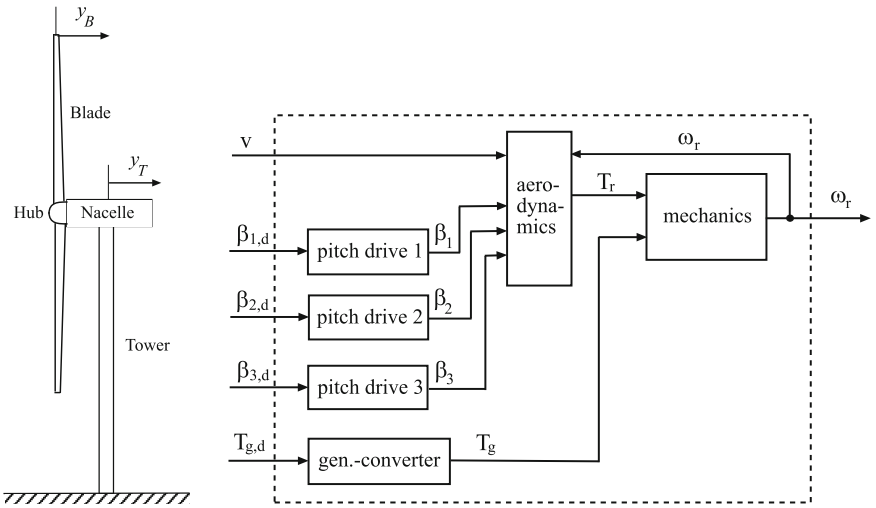


Fig. 2 Schematic side-view and submodels of the complete wind turbine benchmark model with the respective inputs and outputs

Table 3 Variables and parameters of the wind turbine benchmark

Symbol	Description	Value	Unit
T_g	Applied generator torque	–	Nm
$T_{g,d}$	Demanded generator torque	–	Nm
T_r	Aerodynamic rotor torque	–	Nm
v	Wind speed	–	m/s
β_i	Pitch angle of blade $i = 1, 2, 3$	–	rad
$\beta_{i,d}$	Demanded pitch angle of blade $i = 1, 2, 3$	–	rad
λ	Tip speed ratio	–	–
ω_g	Generator angular velocity (generator speed)	–	rad/s
ω_r	Rotor angular velocity (rotor speed)	–	rad/s
θ_g	Generator rotational angle	–	rad
θ_r	Rotor rotational angle	–	rad
θ_s	Shaft torsion angle	–	rad
B_{dt}	Torsional damping coefficient of the drive train	775.49	Nms/rad
B_g	Viscous friction of the high speed shaft	45.6	Nms/rad
B_r	Viscous friction of the low speed shaft	7.11	Nms/rad
C_Q	Aerodynamic rotor torque coefficient	–	–
J_g	Generator inertia	390	$\mathrm{kg\,m^2}$
J_r	Rotor inertia	55×10^6	$\mathrm{kg\,m^2}$
K_{dt}	Torsional stiffness of the drive train	2.7×10^9	Nm/rad
N_g	Gearbox ratio	95	–
R	Rotor radius	57.5	m
η_{dt}	Efficiency of the drive train	0.97	–
ρ	Air density	1.225	$\mathrm{kg/m^3}$
τ_g	Delay time constant for generator-converter dynamics	0.02	s
ω_n	Natural frequency parameter for second-order pitch dynamics	11.11	rad/s
ζ	Damping constant for second-order pitch dynamics	0.6	–

where ρ denotes the air density, R is the rotor radius, v is the average wind speed at the rotor, β_i describes the individual pitch angle of blade $i \in \{1, 2, 3\}$ and $\lambda = \omega_r \; R/v$ characterises the tip speed ratio.

In pitch-controlled wind turbines, the pitch angles of the blades are altered in the full load region to keep the wind turbine at the desired nominal rotor speed by reducing the aerodynamic rotor torque. The blade pitch system with three single pitch drives is modelled by the second-order delay systems [14]

$$\ddot{\beta}_i + 2\,\zeta\,\omega_n\,\dot{\beta}_i + \omega_n^2\,\beta_i = \omega_n^2\,\beta_{i,d}, \tag{27}$$

where β_i and $\beta_{i,d}$ denote the actual and demanded pitch angles of blade i. To take the generator-converter dynamics into account, and for the purpose of actuator fault detection, a first-order delay model is used in [14]

$$\dot{T}_g = -\frac{1}{\tau_g} T_g + \frac{1}{\tau_g} T_{g,d}, \tag{28}$$

where T_g and $T_{g,d}$ denote the applied and demanded generator torque.

3.2.2 TS Model

As in the first case study (inverted pendulum), a TS model is derived using the sector nonlinearity approach [22]. Equations (25)–(28) are combined to a state-space model with the state and input vector

$$\mathbf{x} = \begin{bmatrix} \omega_r & \omega_g & \theta_s & T_g & \dot{\beta}_1 & \beta_1 & \dot{\beta}_2 & \beta_2 & \dot{\beta}_3 & \beta_3 \end{bmatrix}^{\mathrm{T}}, \qquad \mathbf{u} = \begin{bmatrix} T_{g,d} & \beta_{1,d} & \beta_{2,d} & \beta_{3,d} \end{bmatrix}^{\mathrm{T}}$$

and rewritten in the form

$$\dot{\mathbf{x}} = \underbrace{\begin{bmatrix} \tilde{\mathbf{A}}_{11}(x_1, v) & \mathbf{0}_{4\times2} & \mathbf{0}_{4\times2} & \mathbf{0}_{4\times2} \\ \mathbf{0}_{2\times4} & \tilde{\mathbf{A}}_{22} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} \\ \mathbf{0}_{2\times4} & \mathbf{0}_{2\times2} & \tilde{\mathbf{A}}_{33} & \mathbf{0}_{2\times2} \\ \mathbf{0}_{2\times4} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \tilde{\mathbf{A}}_{44} \end{bmatrix}}_{\mathbf{A}(x_1,v)} \mathbf{x} + \underbrace{\begin{bmatrix} & \mathbf{0}_{3\times4} & \\ \frac{1}{\tau_g} & 0 & 0 & 0 \\ 0 & \omega_n^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \omega_n^2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_n^2 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{B}} \mathbf{u}, \tag{29}$$

with

$$\tilde{\mathbf{A}}_{11}(x_1, v) = \begin{bmatrix} f_1(x_1, v) & \frac{B_{dt}}{N_g J_r} & -\frac{K_{dt}}{J_r} & 0 \\ \frac{\eta_{dt} B_{dt}}{N_g J_g} & -\left(\frac{\eta_{dt} B_{dt}}{N_g^2 J_g} + \frac{B_g}{J_g}\right) & \frac{\eta_{dt} K_{dt}}{N_g J_g} & -\frac{1}{J_g} \\ 1 & -\frac{1}{N_g} & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{\tau_g} \end{bmatrix}, \quad \tilde{\mathbf{A}}_{jj} = \begin{bmatrix} -2\,\omega_n\,\zeta & -\omega_n^2 \\ 1 & 0 \end{bmatrix}$$

for $j \in \{2, 3, 4\}$. The scalar-valued nonlinear function f_1 is given by

$$f_1(x_1, v) = -\frac{1}{J_r}\Big(B_{dt} + B_r\Big) + \frac{1}{\omega_r}\,\frac{1}{J_r}\,T_r\big(\lambda(x_1, v), v\big), \quad x_1 \in [\,\underline{x}_1\,,\,\overline{x}_1\,], \tag{30}$$

where $\underline{x}_1 > 0$ and $\overline{x}_1$ denote the lower and upper bounds of the rotor speed in the wind turbine operating range. For the transition to the TS model structure (1) these

functions are written in a different form. Let $\underline{f}_1$ and $\overline{f}_1$ denote the minimum and maximum values of the function f_1. The identity

$$f_1(x_1, v) = w_{11}(x_1, v)\,\overline{f}_1 + w_{12}(x_1, v)\,\underline{f}_1 \tag{31}$$

holds, where the weighting functions w_{1k}, $k \in \{1, 2\}$, are given by

$$w_{11}(x_1, v) := \frac{f_1(x_1, v) - \underline{f}_1}{\overline{f}_1 - \underline{f}_1}, \quad w_{12}(x_1, v) := \frac{\overline{f}_1 - f_1(x_1, v)}{\overline{f}_1 - \underline{f}_1}. \tag{32}$$

Using this formulation (31), the nonlinearity has been shifted to the membership functions, which in this second case study are described by

$$h_1(x_1, v) = w_{11}(x_1, v), \qquad h_2(x_1, v) = w_{12}(x_1, v). \tag{33}$$

All other constant matrix entries of (29) are multiplied by $\sum_{i=1}^{N_r} h_i(x_1, v) = 1$, such that $\mathbf{A}(x_1, v) = \mathbf{A}(z)$ can be written as a weighted sum of $N_r = 2$ matrices with constant coefficient:

$$\mathbf{A}(z) = \sum_{i=1}^{N_r=2} h_i(z)\,\mathbf{A}_i,$$

where $z = f_1(x_1, v)$. The state-space model (29) can thus be transformed into the TS form (1)

$$\dot{\mathbf{x}} = \sum_{i=1}^{N_r=2} h_i(z)\left(\mathbf{A}_i\,\mathbf{x} + \mathbf{B}\,\mathbf{u}\right). \tag{34}$$

3.2.3 LPV Model

Consider the nonlinear system of the wind turbine benchmark, where the nonlinear function $z = T_r/x_1$ is integrated into the system matrix $\mathbf{A}$ as follows:

$$\mathbf{A}(x_1, T_r) = \begin{bmatrix} -\frac{B_{dt}+B_r}{J_r} + \frac{T_r}{x_1 J_r} & a_{12} & a_{13}\,\mathbf{0}_{1\times 7} \\ & \mathbf{A}_{(\mathrm{n\text{-}1})\times \mathrm{n}} & \end{bmatrix}, \tag{35}$$

where a_{12}, a_{13} and $\mathbf{A}_{(\mathrm{n\text{-}1})\times \mathrm{n}}$, $\mathrm{n} = 10$ can be found in the descriptions in Sect. 3.2.2. Just as described in Sect. 2.2, the nonlinear function $\frac{T_r}{x_1}$ can be represented by an affine term θ and two constants resulting in a linear function by the use of its bounds. Note that this model is valid for the operation range of the wind turbine, where it is assumed that $x_1 > 0$. The bounds of the nonlinear function $\frac{T_r}{x_1}$ are $\overline{z} = \sup\{z\} = 1.5458 \times 10^{10}\ \frac{\mathrm{Nms}}{\mathrm{rad}}$ and $\underline{z} = \inf\{z\} = -3.8538 \times 10^{11}\ \frac{\mathrm{Nms}}{\mathrm{rad}}$, which are obtained from the information of

the operating range of the wind turbine and the knowledge of the specific torque coefficients of the benchmark model presented in [15]. Based on these bounds, the description of the dynamics of the wind turbine can be altered to a LPV model of the form

$$\begin{aligned}\dot{\mathbf{x}} &= \underbrace{\left(\begin{bmatrix} -\frac{B_{dt}+B_r}{J_r} + \frac{c_1}{J_r} & a_{12} & a_{13}\ \mathbf{0}_{1\times 7} \\ & \mathbf{A}_{(\mathrm{n-1})\times \mathrm{n}} & \end{bmatrix} + \theta \begin{bmatrix} \frac{c_2}{J_r} & \mathbf{0}_{1\times 9} \\ & \mathbf{0}_{(\mathrm{n-1})\times \mathrm{n}} \end{bmatrix}\right)}_{\mathbf{A}(\theta)} \mathbf{x} + \mathbf{B}\mathbf{u} \\ &= \mathbf{A}(\theta)\mathbf{x} + \mathbf{B}\mathbf{u}, \quad \mathbf{y} = \mathbf{C}\mathbf{x}, \end{aligned} \tag{36}$$

where the coefficients are again defined as $c_1 = \underline{z} + \frac{\Delta z}{2}$, $c_2 = \frac{\Delta z}{2}$ and $\theta = \frac{z-\underline{z}-\Delta z/2}{\Delta z/2}$.

4 Fault Reconstruction by Sliding Mode Observation

4.1 System Description and Canonical Form

We consider two different norm-bounded uncertain model structures: The first is based on the previously introduced TS model (1)

$$\begin{aligned}\dot{\mathbf{x}} &= \sum_{i=1}^{N_r} h_i(\mathbf{z}) \left(\mathbf{A}_i\, \mathbf{x} + \mathbf{B}_i\, \mathbf{u} + \mathbf{F}_i\, \mathbf{f}_a\right), \\ \mathbf{y} &= \mathbf{C}\, \mathbf{x}, \end{aligned} \tag{37}$$

where $\mathbf{F}_i \in \mathbb{R}^{n\times a}$ denotes the fault distribution matrix and the faults are presented by $\mathbf{f}_a \in \mathbb{R}^a$. The common $\mathbf{C}$ in (37) is only a small restriction, since many applications [7, 12, 18] comprise outputs that are linear in the system states. The second exploits the LPV modelling techniques based on the introduced structure (4)

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}(\boldsymbol{\theta})\, \mathbf{x} + \mathbf{B}(\boldsymbol{\theta})\, \mathbf{u} + \underbrace{\mathbf{F}(\boldsymbol{\theta})\mathbf{f}_a}_{\mathbf{H}\,\mathbf{E}(\boldsymbol{\theta})\mathbf{f}_a}, \\ \mathbf{y} &= \mathbf{C}\, \mathbf{x}, \end{aligned} \tag{38}$$

where $\mathbf{f}_v = \mathbf{E}(\boldsymbol{\theta})\mathbf{f}_a$ denotes the virtual fault vector with the special form of the distribution matrix (cf. Sect. 3.1.3). Note that the special form of the distribution matrix and the virtual fault is only needed if the original distribution matrix depends on an affine term θ_i. This is due to the transformations discussed in the following part of this discourse. Otherwise, this allocation can be ignored.

For the observer design, three *existence conditions* have to be fulfilled [4, 9, 10]:

- Condition 1: The faults presented in (37) and (38) are unknown but norm bounded by known positive constants $\Xi_{f_a} \in \mathbb{R}^q$ which satisfy $\|\mathbf{f}_a(t)\| \leq \Xi_{f_a}$. Moreover, the system states and inputs are assumed to be bounded.
- Condition 2:
  - TS (37): Let $q_i = q \ \forall i$ be defined as the number of columns of $\mathbf{F}_i$. Then the condition $q = \text{rank}(\mathbf{C}\,\mathbf{F}_i) = \text{rank}(\mathbf{F}_i)$ must be fulfilled and it must hold that $p > q$, where p is the number of measurable system states.
  - LPV (38): Let q be defined as the number of columns of $\mathbf{H}$. Then the condition $q = \text{rank}(\mathbf{C}\,\mathbf{H}) = \text{rank}(\mathbf{H})$ must be fulfilled and it must hold that $p > q$, where p is the number of measurable system states.
- Condition 3:
  - TS (37): All invariant zeros of $(\mathbf{A}_i, \mathbf{F}_i, \mathbf{C})$ must lie in $\mathbb{C}_-$, which denotes the open left half of the complex plane.
  - LPV (38): All invariant zeros of $(\mathbf{A}(\boldsymbol{\theta}), \mathbf{H}, \mathbf{C})$ must lie in $\mathbb{C}_-$.

The design of a sliding mode observer for fault reconstruction that is applicable for the classes of TS systems (37) or LPV systems (38) is carried out in a special canonical form. With a series of transformations $\mathbf{T}_i$ for each TS submodel

$$\mathbf{T}_i = \mathbf{T}_{L,i}\ \mathbf{T}_{b,i}\ \mathbf{T}_c, \tag{39}$$

or for the LPV model with the common distribution matrix $\mathbf{H}$

$$\mathbf{T} = \mathbf{T}_L\ \mathbf{T}_b\ \mathbf{T}_c, \tag{40}$$

the TS respectively LPV system is brought into a structure where, first, the last p states of the systems are the outputs $\mathbf{y}$ and, second, the faults $\mathbf{f}_a$ only act on the measurable system states (for further details such as the description of the transformation matrices (39) and (40) and proofs see [4, 6, 9]).

The **TS model in canonical form** is described by

$$\begin{aligned} \dot{\mathbf{x}}_1 &= \sum_{i=1}^{N_r} h_i(\mathbf{z}) \left(\mathscr{A}_{11,i}\,\mathbf{x}_1 + \mathscr{A}_{12,i}\,\mathbf{y} + \mathscr{B}_{1,i}\,\mathbf{u} \right), \\ \dot{\mathbf{y}} &= \sum_{i=1}^{N_r} h_i(\mathbf{z}) (\mathscr{A}_{21,i}\mathbf{x}_1 + \mathscr{A}_{22,i}\mathbf{y} + \mathscr{B}_{2,i}\,\mathbf{u} + \mathscr{F}_{2,i}\mathbf{f}_a), \end{aligned} \tag{41}$$

with the non-measurable states $\mathbf{x}_1 \in \mathbb{R}^{(n-p)}$ and the measurable states $\mathbf{y} \in \mathbb{R}^p$. The transformed system matrices in (41) have the following block structures:

$$\mathscr{A}_i = \mathbf{T}_i\,\mathbf{A}_i\,\mathbf{T}_i^{-1} = \begin{bmatrix} \mathscr{A}_{11,i} & \mathscr{A}_{12,i} \\ \mathscr{A}_{21,i} & \mathscr{A}_{22,i} \end{bmatrix}, \quad \mathscr{B}_i = \mathbf{T}_i\,\mathbf{B}_i = \begin{bmatrix} \mathscr{B}_{1,i}^{\mathrm{T}} & \mathscr{B}_{2,i}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}},$$
$$\mathscr{F}_i = \mathbf{T}_i\,\mathbf{F}_i = \begin{bmatrix} \mathbf{0}^{\mathrm{T}} & \mathscr{F}_{2,i}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}.$$

The **LPV model in canonical form** is described by

$$\begin{aligned} \dot{\mathbf{x}}_1 &= \mathscr{A}_{11}(\boldsymbol{\theta})\,\mathbf{x}_1 + \mathscr{A}_{12}(\boldsymbol{\theta})\,\mathbf{y} + \mathscr{B}_1(\boldsymbol{\theta})\,\mathbf{u}, \\ \dot{\mathbf{y}} &= \mathscr{A}_{21}(\boldsymbol{\theta})\,\mathbf{x}_1 + \mathscr{A}_{22}(\boldsymbol{\theta})\,\mathbf{y} + \mathscr{B}_2(\boldsymbol{\theta})\,\mathbf{u} + \mathscr{H}_2\mathbf{f}_v, \end{aligned} \tag{42}$$

with the non-measurable $\mathbf{x}_1 \in \mathbb{R}^{(n-p)}$ and the measurable states $\mathbf{y} \in \mathbb{R}^p$. The transformed system matrices in (41) have the following block structures:

$$\mathscr{A}(\boldsymbol{\theta}) = \mathbf{T}\,\mathbf{A}(\boldsymbol{\theta})\,\mathbf{T}^{-1} = \begin{bmatrix} \mathscr{A}_{11}(\boldsymbol{\theta}) & \mathscr{A}_{12}(\boldsymbol{\theta}) \\ \mathscr{A}_{21}(\boldsymbol{\theta}) & \mathscr{A}_{22}(\boldsymbol{\theta}) \end{bmatrix}, \quad \mathscr{B}(\boldsymbol{\theta}) = \mathbf{T}\,\mathbf{B}(\boldsymbol{\theta}) = \begin{bmatrix} \mathscr{B}_1^{\mathrm{T}}(\boldsymbol{\theta}) & \mathscr{B}_2^{\mathrm{T}}(\boldsymbol{\theta}) \end{bmatrix}^{\mathrm{T}},$$
$$\mathscr{H} = \mathbf{T}\,\mathbf{H} = \begin{bmatrix} \mathbf{0}^{\mathrm{T}} & \mathscr{H}_2^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}.$$

4.2 Sliding Mode Observation

TS Sliding Mode Observation

The TS sliding mode (TS SM) observer for the system (41) in transformed form is given by

$$\begin{aligned} \dot{\hat{\mathbf{x}}}_1 &= \sum_{i=1}^{N_r} h_i\,(\mathbf{z})\,\left(\mathscr{A}_{11,i}\,\hat{\mathbf{x}}_1 + \mathscr{A}_{12,i}\,\hat{\mathbf{y}} + \mathscr{B}_{1,i}\,\mathbf{u} - \mathscr{A}_{12,i}\mathbf{e}_y\right), \\ \dot{\hat{\mathbf{y}}} &= \sum_{i=1}^{N_r} h_i\,(\mathbf{z})\,\left(\mathscr{A}_{21,i}\,\hat{\mathbf{x}}_1 + \mathscr{A}_{22,i}\,\hat{\mathbf{y}} + \mathscr{B}_{2,i}\,\mathbf{u} - \left(\mathscr{A}_{22,i} - \mathscr{A}_{22}^{s}\right)\mathbf{e}_y + \boldsymbol{\nu}\right), \end{aligned} \tag{43}$$

where $\mathbf{e}_y = \hat{\mathbf{y}} - \mathbf{y}$ denotes the error vector and $\mathscr{A}_{22}^{s}$ is a common stable design matrix. An obvious choice for $\mathscr{A}_{22}^{s}$ is a diagonal matrix, where the elements are the desired eigenvalues of the output error dynamics. Using the inverse transformation $\mathbf{T}_i^{-1}$, the TS sliding mode observer can be obtained in the coordinates $\mathbf{x}$ of the original model (37)

$$\dot{\hat{\mathbf{x}}} = \sum_{i=1}^{N_r} h_i(\mathbf{z})\left(\mathbf{A}_i\,\hat{\mathbf{x}} + \mathbf{B}_i\,\mathbf{u} - \mathbf{G}_{l,i}\,\mathbf{e}_y + \mathbf{G}_{n,i}\,\boldsymbol{\nu}\right) \tag{44}$$

with the observer gains

$$\mathbf{G}_{l,i} = \mathbf{T}_i^{-1} \begin{bmatrix} \mathscr{A}_{12,i} \\ \mathscr{A}_{22,i} - \mathscr{A}_{22}^{s} \end{bmatrix}, \qquad \mathbf{G}_{n,i} = \mathbf{T}_i^{-1} \begin{bmatrix} \mathbf{0}_{(n-p)\times p} \\ \mathbf{I}_p \end{bmatrix}.$$

LPV Sliding Mode Observation

The LPV sliding mode (LPV SM) observer for the system (42) in transformed form is given by

$$\begin{aligned} \dot{\hat{\mathbf{x}}}_1 &= \mathscr{A}_{11}(\boldsymbol{\theta})\,\hat{\mathbf{x}}_1 + \mathscr{A}_{12}(\boldsymbol{\theta})\,\hat{\mathbf{y}} + \mathscr{B}_1(\boldsymbol{\theta})\,\mathbf{u} - \mathscr{A}_{12}(\boldsymbol{\theta})\mathbf{e}_y, \\ \dot{\hat{\mathbf{y}}} &= \mathscr{A}_{21}(\boldsymbol{\theta})\,\hat{\mathbf{x}}_1 + \mathscr{A}_{22}(\boldsymbol{\theta})\,\hat{\mathbf{y}} + \mathscr{B}_2(\boldsymbol{\theta})\,\mathbf{u} - \left(\mathscr{A}_{22}(\boldsymbol{\theta}) - \mathscr{A}_{22}^{s}\right)\mathbf{e}_y + \boldsymbol{\nu}. \end{aligned} \tag{45}$$

Using the inverse transformation $\mathbf{T}^{-1}$, the LPV sliding mode observer can be obtained in the coordinates $\mathbf{x}$ of the original model (38)

$$\dot{\hat{\mathbf{x}}} = \mathbf{A}(\boldsymbol{\theta})\,\hat{\mathbf{x}} + \mathbf{B}(\boldsymbol{\theta})\,\mathbf{u} - \mathbf{G}_l(\boldsymbol{\theta})\,\mathbf{e}_y + \mathbf{G}_n\,\boldsymbol{\nu} \tag{46}$$

with the observer gains

$$\mathbf{G}_l(\boldsymbol{\theta}) = \mathbf{T}^{-1} \begin{bmatrix} \mathscr{A}_{12}(\boldsymbol{\theta}) \\ \mathscr{A}_{22}(\boldsymbol{\theta}) - \mathscr{A}_{22}^{s} \end{bmatrix}, \qquad \mathbf{G}_n = \mathbf{T}^{-1} \begin{bmatrix} \mathbf{0}_{(n-p)\times p} \\ \mathbf{I}_p \end{bmatrix}.$$

The discontinuous term $\boldsymbol{\nu}$ is necessary for both observer structures in (43), (44) or (45), (46) to establish and maintain a sliding motion. The sliding motion is given by

$$\boldsymbol{\nu} = -\rho\,\frac{\mathbf{P}_2\,\mathbf{e}_y}{\left\|\mathbf{P}_2\mathbf{e}_y\right\|}, \qquad \text{if} \quad \mathbf{e}_y \neq \mathbf{0}, \tag{47}$$

where ρ is a gain factor and $\mathbf{P}_2$ is the symmetric, positive definite solution of the Lyapunov equation

$$\mathbf{P}_2\,\mathscr{A}_{22}^{s} + \mathscr{A}_{22}^{s\,T}\,\mathbf{P}_2 = -\mathbf{Q}_2, \tag{48}$$

where $\mathbf{Q}_2$ is a symmetric positive definite design matrix. Note that the discontinuous term $\boldsymbol{\nu}$ in (47) is undefined in the case of $\mathbf{e}_y = \mathbf{0}$ [6]. Once the sliding surface

$$\mathscr{S} = \{\mathbf{e}(t) \in \mathbb{R}^n := \mathscr{C}\,\underbrace{[\,\mathbf{e}_1^{\mathrm{T}}\ \mathbf{e}_y^{\mathrm{T}}\,]^{\mathrm{T}}}_{\mathbf{e}} = \mathbf{0}\,\} \tag{49}$$

$$\text{with} \quad \mathscr{C} = \mathbf{C}\,\mathbf{T}_c^{-1} = \left[\mathbf{0}_{p\times(n-p)}\,\mathbf{I}_p\right] \tag{50}$$

is reached at the time $t = t_r$, the TS SM observer attempts to maintain the sliding motion on the surface $\mathscr{S}$.

Actuator Fault Reconstruction by TS SM Observation

For the actuator fault reconstruction we consider first the TS model (41) and the TS SM observer (43) in canonical form. The derivatives of the error of non-measurable

and measurable states are $\dot{\mathbf{e}}_1 = \dot{\hat{\mathbf{x}}}_1 - \dot{\mathbf{x}}_1$, $\dot{\mathbf{e}}_y = \dot{\hat{\mathbf{y}}} - \dot{\mathbf{y}}$ and substituting (41) and (43) it can be verified that

$$\dot{\mathbf{e}}_1 = \sum_{i=1}^{N_r} h_i(\mathbf{z}) \mathscr{A}_{11,i}\, \mathbf{e}_1 \tag{51}$$

and

$$\dot{\mathbf{e}}_y = \sum_{i=1}^{N_r} h_i(\mathbf{z}) \big(\mathscr{A}_{21,i}\, \mathbf{e}_1 + \mathscr{A}^s_{22}\, \mathbf{e}_y + \boldsymbol{\nu} - \mathscr{F}_{2,i}\, \mathbf{f}_a \big). \tag{52}$$

Assume the TS SM observer has been designed and a sliding motion has been established from $t \geq t_r$. This means that $\mathbf{e}_y = \mathbf{0}$, $\dot{\mathbf{e}}_y = \mathbf{0}$. In this case the error equation (52) is simplified to

$$\mathbf{0} = \sum_{i=1}^{N_r} h_i(\mathbf{z}) \big(\mathscr{A}_{21,i}\, \mathbf{e}_1 + \boldsymbol{\nu}_{eq} - \mathscr{F}_{2,i}\, \mathbf{f}_a \big) \tag{53}$$

and the discontinuous term $\boldsymbol{\nu}$ is replaced by the so-called equivalent output injection signal [4]

$$\boldsymbol{\nu}_{eq} = \boldsymbol{\nu}_\delta (t \geq t_r), \tag{54}$$

where $\boldsymbol{\nu}_\delta$ denotes an approximation of (47) by introducing a small positive scalar δ

$$\boldsymbol{\nu}_\delta = -\rho \frac{\mathbf{P}_2\, \mathbf{e}_y}{\| \mathbf{P}_2 \mathbf{e}_y \| + \delta}. \tag{55}$$

It should be noted that the value of δ should be chosen as small as possible, because it influences the quality of the fault reconstruction [6]. Thus, the equivalent output injection signal is given by rearranging Eq. (53) according to

$$\boldsymbol{\nu}_{eq} = \sum_{i=1}^{N_r} h_i(\mathbf{z}) \big(\mathscr{F}_{2,i}\, \mathbf{f}_a - \mathscr{A}_{21,i}\, \mathbf{e}_1 \big). \tag{56}$$

Substituting the steady-state solution of (51) into (56), we get the relation

$$\mathbf{f}_a = \left[\sum_{i=1}^{N_r} h_i(\mathbf{z})\, \mathscr{F}_{2,i} \right]^+ \boldsymbol{\nu}_{eq}, \tag{57}$$

where $(\cdot)^+$ denotes the pseudo-inverse of the convex combination of the matrices $\mathscr{F}_{2,i}$. Note that the pseudo-inverse of the convex combination of matrices exists if the full rank characterization is satisfied by the theorem proposed in [11].

Actuator Fault Reconstruction by LPV SM Observation

Equal to the TS approach, the error equations are defined by $\dot{\mathbf{e}}_1 = \mathscr{A}_{11}(\boldsymbol{\theta})\,\mathbf{e}_1$ and $\dot{\mathbf{e}}_y = \dot{\hat{\mathbf{y}}} - \dot{\mathbf{y}}$. On the sliding surface it simplifies to

$$\dot{\mathbf{e}}_y = \mathbf{0} = \mathscr{A}_{21}(\boldsymbol{\theta})\,\mathbf{e}_1 + \nu_{eq} - \mathscr{H}_2\,\mathbf{f}_v. \tag{58}$$

Substituting the steady-state solution of $\dot{\mathbf{e}}_1$ into (58), which is stable by the design of the transformation matrix $\mathbf{T}_L$ (cf. Sect. 6), leads to

$$\mathbf{f}_v = \left[\mathscr{H}_2\right]^+ \nu_{eq}. \tag{59}$$

5 Simulation Results for the Case Studies

After the canonical form of the observer and the existence conditions have been shown, the main focus is directed on the degrees of freedom in the design process having a direct impact on the reachability of the sliding surface. This is a necessary assumption in the reconstruction of the faults. Furthermore, the quality of the fault reconstruction directly depends on the chosen parameters. The design process can be considered as an iterative procedure [8] using the simulation environment including the nonlinear model and the constructed observer. Note that the simulation itself influences the quality of the reconstruction through parameters like the chosen solver, the fundamental sample time and effects like chattering. The following descriptions of the design process are based on the experiences of the authors and sketch a possible way of achieving a well-operating observer. Anyhow, there might be different approaches, which work as well.

The design matrix $\mathscr{A}^s_{22}$ (cf. (48)) plays an important role in the design process, since it governs the output error dynamics. One way to evaluate the influence of $\mathscr{A}^s_{22}$ in the simulation is to operate the designed observer in the form similar to the Luenberger observer without the switching term

$$\begin{aligned} &\dot{\hat{\mathbf{x}}} = \sum_{i=1}^{N_r} h_i(\mathbf{z})\left[\mathbf{A}_i\hat{\mathbf{x}} + \mathbf{B}_i\mathbf{u} - \mathbf{G}_{l,i}\mathbf{e}_{\tilde{y}}\right], \quad \hat{\mathbf{y}} = \mathbf{C}\hat{\mathbf{x}}, \qquad \text{resp.} \\ &\dot{\hat{\mathbf{x}}} = \mathbf{A}(\boldsymbol{\theta})\hat{\mathbf{x}} + \mathbf{B}(\boldsymbol{\theta})u - \mathbf{G}_l(\boldsymbol{\theta})\mathbf{e}_{\tilde{y}}, \quad \hat{\mathbf{y}} = \mathbf{C}\hat{\mathbf{x}} \end{aligned} \tag{60}$$

until the desired behaviour of the downgraded observer is achieved. Based on the designed Luenberger-like observer and after enabling the switching term again, the parameters $\boldsymbol{\rho}$ and δ for the equivalent output injection signal can be acquired. As a first step, δ can be chosen as a small scalar (cf. Tables 4 and 5) and creates an area

Table 4 Design parameters for case study I

Parameter	Value TS	Value LPV
$\mathscr{A}_{22}^{s}$	$-10\ \mathrm{diag}(1, 1, 1)$	$-10\ \mathrm{diag}(1, 1, 1)$
$\mathbf{Q}_2$	$\mathrm{diag}(1, 1, 1)$	$\mathrm{diag}(1, 1, 1)$
ρ	1000	1000
δ	$10^{-1.3}$	$10^{-1.3}$
Resulting matrices from the design process		
$\mathbf{P}_2$	$0.05\ \mathrm{diag}(1, 1, 1)$	$0.05\ \mathrm{diag}(1, 1, 1)$
$\mathbf{G}_l$	$\begin{bmatrix} 13.09 & 0 & 0 \\ 34.44 & 0 & 0 \\ 0 & 10 & 1 \\ 1.38 & 0 & 9.99 \end{bmatrix}$	$\begin{bmatrix} 20.5 & 0 & 0 \\ 108.4 & 0 & 0 \\ 0 & 10 & 1 \\ 2.8 & 0 & 10 \end{bmatrix}$
$\mathbf{G}_n$	$\begin{bmatrix} 1 & 0 & 0 \\ 3.1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 10.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

around the sliding surface, which reduces effects like chattering. As an indication of the magnitude of the gain matrix ρ, the proof of the stability of $\dot{\mathbf{e}}$ through the Lyapunov function, presented for example in [8] for TS and in [1] for the LPV case, can be utilised. Therefore, the maximum values of the errors, matrices, faults and uncertainties have to be estimated. Afterwards, the magnitude of ρ can be altered iteratively until the desired results are achieved. In case of the wind turbine observer, a matrix $\boldsymbol{\rho}$ was used instead of a scalar, which takes account of the different magnitudes of the outputs of the wind turbine. Especially for multiple faults to detect, the use of a matrix can lead to an improvement of the performance. This allows a higher degree of freedom in the design process, but increases the complexity of finding a combination of entries in $\boldsymbol{\rho}$, which leads to a good reconstruction accuracy. The entries in $\boldsymbol{\rho}$ have to be well matched for a good reconstruction accuracy. To integrate the extended switching term with $\boldsymbol{\rho}$ as a matrix, a weighting matrix $\mathbf{W}$ is introduced into the switching term. The entries of $\mathbf{W}$ contain the reciprocal values of the maximum estimated faults of the outputs and lead to a normalisation [6]. Note that the iterative alteration of $\boldsymbol{\rho}$ (or ρ) does not depend on one specific fault, but rather on the accuracy of the reconstruction over a range of faults is used as a criterion. A well-working combination of parameters for the case studies is shown in Tables 4 and 5. They represent a possible solution for the fault reconstruction; however, other parameter sets might lead to a good reconstruction of faults as well. In case study I the ODE3 Simulink® solver with a fixed sample time of 0.001 s was used. Case study II was calculated by an ODE4 Simulink® solver with a fixed sample time of 0.001 s.

Table 5 Design parameters for case study II

Para.	Value TS	Value LPV
$\mathscr{A}_{22}^{s}$	$-100\ \mathrm{diag}(1,\,1,\,1,\,1,\,1,\,1)$	$-100\ \mathrm{diag}(1,\,1,\,1,\,1,\,1,\,1)$
$\mathbf{Q}_2$	$\mathrm{diag}(1,\,1,\,1,\,1,\,1,\,1)$	$\mathrm{diag}(1,\,1,\,1,\,1,\,1,\,1)$
ρ	$\mathrm{diag}(1.2\mathrm{e}3,\ 75\mathrm{e}3,\ 1\mathrm{e}7,\ 1\mathrm{e}4,\ 1\mathrm{e}4,\ 1\mathrm{e}4)$	$\mathrm{diag}(1.2\mathrm{e}3,\ 75\mathrm{e}3,\ 1\mathrm{e}7,\ 1\mathrm{e}4,\ 1\mathrm{e}4,\ 1\mathrm{e}4)$
$\mathbf{W}$	$\mathrm{diag}(1/1.4,\ 1/20,\ 1/4000,\ 1/5,\ 1/5,\ 1/5)$	$\mathrm{diag}(1/1.4,\ 1/20,\ 1/4000,\ 1/5,\ 1/5,\ 1/5)$
δ	0.005	0.005
Resulting matrices from the design process		
$\mathbf{P}_2$	$0.005\ \mathrm{diag}(1,\,1,\,1,\,1,\,1,\,1)$	$0.005\ \mathrm{diag}(1,\,1,\,1,\,1,\,1,\,1)$
$\mathbf{G}_l$	$\begin{bmatrix} -3.26\mathrm{e}3 & -0.0073 & 0 & & & \\ -10.5 & 110.4 & -0.0026 & \multicolumn{3}{c}{\multirow{3}{*}{$\mathbf{0}_{4\times3}$}} \\ 0.99 & 0.0043 & 0 & & & \\ 0 & 0 & 50 & & & \\ & & & -123.4 & 0 & 0 \\ & & & 100 & 0 & 0 \\ & & & 0 & -123.4 & 0 \\ & \multicolumn{2}{c}{\mathbf{0}_{6\times3}} & 0 & 100 & 0 \\ & & & 0 & 0 & -123.4 \\ & & & 0 & 0 & 100 \end{bmatrix}$	$= \mathbf{G}_l(h_i = 1/N_r\ \forall i)$
$\mathbf{G}_n$	$\begin{bmatrix} 1 & 0 & 0 & & & \\ 0 & 1 & 0 & \multicolumn{3}{c}{\multirow{3}{*}{$\mathbf{0}_{4\times3}$}} \\ -0.0001 & 0.0001 & 0 & & & \\ 0 & 0 & 1 & & & \\ & & & 0 & 0 & 0 \\ & & & 1 & 0 & 0 \\ & \multicolumn{2}{c}{\mathbf{0}_{6\times3}} & 0 & 0 & 0 \\ & & & 0 & 1 & 0 \\ & & & 0 & 0 & 0 \\ & & & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & & & \\ 0 & 1 & 0 & \multicolumn{3}{c}{\multirow{3}{*}{$\mathbf{0}_{4\times3}$}} \\ -0.0001 & 0.0001 & 0 & & & \\ 0 & 0 & 1 & & & \\ & & & 0 & 0 & 0 \\ & & & 1 & 0 & 0 \\ & \multicolumn{2}{c}{\mathbf{0}_{6\times3}} & 0 & 0 & 0 \\ & & & 0 & 1 & 0 \\ & & & 0 & 0 & 0 \\ & & & 0 & 0 & 1 \end{bmatrix}$

Note that the observer matrix $\mathbf{G}_l$ depends on either the membership functions h_i or the affine term θ_i. The matrices shown in Tables 4 and 5 are based on the operating point, where $h_i = 1/N_r\ \forall i$ and $\theta_i = 0\ \forall i$, and give an idea of the structure of $\mathbf{G}_l$.

5.1 Simulation Results for Case Study I

Figure 3 shows the reconstructed fault $\hat{f}_F$ by the LPV sliding mode observer based on the approximated model compared to the real occurring fault f_F. In Fig. 4 the reconstructed fault $\hat{f}_F$ of the TS sliding mode observer compared to the real occurring fault f_F is plotted. Both designed observers show a good accuracy of reconstruction.

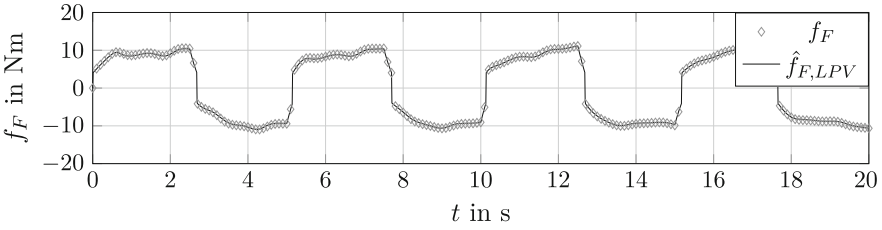


Fig. 3 Reconstructed fault $\hat{f}_F$ by the LPV SM observer of the inverted pendulum benchmark

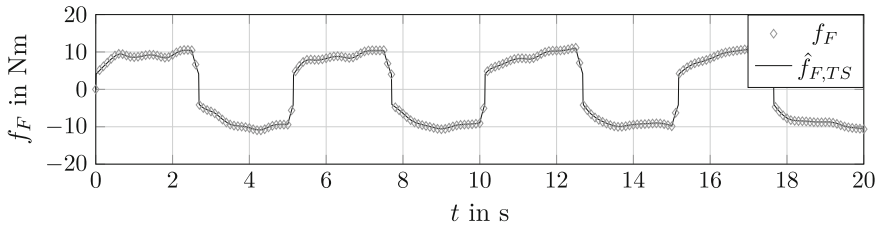


Fig. 4 Reconstructed fault $\hat{f}_F$ by the TS SM observer of the inverted pendulum benchmark

5.2 *Simulation Results for Case Study II*

The benchmark model of the wind turbine by Odgaard et al., presented in [15], was created for the evaluation of FDI and FTC methods. For this reason, fault scenarios are implemented in the MATLAB model of the wind turbine. As an example, Fault 8 (cf. [15]), which results in an offset of $f_a = 100\,\text{Nm}$ on T_g and is active from $3800\,\text{s} \leq t \leq 3900\,\text{s}$, was chosen to give an impression of the performance of the designed sliding mode observer. The observers based on the TS and on the LPV model were simulated in parallel to the nonlinear benchmark model of the wind turbine with the faults included. From $t = 3800\,\text{s}$ to $t = 3900\,\text{s}$ an offset of 100 Nm on the generator torque T_g occurred. The benchmark provides additional noise to the output of the model, which leads to a fluctuation in the fault-free and fault-afflicted reconstruction signal, cf. Figs. 5 and 6. The reconstructed signals of the faults are based on the signals of the noisy benchmark. This is handled by the use of a filter with a transfer function $H(s) = \frac{10^{-5}s^2+129.33}{s^2+14.91s+130.83}$, which is applied to the noisy reconstructed signals. The plots in Figs. 5 and 6 show the filtered reconstructed signals. Note that, for example in the moment of occurrence of the fault, the fault to the nominal torque magnitude ratio is approx. $f_a/T_g = 0.8\,\%$. The sensor noise itself can reach magnitudes of 30 Nm, which makes it harder for the used observers to reconstruct the induced fault. Anyhow, the observers for the LPV and TS model show an identical accuracy of reconstruction of approx. 95 %. This result is not surprising, since the LPV and TS models are an exact representation of the nonlinear model and, as shown in Table 5, the design leads to exactly the same sliding mode observer.

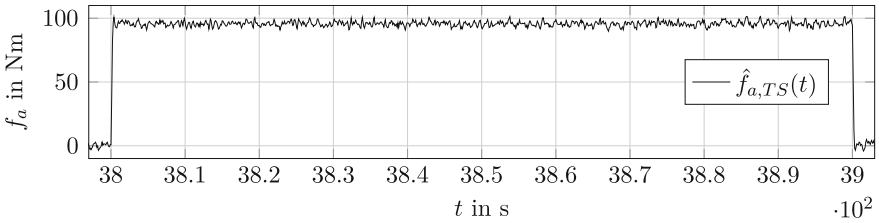


Fig. 5 Reconstructed fault f_a by the TS SMO of the wind turbine benchmark

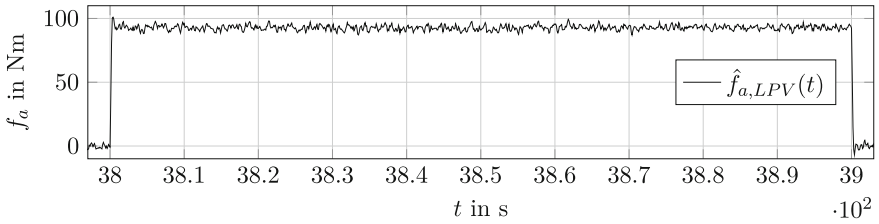


Fig. 6 Reconstructed fault f_a by the LPV SMO of the wind turbine benchmark

6 Stability of TS and LPV Systems: A Comparison

Consider the autonomous system in the LPV description as it was introduced in Sect. 2.2

$$\dot{\mathbf{x}} = \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x}. \tag{61}$$

As described by Shamma in [13], the system is quadratically stable if there exists a symmetric, positive definite solution $\mathbf{P}$ of

$$\mathbf{P}\,\mathbf{A}(\boldsymbol{\theta}) + \mathbf{A}(\boldsymbol{\theta})^{\mathrm{T}}\,\mathbf{P} < 0 \tag{62}$$

for all possible trajectories $\boldsymbol{\theta}$. This condition is based on the exploitation of the characteristics of the Lyapunov function $V = \mathbf{x}^{\mathrm{T}}\mathbf{P}\,\mathbf{x}$, which ensures quadratic stability. Since the LPV and TS system descriptions lead to a convex formulation, this approach can be handled by the use of linear matrix inequalities (LMI). When comparing LPV and TS models the way they are introduced in this discourse (cf. Sect. 2), it is possible to show that, with the same nonlinear functions, the LMI constraints become the same. Consider the system matrix in (61), where the structure can be arranged to

$$\mathbf{A}(\boldsymbol{\theta}) = \mathbf{A}_0\big|_{\underline{z}_i + \frac{\Delta z_i}{2}} + \mathbf{A}_1(\theta_1) + \mathbf{A}_2(\theta_2) + \cdots + \mathbf{A}_{n_l}(\theta_{n_l}). \tag{63}$$

This model has n_l nonlinearities ($\boldsymbol{\theta} \in \mathbb{R}^{n_l}$). Note that due to lack of space in the following considerations $\mathbf{A}_0\big|_{\underline{z}_i + \frac{\Delta z_i}{2}}$ defines $\mathbf{A}_0\big|_{\underline{z}_1 + \frac{\Delta z_1}{2},\ \underline{z}_2 + \frac{\Delta z_2}{2},\ \ldots,\ \underline{z}_{n_l} + \frac{\Delta z_{n_l}}{2}}$. Since the prob-

lem was formulated as in Eq. (5), it holds that $\theta_i \in [-1, 1]\,\forall i$. Because this is a convex problem formulation, the LMI constraints are governed by the bounds of each θ_i. Therefore, there are $N_r = 2^{n_l}$ possible combinations. Setting the matrices according to the possible combinations leads to

$$\begin{aligned}
\tilde{\mathbf{A}}_1 &= \mathbf{A}_0\big|_{\underline{z}_i+\frac{\Delta z_i}{2}} + \mathbf{A}_1(\theta_1 = 1) + \mathbf{A}_2(\theta_2 = 1) + \cdots + \mathbf{A}_{n_l}(\theta_{n_l} = 1)\\
\tilde{\mathbf{A}}_2 &= \mathbf{A}_0\big|_{\underline{z}_i+\frac{\Delta z_i}{2}} + \mathbf{A}_1(\theta_1 = 1) + \mathbf{A}_2(\theta_2 = 1) + \cdots + \mathbf{A}_{n_l}(\theta_{n_l} = -1)\\
&\vdots \qquad\qquad\qquad\qquad \vdots\\
\tilde{\mathbf{A}}_{N_r} &= \mathbf{A}_0\big|_{\underline{z}_i+\frac{\Delta z_i}{2}} + \mathbf{A}_1(\theta_1 = -1) + \mathbf{A}_2(\theta_2 = -1) + \cdots + \mathbf{A}_{n_l}(\theta_{n_l} = -1).
\end{aligned} \tag{64}$$

From the definition of the LPV model in (6), it is easy to verify that $\mathbf{A}_i(\theta_i = 1) = \mathbf{A}_i\big|_{\frac{\Delta z_i}{2}}$ and $\mathbf{A}_i(\theta_i = -1) = \mathbf{A}_i\big|_{-\frac{\Delta z_i}{2}}$ holds. Using this knowledge, (64) can be described by

$$\begin{aligned}
&\mathbf{A}_0\big|_{\underline{z}_i+\frac{\Delta z_i}{2}} + \mathbf{A}_1\big|_{\frac{\Delta z_1}{2}} + \mathbf{A}_2\big|_{\frac{\Delta z_2}{2}} + \cdots + \mathbf{A}_{n_l}\big|_{\frac{\Delta z_{n_l}}{2}} = \tilde{\mathbf{A}}_1\big|_{\overline{z}_1,\,\overline{z}_2,\ldots,\overline{z}_{n_l}}\\
&\mathbf{A}_0\big|_{\underline{z}_i+\frac{\Delta z_i}{2}} + \mathbf{A}_1\big|_{\frac{\Delta z_1}{2}} + \mathbf{A}_2\big|_{\frac{\Delta z_2}{2}} + \cdots + \mathbf{A}_{n_l}\big|_{-\frac{\Delta z_{n_l}}{2}} = \tilde{\mathbf{A}}_2\big|_{\overline{z}_1,\,\overline{z}_2,\ldots,\underline{z}_{n_l}}\\
&\qquad\qquad\vdots \qquad\qquad\qquad\qquad \vdots\\
&\mathbf{A}_0\big|_{\underline{z}_i+\frac{\Delta z_i}{2}} + \mathbf{A}_1\big|_{-\frac{\Delta z_1}{2}} + \mathbf{A}_2\big|_{-\frac{\Delta z_2}{2}} + \cdots + \mathbf{A}_{n_l}\big|_{-\frac{\Delta z_{n_l}}{2}} = \tilde{\mathbf{A}}_{N_r}\big|_{\underline{z}_1,\,\underline{z}_2,\ldots,\underline{z}_{n_l}}.
\end{aligned} \tag{65}$$

Thus, the LMI constraints to solve equal

$$\begin{aligned}
&\tilde{\mathbf{A}}_1\big|^{\mathrm{T}}_{\overline{z}_1,\,\overline{z}_2,\ldots,\overline{z}_{n_l}}\mathbf{P} + \mathbf{P}\tilde{\mathbf{A}}_1\big|_{\overline{z}_1,\,\overline{z}_2,\ldots,\overline{z}_{n_l}} < 0\\
&\tilde{\mathbf{A}}_2\big|^{\mathrm{T}}_{\overline{z}_1,\,\overline{z}_2,\ldots,\underline{z}_{n_l}}\mathbf{P} + \mathbf{P}\tilde{\mathbf{A}}_2\big|_{\overline{z}_1,\,\overline{z}_2,\ldots,\underline{z}_{n_l}} < 0\\
&\vdots \qquad\qquad\qquad\qquad \vdots\\
&\tilde{\mathbf{A}}_{N_r}\big|^{\mathrm{T}}_{\underline{z}_1,\,\underline{z}_2,\ldots,\underline{z}_{n_l}}\mathbf{P} + \mathbf{P}\tilde{\mathbf{A}}_{N_r}\big|_{\underline{z}_1,\,\underline{z}_2,\ldots,\underline{z}_{n_l}} < 0.
\end{aligned} \tag{66}$$

The comparison of the matrices $\tilde{\mathbf{A}}_i$, $i \in \{1, 2, \ldots, N_r\}$ to the individual matrices of the submodels in the TS formulation

$$\dot{\mathbf{x}} = \sum_{i=1}^{N_r} h_i(\mathbf{z})\,\mathbf{A}_i\,\mathbf{x} \tag{67}$$

leads to the realisation that they are the same. When using the Lyapunov function approach for the TS model, it is easy to verify that LMI formulation of the TS and the LPV system result in the same constraints.

As described in [1, 8], the error dynamics for the non-measurable states $\dot{\mathbf{e}}_1 = \mathscr{A}_{11}\,\mathbf{e}_1$ are ensured to be stable by the use of an LMI problem formulation

$$\begin{aligned}
\mathbf{P}_1 &> 0,\\
\mathbf{A}_{a,11}(\boldsymbol{\theta})\mathbf{P}_1 + \mathbf{P}_1\mathbf{A}_{a,11}(\boldsymbol{\theta}) + \mathbf{A}_{a,211}(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{N}^{\mathrm{T}} + \mathbf{N}\mathbf{A}_{a,211}(\boldsymbol{\theta}) &< 0,\\
&\text{resp.}\\
\mathbf{P}_1 &> 0,\\
\mathbf{A}_{a,11,i}\,\mathbf{P}_1 + \mathbf{P}_1\mathbf{A}_{a,11,i} + \mathbf{A}_{a,211,i}{}^{\mathrm{T}}\mathbf{N}_i^{\mathrm{T}} + \mathbf{N}_i\mathbf{A}_{a,211,i} &< 0, \text{ for } i \in \{1,\ldots,N_r\}.
\end{aligned} \tag{68}$$

As outlined, the LMI constraints for the TS and LPV model are equal, when based on the same nonlinear functions. However, there *might be* a difference in the solution. This is due to the fact that for the TS observer the solver is allowed to find N_r different solutions for $\mathbf{N}_i$, since the resulting transformation from the design process can be applied to each individual subsystem $\{\mathbf{A}_i, \mathbf{B}_i, \mathbf{C}_i\}$ of the TS system description. In case of the LPV design process one solution for $\mathbf{N}$ is accepted based on the same constraints. This is due to the fact that no obvious assignment of a solution to individual matrices results from the design process.

7 Conclusions

In this chapter, a LPV and Takagi–Sugeno model-based sliding mode observer design approach was investigated. After a brief introduction of both model structures, the entire modelling process for the observer design was studied in detail by means of two case studies using LPV and TS techniques.

As a result, it can be noted that there exist wide similarities between the LPV and the TS extension of the canonical LTI form of sliding mode observers. Both approaches of the polytopic extension of uncertain LTI systems are suitable for the consideration of nonlinearities of the nominal system dynamics. In particular, there are few differences which can lead to different dynamics of the reconstructed unknown inputs, respectively, occurring faults:

- In the case of non-factorizable fault distribution matrices $\mathbf{F}(\boldsymbol{\theta})$ (inverted pendulum), the use of the LPV approach requires a model approximation. In contrast, the TS model approach does not require any approximation and it is therefore straightforward to implement without loss of accuracy.
- The TS model structure is characterised by weighted convex combinations of linear submodels. This can be exploited in the design process, because the LMI problem formulation allow for different solutions for $\mathbf{N}_i$ from which $i \in \{1,\ldots,N_r\}$ observer gains $\mathbf{G}_{l,i}$ and $\mathbf{G}_{n,i}$ follow for each individual subsystem. In contrast, the individual matrices in the LPV structure cannot be assigned to related submodels. Based on this fact, the LMI problem formulation caused a common solution $\mathbf{N}$ whereby the sliding mode observer design is also restricted to a common transformation matrix $\mathbf{T}$.

However, it must be noted that the performance of the LPV sliding mode observer can be seen as equivalent to the TS sliding mode observer. In both case studies the designed observers achieved a high accuracy of the reconstructed fault signals.

References

1. Alwi H, Edwards C (2010) Robust actuator fault reconstruction for LPV systems using sliding mode observers. In: IEEE conference on decision and control. Hilton Atlanta Hotel, Atlanta
2. Bianchi FD, De Battista H, Mantz RJ (2007) Wind turbine control systems - principles modelling and gain scheduling design. Springer, London Limited, London
3. Donath H, Georg S, Schulte H (2013) Takagi-Sugeno sliding mode observer for friction compensation with application to an inverted pendulum. In: IEEE international conference on fuzzy systems (FUZZ). Hyderabad, India. doi:10.1109/FUZZ-IEEE.2013.6622558
4. Edwards C, Spurgeon SK (1998) Sliding mode control: theory and applications. Taylor & Francis, Boca Raton
5. Gasch R, Twele J (2012) Wind power plants, 2nd edn. Springer, Berlin
6. Georg S (2015) Fault diagnosis and fault-tolerant control of wind turbines nonlinear Takagi-Sugeno and sliding mode techniques. Ph.D. thesis, University of Rostock, Faculty of Mechanical Engineering and Marine Technology
7. Georg S, Schulte H, Aschemann H (2012) Control-oriented modelling of wind turbines using a Takagi-Sugeno model structure. In: IEEE international conference on fuzzy systems. Brisbane, Australia, pp. 1737–1744
8. Gerland P (2011) Klassifikationsgestützte on-line Adaption eines robusten beobachterbasierten Fehlerdiagnoseansatzes für nichtlineare Systeme. Ph.D. thesis, Universität Kassel
9. Gerland P, Groß D, Schulte H, Kroll A (2010) Design of sliding mode observers for TS fuzzy systems with application to disturbance and actuator fault estimation. In: IEEE conference on decision and control. Hilton Atlanta Hotel, Atlanta, pp. 4373–4378
10. Gerland P, Groß D, Schulte H, Kroll A (2010) Robust adaptive fault detection using global state information and application to mobile working machines. In: Conference on control and fault-tolerant systems. Nice, France, pp. 813–818
11. Kolodziejczak B, Szulc T (1999) Convex combinations of matrices - full rank characterization. Linear Algebra Appl 287:215–222
12. Kroll A, Schulte H (2014) Benchmark problems for nonlinear system identification and control using soft computing methods: need and overview. Appl Soft Comput 25(12):496–513
13. Mohammadpour J, Scherer CW (eds) (2012) Control of linear parameter varying systems with applications, 1st edn. Springer, New York
14. Odgaard PF, Stoustrup J, Kinnaert M (2009) Fault tolerant control of wind turbines - a benchmark model. In: IFAC symposium on fault detection, supervision and safety of technical processes. Barcelona, Spain, pp. 155–160
15. Odgaard PF, Stoustrup J, Kinnaert M (2013) Fault-tolerant control of wind turbines: a benchmark model. IEEE Trans Control Syst Technol 21(4):1168–1182
16. Ohtake H, Tanaka K, Wang HO (2001) Fuzzy modeling via sector nonlinearity concept. In: Joint 9th IFSA world congress and 20th NAFIPS international conference. Vancouver, Canada, pp. 127–132
17. Pöschke F, Georg S, Schulte H (2014) Fault reconstruction using a Takagi-Sugeno sliding mode observer for the wind turbine benchmark. In: UKACC international conference on control (CONTROL). Loughborough, pp. 456–461. doi:10.1109/CONTROL.2014.6915183
18. Schulte H, Gerland P (2010) Observer-based estimation of pressure signals in hydrostatic transmissions. In: IFAC symposium advances in automotive control (AAC). Munich, Germany
19. Shamma JS (1988) Analysis and desgin of gain scheduled control systems. Ph.D. thesis, Massachusetts Institute of Technology

20. Sugeno M, Kang GT (1988) Structure identification of fuzzy models. Fuzzy Sets Syst 28:15–33
21. Takagi T, Sugeno M (1985) Fuzzy identification of systems and its application to modeling and control. IEEE Trans Syst Man Cybern 15(1):116–132
22. Tanaka K, Sano M (1994) On the concept of fuzzy regulators and fuzzy observers. In: IEEE conference on fuzzy systems, pp. 767–772
23. Tanaka K, Sano M (1994) A robust stabilization problem of fuzzy control systems and its application to backing up control of a truck-trailer. IEEE Trans Fuzzy Syst 2(2):119–134
24. Tanaka K, Wang HO (2001) Fuzzy control systems design and analysis: a linear matrix inequality approach. Wiley, New York
25. Utkin VI (1979) Variable structure systems with sliding mode. IEEE Trans Autom Control 22(2):212–222
26. Utkin VI (1992) Sliding modes in control optimization. Springer, Berlin
27. Wang HO, Tanaka K, Griffin MF (1996) An approach to fuzzy control of nonlinear systems: stability and design issues. IEEE Trans Fuzzy Syst 4(1):14–23

Part III
Variable-Structure Methods and Models in Control and Estimation

The focus of the first two parts of this book was set on the use of sliding mode techniques, both in the frame of control synthesis and observer design. In addition to these tasks, sliding mode and related variable-structure approaches are also relevant for modeling and design of decentralized control systems, predictive control, and further (stochastic) fault diagnosis techniques. In Chap. 9, Zheng Huang, Ron J. Patton, and Jianglin Lan present sliding mode state and fault estimation for decentralized systems, including a detailed case study highlighting the most important methodological aspects. Hybrid estimation tasks, due to transitions between various dynamic system models, are discussed by Dirk Weidemann and Ilja Alkov in Chap. 10 for the fault diagnosis of nonlinear differential-algebraic equation systems. Here, Kalman filtering techniques are adapted in such a way that they become applicable for systems of differential-algebraic equations with a variable structure. In order to make a control design insensitive against faults, Piotr Witczak and Marcin Witczak deal with fault-tolerant model predictive control techniques under consideration of constraints for Takagi–Sugeno systems in Chap. 11. This book is concluded by a contribution of Jens Tonne and Olaf Stursberg who focus on a constrained model predictive control of processes with an uncertain structure. This structural uncertainty is described suitably by jump Markov linear systems.

Sliding Mode State and Fault Estimation for Decentralized Systems

Zheng Huang, Ron J. Patton and Jianglin Lan

Abstract The interconnection of dynamical systems gives rise to interesting challenges for control in terms of stability, robustness and the overall performance of the global interconnected systems as well as the fault tolerance of the individual subsystems. Interconnected systems can be developed either from a standpoint of centrality of control based on the construction and design of a global system that satisfies the above requirements. Alternatively, the interconnected system can be decentralized which means that the stability, performance, etc., requirements are achieved at the local (subsystem) levels. To develop a good "fault-tolerant control" strategy for decentralized systems it is necessary to take account of various faults or uncertainties that may occur throughout all local levels of the system. A powerful way to achieve this is to use robust state and fault estimation methods accounting for the model–reality mismatch that is inevitable when (a) systems are linearized and (b) when faults occur in subsystem components such as actuators, sensors, etc. The chapter develops a strategy for decentralized state and fault estimation based on the Walcott–Żak form of sliding mode observer (SMO) with linear matrix inequality (LMI) formulation. This strategy is shown to be advantageous when considering the estimation problem for a large number of interconnected subsystems. After developing the design procedure a tutorial example of two interconnected linear systems with nonlinear interconnection functions shows that the states as well as actuator and sensor faults can be robustly estimated. Finally, an application-oriented example of a three-machine power system is given which has actuator faults as well as nonlinear machine interconnections.

Z. Huang · R.J. Patton (✉) · J. Lan
Control and Intelligent Systems Engineering Research Laboratory,
University of Hull, Hull, UK
e-mail: r.j.patton@hull.ac.uk

Z. Huang
e-mail: hzcadmean@hotmail.com

J. Lan
e-mail: lanjianglin@126.com

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*,
Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_9

1 Introduction

The complexity of industrial, process, banking and IT systems increases rapidly as modern technology makes more and more use of interconnected, embedded, networked and distributed architectures [2, 16, 22, 25, 26, 28]. Šiljak [25] states that the complexity of real systems might not be well organized, whilst for control to be effective a good structural system organization is required. This can be achieved using a decentralized system structure in which local interconnected subsystems are well-defined.

A potentially significant number of variables can be involved in a decentralized system, along with nonlinear interrelationships, and various kinds of uncertainty. This means that mathematical models corresponding to the dynamics of the interconnected system may be hard to define precisely. Hence, in the control systems community, methods of handling modelling uncertainty stand out as significant challenges for the design of model-based methods of control and estimation of interconnected system. Some researchers state that the complexity and difficulties of control of interconnected systems arise mainly from the dimensionality, uncertainty, delay and information constraints [2, 25].

Decentralized control is economical and can be reliable. When the system is too large to be dealt with by centralized control, it is computationally efficient to use only local information, i.e. local states or outputs, to make the control decision. This method is also economical since it is easier to implement and it can effectively reduce the communication cost [26]. Decentralized control also facilitates the development of good robustness and provides an opportunity to achieve good closed-loop system stability. Furthermore, the decentralized system can be made robust to a broad range of subsystem or interconnection uncertainties [26, 28].

However, the disturbance from interactions should be handled by combined use of state estimation and control and some researchers use decentralized observer-based control, e.g. [1, 3, 17–19, 23, 32], to achieve some control performance goals (stability, robustness, time and/or frequency domain objectives, etc.), whilst making use of joint robustness in control and estimation.

A further challenge arises, even for centralized systems, when actuator, sensor or process faults are present. The design problem is further complicated by the combined presence of faults and uncertainties. The fact that faults and modelling uncertainties and/or disturbance may act together in a dynamical system leads to the important requirement that all these unwanted system effects need to be minimized, e.g. using estimation combined with control compensation. This challenge goes far beyond the usual design requirements of robust control, since the faults have to be considered as new forms of uncertainty/disturbance acting on the system. For interconnected and decentralized systems the combined presence of faults and modelling uncertainty/disturbance becomes an even larger challenge.

Some studies in the control community focus only on the details of the decentralized state estimation problem based on modern robust estimation methods [7, 27]. There is a requirement for this due to the difficulty of achieving good state estimation

performance in the presence of unavoidable interactions. This is the motivation for this current study. Interactions which act as uncertainties can be tolerated or compensated by fault estimation. There are several powerful approaches in the literature for robust fault estimation using specialized observers, e.g. [5, 11, 24, 36]. If a fault signal can be reconstructed or estimated robustly then there is also a possibility of making direct use of the fault estimate in a robust control scheme to compensate the fault effect, providing an opportunity for good fault-tolerant control performance [14]. Hence, the main challenge of fault estimation for interconnected systems is to understand and take into account the separate effects and influences from interactions and uncertainties.

The SMO can track the state trajectories with insensitivity to certain so-called matched and bounded perturbations during sliding mode dynamic behaviour, which may be considered as "actuator faults" and the SMO provides an excellent framework for robust estimation and reconstruction of faults [9], using the idea of equivalent injection signals. During sliding the effect of the fault on the sliding motion is eliminated but it can be shown that the equivalent injection signal becomes a very good estimate of the fault signal itself. The original formulation applies only to matched actuator faults, which are matched in the sense that they act within one or more of the system control input channels. The work of [9] was further developed by [31] to enable sensor faults to be estimated using an augmented system to make the sensor faults appear as matched actuator faults. Further research discussion on fault estimation for decentralized systems based on the so-called canonical form SMO can be found in [36] describing how local decentralized system faults can be estimated. They also discuss the relationship between the effects of interactions and faults in the estimation. The design procedure for the canonical form of SMO requires several state transformations making the algorithm complicated and not easy to implement for interconnected systems. This chapter describes an important development of sliding mode fault estimation based on the Walcott–Żak SMO which does not require state transformations and is also directly applicable to estimation for both actuator and sensor faults.

The remainder of this chapter is organized as follows. Brief reviews of the canonical and Walcott–Żak SMOs are given in Sect. 2. In Sect. 3 a decentralized SMO is described. Section 4 develops the fault estimation design. To illustrate the power of the design approach a tutorial example and a multimachine power system application example are given in Sects. 5 and 6, respectively. A conclusion to the chapter is given in Sect. 7.

In the chapter the symbol † denotes the pseudoinverse; $\|\cdot\|$ denotes the Euclidean vector norm and induced matrix norm; $\star$ denotes the symmetric part of a matrix; $\mathbf{I}_m$ denotes the m dimensional identity matrix; *s.p.d.* denotes that a matrix is symmetric positive definite; $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimal and maximal values of an eigenvalue λ, respectively.

2 Sliding Mode Observer

By duality with the sliding mode control (SMC), the introduction of a nonlinear feedback component in the SMO allows the error system trajectories to reach the sliding surface in finite time, such that the error system is insensitive to faults and perturbations satisfying certain conditions. As useful background, this section provides an outline of two SMO methods: the canonical form of SMO and the Walcott–Żak SMO. Both of these SMO strategies follow the same basic sliding mode concepts of existence and reachability and design of a sliding surface with respect to the output error. Once the sliding surface is reached, the state estimation error will be asymptotically stable, subject to bounded disturbance and uncertainty. When dealing with systems with faults (sensors, actuators, etc.) the faults must also have prescribed bounds.

Consider a linear system expressed in the following state variable form:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{G}\mathbf{f}(\mathbf{y}, \mathbf{u}, t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) \end{aligned} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\mathbf{y} \in \mathbb{R}^p$ is the output vector, $\mathbf{u} \in \mathbb{R}^m$ is the control input, and $\mathbf{f}(\mathbf{y}, \mathbf{u}, t) \in \mathbb{R}^q$ represents the perturbation including all the disturbance and uncertainty in the system. The system matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$, $\mathbf{C}$ are of appropriate dimensions. The following Assumptions are made here:

Assumption 1 The pair $(\mathbf{A}, \mathbf{C})$ is observable.

Assumption 2 The matrices $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G}$ are of full rank and $q \leq p < n$.

Assumption 3 The perturbation $\mathbf{f}(\mathbf{y}, \mathbf{u}, t)$ is unknown but bounded by

$$\|\mathbf{f}(\mathbf{y}, \mathbf{u}, t)\| \leq \alpha(\mathbf{y}, \mathbf{u}, t)$$

where $\alpha(\mathbf{y}, \mathbf{u}, t)$ is a known continuous scalar function.

Assumption 4 $\operatorname{rank}(\mathbf{CG}) = \operatorname{rank}(\mathbf{G}) = q$ and the invariant zeros of the triple $(\mathbf{A}, \mathbf{G}, \mathbf{C})$ are in the left-hand complex plane, that is,

$$\operatorname{rank} \begin{bmatrix} s\mathbf{I} - \mathbf{A} & \mathbf{G} \\ \mathbf{C} & \mathbf{0} \end{bmatrix} = n + q, \ \forall\, \mathbb{R}(s) \geq 0.$$

For simplicity, the time index is omitted and $\mathbf{f}(\mathbf{y}, \mathbf{u}, t)$ is replaced by $\mathbf{f}$ in the remainder of this section.

2.1 Canonical Form Sliding Mode Observer

In this section, the canonical form of SMO is reviewed briefly. This type of observer was first developed by Uktin [33], i.e. it can be referred to as the Uktin observer. Later, it was extended by many other researchers, e.g. [8, 37].

Suppose there exists a linear non-singular change of coordinates: $\mathbf{x} \mapsto \mathbf{T}_0\mathbf{x}$ where

$$\mathbf{T}_0 = \begin{bmatrix} \mathbf{N}_0^\top \\ \mathbf{C} \end{bmatrix}$$

and the columns of $\mathbf{N}_0 \in \mathbb{R}^{n\times(n-p)}$ span the null space of the output distribution matrix $\mathbf{C}$. With this change of coordinates, the system matrices become

$$\mathbf{T}_0\mathbf{A}\mathbf{T}_0^\top = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},\ \mathbf{T}_0\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix},\ \mathbf{T}_0\mathbf{G} = \begin{bmatrix} \mathbf{0} \\ \mathbf{G}_2 \end{bmatrix},\ \mathbf{C}\mathbf{T}_0^\top = [\mathbf{0}\ \ \mathbf{I}_\mathrm{p}].$$

Since $\mathbf{T}_0\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{y} \end{bmatrix}$, the new coordinate system can be written as

$$\begin{aligned} \dot{\mathbf{x}}_1 &= \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}\mathbf{y} + \mathbf{B}_1\mathbf{u} \\ \dot{\mathbf{y}} &= \mathbf{A}_{21}\mathbf{x}_1 + \mathbf{A}_{22}\mathbf{y} + \mathbf{B}_2\mathbf{u} + \mathbf{G}_2\mathbf{f}. \end{aligned}$$

Consider the observer of the form

$$\begin{aligned} \dot{\hat{\mathbf{x}}}_1 &= \mathbf{A}_{11}\hat{\mathbf{x}}_1 + \mathbf{A}_{12}\hat{\mathbf{y}} + \mathbf{B}_1\mathbf{u} - \mathbf{A}_{12}\mathbf{e}_\mathrm{y} \\ \dot{\hat{\mathbf{y}}} &= \mathbf{A}_{21}\hat{\mathbf{x}}_1 + \mathbf{A}_{22}\hat{\mathbf{y}} + \mathbf{B}_2\mathbf{u} - (\mathbf{A}_{22} - \mathbf{A}_{22}^\mathrm{s})\mathbf{e}_\mathrm{y} + \mathbf{v} \end{aligned} \tag{2}$$

where $\mathbf{A}_{22}^\mathrm{s}$ is a stable design matrix and $\mathbf{e}_\mathrm{y} = \mathbf{y} - \hat{\mathbf{y}}$. The switching component $\mathbf{v}$ of the observer is designed as

$$\mathbf{v} = \begin{cases} -\rho(\mathbf{y}, \mathbf{u}, t)\|\mathbf{G}\|\frac{\mathbf{P}_2\mathbf{e}_\mathrm{y}}{\|\mathbf{P}_2\mathbf{e}_\mathrm{y}\|}, & \text{if } \mathbf{e}_\mathrm{y} \neq \mathbf{0} \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

where the scalar function $\rho(\mathbf{y}, \mathbf{u}, t)$ should be chosen such that it is larger than the upper bound of the perturbation $\mathbf{f}$, i.e.

$$\rho(\mathbf{y}, \mathbf{u}, t) \geq \alpha(\mathbf{y}, \mathbf{u}, t) + \eta$$

with a positive scalar η. The matrix $\mathbf{P}_2$ is an *s.p.d* matrix satisfying the Lyapunov equation

$$\mathbf{A}_{22}^{\mathrm{s}\top}\mathbf{P}_2 + \mathbf{P}_2\mathbf{A}_{22}^\mathrm{s} = -\mathbf{Q}^\mathrm{s}$$

where $\mathbf{Q}^\mathrm{s}$ is an *s.p.d.* matrix of compatible dimension.

From (2), the reduced-order error dynamics can be derived

$$\begin{aligned}\dot{\mathbf{e}}_1 &= \mathbf{A}_{11}\mathbf{e}_1 \\ \dot{\mathbf{e}}_y &= \mathbf{A}_{21}\mathbf{e}_1 + \mathbf{A}_{22}^{\mathrm{s}}\mathbf{e}_{\mathrm{y}} + \mathbf{v} - \mathbf{G}_2\mathbf{f}.\end{aligned} \tag{3}$$

In [8], it has been proved that if the observer system has the form of (2), then the error system (3) is quadratically stable. Defining the state estimation error as $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$, then the canonical form SMO is summarized as

$$\dot{\mathbf{x}} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} - \mathbf{G}_{\mathrm{l}}\mathbf{e}_{\mathrm{y}} + \mathbf{G}_{\mathrm{n}}\mathbf{v}$$

with the linear gain

$$\mathbf{G}_{\mathrm{l}} = \mathbf{T}_0^{-1}\begin{bmatrix}\mathbf{A}_{12} \\ \mathbf{A}_{22} - \mathbf{A}_{22}^{\mathrm{s}}\end{bmatrix},$$

the nonlinear gain

$$\mathbf{G}_{\mathrm{n}} = \|\mathbf{G}_2\|\mathbf{T}_0^{-1}\begin{bmatrix}\mathbf{0} \\ \mathbf{I}_{\mathrm{p}}\end{bmatrix},$$

and the nonlinear switching function

$$\mathbf{v} = \begin{cases} -\rho(\mathbf{y}, \mathbf{u}, t)\frac{\mathbf{P}_2\mathbf{e}_{\mathrm{y}}}{\|\mathbf{P}_2\mathbf{e}_{\mathrm{y}}\|}, & \text{if } \mathbf{e}_{\mathrm{y}} \neq \mathbf{0} \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{4}$$

The existence conditions of the canonical form SMO are summarized in the following Lemma whose proof can be found in [8].

Lemma 1 *Given the system (1) represented by* $(\mathbf{A}, \mathbf{G}, \mathbf{C})$ *with* $p > q$ *and* $\mathrm{rank}(\mathbf{CG}) = q$*, then there exists a linear non-singular change of coordinates:* $\mathbf{x} \mapsto \mathbf{T}_0 x$ *such that the system* $(\bar{\mathbf{A}}, \bar{\mathbf{G}}, \bar{\mathbf{C}})$ *in the new coordinates has the following structure:*

(*1*) *The system matrix could be written as*

$$\bar{\mathbf{A}} = \begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12} \\ \begin{bmatrix}\mathbf{A}_{211} \\ \mathbf{A}_{212}\end{bmatrix} & \mathbf{A}_{22}\end{bmatrix}, \quad \mathbf{A}_{211} = [\mathbf{0}\ \mathbf{A}_{21}^0], \quad \mathbf{A}_{11} = \begin{bmatrix}\mathbf{A}_{11}^0 & \mathbf{A}_{12}^0 \\ \mathbf{0} & \mathbf{A}_{22}^0\end{bmatrix}$$

where $\mathbf{A}_{11}^0 \in \mathbb{R}^{r\times r}$ *and* $\mathbf{A}_{21}^0 \in \mathbb{R}^{(p-q)\times(n-q-r)}$ *for some* $r \geq 0$ *and the pair* $(\mathbf{A}_{22}^0, \mathbf{A}_{21}^0)$ *is observable. Furthermore, the eigenvalues of* $\mathbf{A}_{11}^0$ *are the invariant zeros of* $(\mathbf{A}, \mathbf{G}, \mathbf{C})$.

(*2*) *The disturbance distribution matrix has the form*

$$\bar{\mathbf{G}} = \begin{bmatrix}\mathbf{0} \\ \mathbf{G}_2\end{bmatrix}$$

where $\mathbf{G}_2 \in \mathbb{R}^{q\times q}$ *is a non-singular matrix.*

(*3*) *The output distribution matrix has the form*

$$\bar{\mathbf{C}} = [\mathbf{0} \;\; \mathbf{T}]$$

where $\mathbf{T} \in \mathbb{R}^{q\times q}$ *is an orthogonal matrix.*

It follows from Lemma 1 that the system in the new coordinates is

$$\begin{aligned}\dot{\bar{\mathbf{x}}} &= \bar{\mathbf{A}}\bar{\mathbf{x}} + \bar{\mathbf{B}}\mathbf{u} + \bar{\mathbf{G}}\mathbf{f}\\ \bar{\mathbf{y}} &= \bar{\mathbf{C}}\bar{\mathbf{x}}.\end{aligned}$$

Recently, [30] developed the following LMI-based algorithm for the canonical form SMO.

Step 1. Check that $\text{rank}(\mathbf{CG}) = q$ and the eigenvalues of $\mathbf{A}_{11}^0$ have negative real parts. If not, the approach is not applicable.

Step 2. Define two symmetric matrices $\bar{\mathbf{P}} \in \mathbb{R}^{n\times n}$ and $\bar{\mathbf{X}} \in \mathbb{R}^{n\times n}$, and an additional matrix $\bar{\mathbf{Y}} \in \mathbb{R}^{n\times p}$ for the linear gain matrix design.

Step 3. Form the following LMIs:

$$\begin{gathered}\text{Minimize trace}(\bar{\mathbf{X}}),\;\text{subject to}\\ \bar{\mathbf{P}} = \begin{bmatrix}\mathbf{P}_{11} & \mathbf{P}_{12}\\ \mathbf{P}_{12}^\top & \mathbf{P}_{22}\end{bmatrix} > 0,\;\; \begin{bmatrix}-\bar{\mathbf{P}} & \mathbf{I}\\ \mathbf{I} & -\bar{\mathbf{X}}\end{bmatrix} < 0\\ \begin{bmatrix}\bar{\mathbf{P}}\bar{\mathbf{A}} + \bar{\mathbf{A}}^\top\bar{\mathbf{P}} - \bar{\mathbf{C}}^\top\bar{\mathbf{Y}} & \bar{\mathbf{P}}\\ \star & -\mathbf{W}^{-1}\end{bmatrix} < 0\end{gathered}$$

where the matrices $\mathbf{W}$ and $\mathbf{V}$ are *s.p.d.* matrices.

Step 4. Part the resulting matrix $\bar{\mathbf{P}}$ to obtain $\mathbf{P}_{11}, \mathbf{P}_{12}$ and $\mathbf{P}_{22}$ as defined in (5). Compute $\bar{\mathbf{L}} = \mathbf{P}_{11}^{-1}\mathbf{P}_{12}$, $\bar{\mathbf{P}}_2 = \mathbf{P}_{22} - \mathbf{P}_{12}^\top\mathbf{P}_{11}^{-1}\mathbf{P}_{12}$ and $\mathbf{P}_2 = \mathbf{T}\bar{\mathbf{P}}_2\mathbf{T}^\top$ where $\mathbf{T}$ is the orthogonal matrix defined in Lemma 1.

Step 5. The observer gain matrices (in the new coordinates) are calculated by

$$\mathbf{G}_\text{l} = \mathbf{T}_0^\top\bar{\mathbf{P}}^{-1}\bar{\mathbf{Y}},\;\; \mathbf{G}_\text{n} = \|\mathbf{G}_2\|\mathbf{T}_0^{-1}\begin{bmatrix}-\bar{\mathbf{L}}\mathbf{T}^\top\\ \mathbf{T}^\top\end{bmatrix}.$$

The nonlinear switching part is the same as (4).

2.2 Walcott–Żak Observer

The Walcott–Żak observer is first developed in [34]. It is also referred to as a "Lyapunov-based" observer because the observer design is directly based on Lyapunov functions. However, the Walcott–Żak observer includes, in addition to the Lyapunov function formulation, an extra matrix equation constraint. In this section, the basic idea of the Walcott–Żak observer is reviewed.

Consider an SMO for the system (1) as

$$\begin{aligned} \dot{\hat{\mathbf{x}}} &= \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{L}\mathbf{C}\mathbf{e} + \mathbf{K}_{\mathrm{n}}\mathbf{v} \\ \hat{\mathbf{y}} &= \mathbf{C}\hat{\mathbf{x}} \end{aligned} \tag{5}$$

where $\hat{\mathbf{x}} \in \mathbb{R}^n$ and $\hat{\mathbf{y}} \in \mathbb{R}^p$ are the estimate state and output vector, respectively, $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ is the estimation error. $\mathbf{L}$ and $\mathbf{K}_{\mathrm{n}}$ are the matrices to be determined. v is the switching component of the SMO gain with the following form

$$\mathbf{v} = \begin{cases} \rho(\mathbf{y}, \mathbf{u}, t)\frac{\sigma}{\|\sigma\|}, & \text{if } \sigma \neq \mathbf{0} \\ \mathbf{0}, & \text{others} \end{cases} \tag{6}$$

where $\rho(\mathbf{y}, \mathbf{u}, t)$ is chosen such that $\rho(\mathbf{y}, \mathbf{u}, t) \geq \alpha(\mathbf{y}, \mathbf{u}, t) + \eta$ for a given positive scalar η and a design matrix $\mathbf{F}$. $\sigma \in \mathbb{R}^p$ is the sliding surface defined as

$$\sigma = \mathbf{F}\mathbf{C}\mathbf{e} = \mathbf{F}(\mathbf{y} - \mathbf{C}\hat{\mathbf{x}}) = \mathbf{F}\mathbf{e}_{\mathrm{y}}. \tag{7}$$

Combining (1) with (5) gives the error dynamics

$$\dot{\mathbf{e}} = (\mathbf{A} - \mathbf{L}\mathbf{C})\mathbf{e} + \mathbf{G}\mathbf{f} - \mathbf{K}_{\mathrm{n}}\mathbf{v}. \tag{8}$$

The basic idea of the Walcott–Żak observer is to calculate the matrices $\mathbf{L}$ and $\mathbf{F}$ such that $(\mathbf{A} - \mathbf{L}\mathbf{C})$ is stable and satisfies the conditions

$$\mathbf{G}^{\top}\mathbf{P} = \mathbf{F}\mathbf{C} \tag{9}$$

$$\mathbf{P}(\mathbf{A} - \mathbf{L}\mathbf{C}) + (\mathbf{A} - \mathbf{L}\mathbf{C})^{\top}\mathbf{P} = -\mathbf{Q} \tag{10}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are *s.p.d.* matrices. The main problem of the Walcott–Żak observer is how to develop the most effective solution approach for (9) and (10). In the original paper [34], the solution of (9) was assumed to be known. Later, [6] pointed out that the solvability of (9) and (10) was equivalent to the conditions outlined in the following Lemma 2.

Lemma 2 *([6]) The solution to (9) and (10) could be found if and only if* $\mathrm{rank}(\mathbf{C}\mathbf{G}) = \mathrm{rank}(\mathbf{G})$ *and the invariant zeros of* $(\mathbf{A}, \mathbf{G}, \mathbf{C})$ *are in the left-hand complex plane.*

Design the switching gain matrix $\mathbf{K}_{\mathrm{n}}$ of the observer (5) as

$$\mathbf{K}_{\mathrm{n}} = \mathbf{P}^{-1}\mathbf{C}^{\top}\mathbf{F}^{\top}. \tag{11}$$

Theorem 1 is given to obtain the observer design gains.

Theorem 1 *Given the system (1), design an observer of the form (5) with (6), (7), and (11). If the conditions (9) and (10) are satisfied, the observer (5) will track the system state trajectories robustly and become insensitive to the perturbations.*

Proof 1 Following the derivation of the classical sliding mode system theory, two steps are required to prove the Theorem: (1) Ensure that the sliding surface can be reached and (2) Prove that the error system is stable after the sliding surface is reached.

Consider a Lyapunov function

$$V_{\mathrm{e}} = \mathbf{e}^{\top}\mathbf{P}\mathbf{e}.$$

It follows from Assumption 3, (6) and (9)–(11) that the time derivative of V_{e} along the error dynamics (8) is

$$\begin{aligned}
\dot{V}_{\mathrm{e}} &= \mathbf{e}^{\top}[\mathbf{P}(\mathbf{A}-\mathbf{L}\mathbf{C}) + (\mathbf{A}-\mathbf{L}\mathbf{C})^{\top}\mathbf{P}]\mathbf{e} + 2\mathbf{e}^{\top}\mathbf{P}\mathbf{G}\mathbf{f} - 2\mathbf{e}^{\top}\mathbf{P}\mathbf{K}_{\mathrm{n}}\mathbf{v} \\
&= -\mathbf{e}^{\top}\mathbf{Q}\mathbf{e} + 2\mathbf{e}^{\top}\mathbf{C}^{\top}\mathbf{F}^{\top}\mathbf{f} - 2\mathbf{e}^{\top}\mathbf{C}^{\top}\mathbf{F}^{\top}\mathbf{v} \\
&= -\mathbf{e}^{\top}\mathbf{Q}\mathbf{e} + 2\sigma^{\top}\mathbf{f} - 2\sigma^{\top}\mathbf{v} \\
&\leq -\mathbf{e}^{\top}\mathbf{Q}\mathbf{e} + 2\|\sigma\|(\|\mathbf{f}\| - \rho(\mathbf{y},\mathbf{u},t))
\end{aligned}$$

Since $\|\mathbf{f}(\mathbf{y},\mathbf{u},t)\| \leq \alpha(\mathbf{y},\mathbf{u},t)$ and $\rho(\mathbf{y},\mathbf{u},t) \geq \alpha(\mathbf{y},\mathbf{u},t) + \eta$, then $\dot{V}_e$ can be further written as

$$\dot{V}_{\mathrm{e}} \leq -\mathbf{e}^{\top}\mathbf{Q}\mathbf{e} - 2\|\sigma\|\eta \leq 0.$$

This implies that the error dynamics are quadratically stable, and it also explains the reason why the solution of Eq. (9), i.e. $\mathbf{G}^{\top}\mathbf{P} = \mathbf{F}\mathbf{C}$, is required.

Next, consider the Lyapunov function for the sliding surface

$$V_{\mathrm{y}} = \frac{1}{2}\mathbf{e}_{\mathrm{y}}^{\top}(\mathbf{C}\mathbf{P}^{-1}\mathbf{C}^{\top})^{-1}\mathbf{e}_{\mathrm{y}}$$

whose time derivative along the error dynamics (8) is

$$\dot{V}_{\mathrm{y}} = \mathbf{e}_{\mathrm{y}}^{\top}(\mathbf{C}\mathbf{P}^{-1}\mathbf{C}^{\top})^{-1}\mathbf{C}(\mathbf{A}-\mathbf{L}\mathbf{C})\mathbf{e} + \mathbf{e}_{\mathrm{y}}^{\top}(\mathbf{C}\mathbf{P}^{-1}\mathbf{C}^{\top})^{-1}\mathbf{C}\mathbf{G}\mathbf{f} - \mathbf{e}_{\mathrm{y}}^{\top}(\mathbf{C}\mathbf{P}^{-1}\mathbf{C}^{\top})^{-1}\mathbf{C}\mathbf{K}_{\mathrm{n}}\mathbf{v}.$$

Note that $\mathbf{G} = \mathbf{P}^{-1}\mathbf{C}^{\top}\mathbf{F}^{\top}$, it follows from (6) and (11) that

$$\begin{aligned}\dot{V}_{\mathrm{y}} &= \mathbf{e}_{\mathrm{y}}^{\top}(\mathbf{C}\mathbf{P}^{-1}\mathbf{C}^{\top})^{-1}\mathbf{C}(\mathbf{A}-\mathbf{L}\mathbf{C})\mathbf{e} + \mathbf{e}_{\mathrm{y}}^{\top}\mathbf{F}^{\top}\mathbf{f} - \mathbf{e}_{\mathrm{y}}^{\top}\mathbf{F}^{\top}\mathbf{v} \\ &= \mathbf{e}_{\mathrm{y}}^{\top}(\mathbf{C}\mathbf{P}^{-1}\mathbf{C}^{\top})^{-1}\mathbf{C}(\mathbf{A}-\mathbf{L}\mathbf{C})\mathbf{e} + \sigma^{\top}\mathbf{f} - \sigma^{\top}\mathbf{v} \\ &\leq \mathbf{e}_{\mathrm{y}}^{\top}(\mathbf{C}\mathbf{P}^{-1}\mathbf{C}^{\top})^{-1}\mathbf{C}(\mathbf{A}-\mathbf{L}\mathbf{C})\mathbf{e} + \|\sigma\|(\|\mathbf{f}\| - \rho(\mathbf{y},\mathbf{u},t)) \\ &\leq \|\mathbf{e}_{\mathrm{y}}\|[(\mathbf{C}\mathbf{P}^{-1}\mathbf{C}^{\top})^{-1}\mathbf{C}(\mathbf{A}-\mathbf{L}\mathbf{C})\mathbf{e} - \|\sigma\|\eta].\end{aligned}$$

Define $0 < \tilde{\eta} \leq \|\sigma\|\eta$. Since the error dynamics (8) are asymptotic stable, the trajectory of the error dynamics will enter the following domain in finite time

$$\Omega = \{\mathbf{e} | \|(\mathbf{C}\mathbf{P}^{-1}\mathbf{C}^{\top})^{-1}\mathbf{C}(\mathbf{A}-\mathbf{L}\mathbf{C})\|\|\mathbf{e}\| < \tilde{\eta}\}.$$

Thus, $\dot{V}_{\mathrm{y}} < -\|\mathbf{e}_{\mathrm{y}}\|\tilde{\eta} < 0$, which implies that the sliding surface $\sigma = \mathbf{F}\mathbf{e}_{\mathrm{y}} = \mathbf{0}$ can be reached in finite time and remain there subsequently. This completes the proof.

2.3 *Comparison of the Two Types of SMOs*

The main differences between the canonical form SMO and Walcott–Żak observer are as follows:

• The choice of the gain matrix in the sliding surface. In the canonical form SMO, when the sliding surface is reached, it follows that $\mathbf{e}_{\mathrm{y}} = \mathbf{0}$ since $\mathbf{P}_2$ is an *s.p.d.* matrix (invertible and full rank). However, as $\mathbf{F} \in \mathbb{R}^{q\times p}$ in the conventional Walcott–Żak observer, $\mathbf{F}$ is not required to be full rank. Hence, once the sliding surface is reached, $\mathbf{e}_{\mathrm{y}} = \mathbf{0}$ cannot be guaranteed when the system reaches the sliding surface. For this reason, [8] pointed out that the Walcott–Żak observer does not necessarily track the system outputs perfectly.

• Compared with the Walcott–Żak observer, the canonical form SMO does not require the solution of (9), i.e. $\mathbf{G}^{\top}\mathbf{P} = \mathbf{F}\mathbf{C}$. However, the canonical form SMO does require triple state transformations during the design procedure [8]. A consequence of the use of these transformations is that after designing the observer gain matrices the system state must be transferred back into the original coordinates. Although the complexity of the triple transformation gives enhanced accuracy and a better understanding of the observer structure [8], it does lead to the increased SMO design complexity and this is particularly important when considering practical decentralized system design problems. This is the main reason why the Walcott–Żak observer is used here to achieve state and fault estimation for decentralized systems.

To summarize, the advantages of the Walcott–Żak observer, compared with the canonical form SMO, is that it requires no state transformation and thus has simpler structure if (9) can be solved. Recently, [4, 35] proposed a direct solution approach via the use of LMIs to this problem. It can be shown that the observer parameters are

determined by solving only a single LMI formulation. Although this method has less design freedom compared with the canonical form SMO, to a certain extent it facilitates the solvability of the Walcott–Żak observer. In the next section, a modification as well as an LMI formulation is given to the Walcott–Żak observer. Following that, the Walcott–Żak observer is extended to decentralized system problems.

3 Decentralized Walcott–Żak Observer

3.1 A Modification and an LMI Approach to Walcott–Żak Observer

It was stated in [8] that the null space of the sliding surface gain matrix $\mathbf{F}$ spanned by $\mathbf{e}_y$ in the original formulation of the Walcott–Żak observer is non-empty, i.e. $\mathbf{F}\mathbf{e}_y = \mathbf{0}$ for some $\mathbf{e}_y \neq \mathbf{0}$ (the trivial solution). Hence, if the sliding surface is designed by $\sigma = \mathbf{F}\mathbf{e}_y$, the observer does not necessarily track the system outputs perfectly. Actually, this statement is not complete, as the system will still track the output perfectly in finite time. But, when the sliding surface is reached, the output error $\mathbf{e}_y$ might not be identically zero.

In the method proposed in this section, instead of using $\mathbf{K}_n = \mathbf{P}^{-1}\mathbf{C}^\top\mathbf{F}^\top$, the gain matrix $\mathbf{K}_n = \mathbf{P}^{-1}\mathbf{C}^\top$ is chosen. The design of the Walcott–Żak observer directly follows the procedure in Sect. 2.2. In this case, the sliding surface can be designed as $\sigma = \mathbf{e}_y$. This modification does not change the essential properties of the Walcott–Żak observer whilst guaranteeing that the output error becomes zero after reaching the sliding surface.

It could be argued that the equality condition, i.e. $\mathbf{G}^\top\mathbf{P} = \mathbf{F}\mathbf{C}$, adds some complexity to the derivation and solution approach of the Walcott–Żak observer. This equality condition is complicated to solve using algebraic methods, for which more details can be found in [15]. However, [4, 35] propose a structured LMI approach by pre-structuring the *s.p.d.* matrices in a Lyapunov function. This leads to an effective solution to the Walcott–Żak observer problem.

According to [4, 35], an LMI approach is given in Theorem 2 to obtain the design matrices of the Walcott–Żak observer (5) with $\sigma = \mathbf{e}_y$.

Theorem 2 *Equations (9) and (10) can be solved by finding s.p.d. matrices* $\mathbf{W}_1 \in \mathbb{R}^{(n-m)\times(n-m)}$, $\mathbf{W}_2 \in \mathbb{R}^{p\times p}$ *and a general matrix* $\mathbf{Y} \in \mathbb{R}^{n\times p}$ *such that*

$$\tilde{\mathbf{G}}\mathbf{W}_1\tilde{\mathbf{G}}^\top + \mathbf{C}^\top\mathbf{W}_2\mathbf{C} > 0 \tag{12}$$

$$(\tilde{\mathbf{G}}\mathbf{W}_1\tilde{\mathbf{G}}^\top + \mathbf{C}^\top\mathbf{W}_2\mathbf{C})\mathbf{A} + \mathbf{A}^\top(\tilde{\mathbf{G}}\mathbf{W}_1\tilde{\mathbf{G}}^\top + \mathbf{C}^\top\mathbf{W}_2\mathbf{C}) - \mathbf{Y}\mathbf{C} - \mathbf{C}^\top\mathbf{Y}^\top < 0. \tag{13}$$

Then one can get the solution

$$\mathbf{P} = \tilde{\mathbf{G}}\mathbf{W}_1\tilde{\mathbf{G}}^\top + \mathbf{C}^\top\mathbf{W}_2\mathbf{C},\ \mathbf{L} = \mathbf{P}^{-1}\mathbf{Y},\ \mathbf{F} = \mathbf{G}^\top\mathbf{C}^\top\mathbf{W}_2$$

where $\tilde{\mathbf{G}}$ *is the orthogonal matrix of* $\mathbf{G}$*, i.e.* $\tilde{\mathbf{G}}^\top\mathbf{G} = \mathbf{0}$.

Proof 2 Following the design procedure in Sect. 2.2, the proof is similar to that given in [4, 35] and the details are omitted here.

Remark 1 The LMIs (12) and (13) can be solved directly using MATLAB LMI tools. Compared with [30], Theorem 2 simplifies the design algorithm significantly. It should also be noted that the design of the matrices $\mathbf{L}$ and $\mathbf{P}$ can be separated. From Theorem 2, it can be noted that the LMI (13) contains two variables. If there exists a matrix $\mathbf{L}$ which makes $(\mathbf{A} - \mathbf{LC})$ stable, $\mathbf{P}$ can be obtained by solving

$$(\tilde{\mathbf{G}}\mathbf{W}_1\tilde{\mathbf{G}}^\top + \mathbf{C}^\top\mathbf{W}_2\mathbf{C})\mathbf{A} + \mathbf{A}^\top(\tilde{\mathbf{G}}\mathbf{W}_1\tilde{\mathbf{G}}^\top + \mathbf{C}^\top\mathbf{W}_2\mathbf{C}) < 0.$$

This means that the linear component of the SMO and the discontinuous (nonlinear) component of the SMO can be designed separately. However, although the design of $\mathbf{L}$ does not affect the solvability of the LMIs (12) and (13), it does affect the feasible solution region of (13).

3.2 Decentralized Walcott–Żak Observer Design

For the decentralized system problem, in contrast to the single (centralized) system structure used in (1), there always exist interaction terms. However, the interaction terms of the overall system can always be regarded as bounded uncertainties if the overall system has proper local controllers. The influence of the interactions in the estimation error of each local observer is unavoidable unless each interaction satisfies the corresponding (local) rather restrictive observer matching condition: $\text{rank}(\mathbf{CG}) = \text{rank}(\mathbf{G})$. The constraint of the known upper bound of the perturbations can be relaxed by an adaptive mechanism described by [12, 13].

Now, assuming that for a decentralized system with $\mathbf{N}$ subsystems, the local controllers are well designed such that the overall system is stable and the interactions are bounded. Consider the ith subsystem in the form of

$$\begin{aligned}\dot{\mathbf{x}}_i &= \mathbf{A}_i\mathbf{x}_i + \mathbf{B}_i\mathbf{u}_i + \mathbf{G}_i\mathbf{f}_i(\mathbf{x}_i, \mathbf{u}, t) + \mathbf{M}_i\mathbf{h}_i(\mathbf{x}, \mathbf{u}, t)\\ \mathbf{y}_i &= \mathbf{C}_i\mathbf{x}_i\end{aligned} \tag{14}$$

where $\mathbf{x}_i \in \mathbb{R}^{n_i}$ is the state vector, $\mathbf{y}_i \in \mathbb{R}^{p_i}$ is the output signal, and $\mathbf{u}_i \in \mathbb{R}^m$ is the control input. $\mathbf{f}_i(\mathbf{x}_i, \mathbf{u}_i, t) \in \mathbb{R}^{q_i}$ represents the system perturbation. The matrices $\mathbf{G}_i$, $\mathbf{M}_i$ and $\mathbf{C}_i$ are of compatible dimensions. $\mathbf{h}_i(\mathbf{x}, \mathbf{u}, t) \in \mathbb{R}^{r_i}$ represents the interaction between subsystems, which do not satisfy the observer matching condition, i.e. $\text{rank}(\mathbf{C}_i\mathbf{M}_i) \neq \text{rank}(\mathbf{M}_i)$. The following Assumptions are made here.

Assumption 5 The invariant zeros of the triple $(\mathbf{A}_i, \mathbf{G}_i, \mathbf{C}_i)$ are in the left-hand complex plane, $q_i \leq p_i < n_i$, and $\text{rank}(\mathbf{C}_i\mathbf{G}_i) = \text{rank}(\mathbf{G}_i)$.

Assumption 6 The perturbation is bounded by a large enough unknown positive scalar θ_i, i.e. $\|\mathbf{f}_i\| \leq \theta_i < \infty$.

Assumption 7 The interaction term $\mathbf{h}_i(\mathbf{x}, \mathbf{u}, t)$ satisfies $\|\mathbf{h}_i(\mathbf{x}, \mathbf{u}, t)\| \leq \varpi_i$ for some unknown positive scalar ϖ_i.

It follows from (14) that the overall system can be written as

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbf{f}(\mathbf{x}, \mathbf{u}, t) + \mathbf{M}\mathbf{h}(\mathbf{x}, \mathbf{u}, t) \\ \mathbf{y} &= \mathbf{C}\mathbf{x} \end{aligned} \tag{15}$$

where $\mathbf{x} = [\mathbf{x}_1^\top, \ldots, \mathbf{x}_N^\top]^\top \in \mathbb{R}^n$, $\mathbf{y} = [\mathbf{y}_1^\top, \ldots, \mathbf{y}_N^\top]^\top \in \mathbb{R}^p$, $\mathbf{u} = [\mathbf{u}_1^\top, \ldots, \mathbf{u}_N^\top]^\top \in \mathbb{R}^m$, and

$$\begin{aligned} &\mathbf{A} = \text{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_N),\ \mathbf{B} = \text{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_N),\ \mathbf{C} = \text{diag}(\mathbf{C}_1, \ldots, \mathbf{C}_N), \\ &\mathbf{M} = \text{diag}(\mathbf{M}_1, \ldots, \mathbf{M}_N),\ \mathbf{f}(\mathbf{x}, \mathbf{u}, t) = [\mathbf{f}_1^\top(\mathbf{x}_1, \mathbf{u}_1, t), \ldots, \mathbf{f}_N^\top(\mathbf{x}_N, \mathbf{u}_N, t)]^\top, \\ &\mathbf{h}(\mathbf{x}, t) = [\mathbf{h}_1^\top(\mathbf{x}, \mathbf{u}, t), \ldots, \mathbf{h}_N^\top(\mathbf{x}, \mathbf{u}, t)]^\top,\ \|\mathbf{h}(\mathbf{x}, \mathbf{u}, t)\| \leq \varpi,\ \varpi = \text{diag}(\varpi_1, \ldots, \varpi_N), \\ &\|\mathbf{f}\| \leq \Psi < \infty,\ \Psi = [\theta_1^\top, \ldots, \theta_N^\top]^\top. \end{aligned}$$

Without causing confusion, $\mathbf{f}_i(\mathbf{x}_i, \mathbf{u}, t)$, $\mathbf{f}(\mathbf{x}, \mathbf{u}, t)$ and $\mathbf{h}(\mathbf{x}, \mathbf{u}, t)$ are replaced by $\mathbf{f}_i$, $\mathbf{f}$ and $\mathbf{h}$ in the remainder of this section.

A local observer for the subsystem (14) is designed as

$$\begin{aligned} \dot{\hat{\mathbf{x}}}_i &= \mathbf{A}_i\hat{\mathbf{x}}_i + \mathbf{B}_i\mathbf{u}_i + \mathbf{L}_i(\mathbf{y}_i - \hat{\mathbf{y}}_i) + \mathbf{K}_{\text{n}i}\mathbf{v}_i \\ \hat{\mathbf{y}}_i &= \mathbf{C}_i\hat{\mathbf{x}}_i \end{aligned} \tag{16}$$

with a nonlinear switching term $\mathbf{v}_i \in \mathbb{R}^{p_i}$ and design matrices $\mathbf{L}_i$ and $\mathbf{K}_{\text{n}i}$.

Thus, the observer system of the overall system is given by

$$\begin{aligned} \dot{\hat{\mathbf{x}}} &= \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{L}(\mathbf{y} - \hat{\mathbf{y}}) + \mathbf{K}_{\text{n}}\mathbf{v} \\ \hat{\mathbf{y}} &= \mathbf{C}\hat{\mathbf{x}} \end{aligned} \tag{17}$$

where $\mathbf{L} = \text{diag}(\mathbf{L}_1, \ldots, \mathbf{L}_N)$, $\mathbf{K}_{\text{n}} = \text{diag}(\mathbf{K}_{\text{n}1}, \ldots, \mathbf{K}_{\text{n}N})$, and $\mathbf{v} = [\mathbf{v}_1^\top, \ldots, \mathbf{v}_N^\top]^\top$.

Define the estimation errors as $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ and $\mathbf{e}_{\text{y}} = \mathbf{y} - \hat{\mathbf{y}}$, then the error system is

$$\begin{aligned} \dot{\mathbf{e}} &= (\mathbf{A} - \mathbf{L}\mathbf{C})\mathbf{e} + \mathbf{G}\mathbf{f} - \mathbf{K}_{\text{n}}\mathbf{v} + \mathbf{M}\mathbf{h} \\ \mathbf{e}_{\text{y}} &= \mathbf{C}\mathbf{e}. \end{aligned} \tag{18}$$

The LMI approach for the SMO (17) is designed using Theorem 3.

Theorem 3 *The aggregated state estimation error* $\mathbf{e}$ *is ultimately bounded if there exist matrices* $\mathbf{W}_1 = \text{diag}(\mathbf{W}_{1i}, \ldots, \mathbf{W}_{1N})$ *and* $\mathbf{W}_2 = \text{diag}(\mathbf{W}_{2i}, \ldots, \mathbf{W}_{2N})$ *with* $\mathbf{W}_{1i} \in \mathbb{R}^{(n_i - m_i)\times(n_i - m_i)}$ *and* $\mathbf{W}_{2i} \in \mathbb{R}^{p_i \times p_i}$, *and a matrix* $\mathbf{Y} = \text{diag}(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$ *with* $\mathbf{Y}_i \in \mathbb{R}^{n_i \times p_i}$, *such that*

$$\begin{aligned} &\tilde{\mathbf{G}}\mathbf{W}_1\tilde{\mathbf{G}}^\top + \mathbf{C}^\top\mathbf{W}_2\mathbf{C} > 0 \\ &(\tilde{\mathbf{G}}\mathbf{W}_1\tilde{\mathbf{G}}^\top + \mathbf{C}^\top\mathbf{W}_2\mathbf{C})\mathbf{A} + \mathbf{A}^\top(\tilde{\mathbf{G}}\mathbf{W}_1\tilde{\mathbf{G}}^\top + \mathbf{C}^\top\mathbf{W}_2\mathbf{C}) - \mathbf{Y}\mathbf{C} - \mathbf{C}^\top\mathbf{Y}^\top < 0. \end{aligned} \quad (19)$$

Then the observer gains are obtained by

$$\mathbf{P}_i = \tilde{\mathbf{G}}_i\mathbf{W}_{1i}\tilde{\mathbf{G}}_i^\top + \mathbf{C}_i^\top\mathbf{W}_{2i}\mathbf{C}_i,\ \mathbf{L}_i = \mathbf{P}_i^{-1}\mathbf{Y}_i,\ \mathbf{F}_i = \mathbf{G}_i^\top\mathbf{C}_i^\top\mathbf{W}_{2i}$$

and the nonlinear switching term is

$$\mathbf{v}_i = (\|\mathbf{F}_i\|\Psi_i + \eta_i)\frac{\mathbf{e}_{yi}}{\|\mathbf{e}_{yi}\|},\ \dot{\Psi}_i = \|\mathbf{e}_{yi}\|,\ \Psi_i(0) \geq 0 \quad (20)$$

with a constant scalar $\eta_i > 0$.

Proof 3 The proof is achieved by contradiction in the following.

Consider the following Lyapunov function for the aggregate error system (18)

$$V_e = \sum_{i=1}^{N} \mathbf{e}_i^\top\mathbf{P}_i\mathbf{e}_i. \quad (21)$$

It follows from (20) that the time derivative of (21) is

$$\begin{aligned} \dot{V}_e &= \sum_{i=1}^{N}\Big\{\mathbf{e}_i^\top[\mathbf{P}_i(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i) + (\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i)^\top\mathbf{P}_i]\mathbf{e}_i + 2\mathbf{e}_i^\top\mathbf{P}_i\mathbf{G}_i\mathbf{f}_i - 2\mathbf{e}_i^\top\mathbf{C}_i^\top\mathbf{v}_i \\ &\qquad +2\mathbf{e}_i^\top\mathbf{P}_i\mathbf{M}_i\mathbf{h}_i\Big\} \\ &= \sum_{i=1}^{N}\Big[-\mathbf{e}_i^\top\mathbf{Q}_i\mathbf{e}_i + 2\mathbf{e}_{yi}^\top\mathbf{F}_i^\top\mathbf{f}_i - 2\mathbf{e}_{yi}^\top\mathbf{v}_i + 2\mathbf{e}_i^\top\mathbf{P}_i\mathbf{M}_i\mathbf{h}_i\Big] \\ &\leq \sum_{i=1}^{N}\big[-\mathbf{e}_i^\top\mathbf{Q}_i\mathbf{e}_i + 2\|\mathbf{e}_{yi}\|\|\mathbf{F}_i\|\theta_i - 2\|\mathbf{e}_{yi}\|(\|\mathbf{F}_i\|\Psi_i + \eta_i) + 2\mathbf{e}_i^\top\mathbf{P}_i\mathbf{M}_i\mathbf{h}_i\big] \\ &\leq \sum_{i=1}^{N}\big[-\mathbf{e}_i^\top\mathbf{Q}_i\mathbf{e}_i + 2\mathbf{e}_i^\top\mathbf{P}_i\mathbf{M}_i\varpi_i + 2\|\mathbf{e}_{yi}\|(\|\mathbf{F}_i\|\theta_i - \|\mathbf{F}_i\|\Psi_i - \eta_i)\big] \end{aligned}$$

where $-\mathbf{Q}_i = \mathbf{P}_i(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i) + (\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i)^\top\mathbf{P}_i$ and according to (19), $\mathbf{Q}_i$ is an *s.p.d.* matrix.

Since $\dot{\Psi}_i = \|\mathbf{e}_y\| \geq 0$ and $\Psi_i(0) \geq 0$, the adaptive gain Ψ_i is positive and increasing (if $\|\mathbf{e}_y\| \equiv 0$, then the error system is already stable). **Assume that the adaptive gain Ψ_i is unbounded**, then there exists a time $t = t_0$ such that, for all $t > t_0$, then the adaptive gain satisfies

$$\|\mathbf{F}_i\|\Psi_i \geq \|\mathbf{F}_i\|\theta_i - \eta_i$$

which implies that

$$\begin{aligned}\dot{V}_e &\leq \sum_{i=1}^{N} \left(-\mathbf{e}_i^\top \mathbf{Q}_i \mathbf{e}_i + 2\mathbf{e}_i^\top \mathbf{P}_i \mathbf{M}_i \varpi_i\right) \\ &\leq \sum_{i=1}^{N} \|\mathbf{e}_i\| \left[-\lambda_{\min}(\mathbf{Q}_i)\|\mathbf{e}_i\| + 2\|\mathbf{P}_i \mathbf{M}_i\|\varpi_i\right]. \end{aligned} \tag{22}$$

Note that the interaction term ϖ_i is bounded and $\lambda_{\min}(\mathbf{Q}_i) > 0$. If $\|\mathbf{e}_i\|$ is unbounded, then $\dot{V}_e < 0$ (i.e. $\|\mathbf{e}_i\|$ is bounded), which leads to a contradiction. Thus, $\|\mathbf{e}_i\|$ is bounded.

Moreover, define $\mu_{0i} = \lambda_{\min}(\mathbf{Q}_i)$ and $\mu_{1i} = \sqrt{\lambda_{\max}(\mathbf{M}_i^\top \mathbf{P}_i \mathbf{P}_i \mathbf{M}_i)}$, it follows that

$$\dot{V}_e \leq \sum_{i=1}^{N} \|\mathbf{e}_i\|(-\mu_{0i} + 2\mu_{1i}\varpi_i).$$

Following [30], the estimation error $\mathbf{e}$ is ultimately bounded with respect to the set

$$\Omega_\varepsilon = \{\mathbf{e}_i : \|\mathbf{e}_i\| < 2\mu_{1i}\varpi_i/\mu_{0i} + \varepsilon_i\}$$

where $\varepsilon_i > 0$ is an arbitrarily small scalar.

Define the sliding surface as $\sigma_i = \mathbf{C}_i \mathbf{e}_i$. It follows from (18) that

$$\dot{\sigma}_i = \mathbf{C}_i \dot{\mathbf{e}}_i = \mathbf{C}_i(\mathbf{A}_i - \mathbf{L}_i \mathbf{C}_i)\mathbf{e}_i + \mathbf{C}_i \mathbf{G}_i \mathbf{f}_i - \mathbf{C}_i \mathbf{K}_{ni} \mathbf{v}_i + \mathbf{C}_i \mathbf{M}_i \mathbf{h}_i. \tag{23}$$

By defining an unknown positive scalar $\varphi_i \geq \|(\mathbf{C}_i \mathbf{P}_i \mathbf{C}_i^\top)^{-1}\mathbf{F}_i \mathbf{C}_i \mathbf{M}_i\|\varpi_i + \theta_i > 0$, the computation of the adaptive gain Ψ_i is equivalent to the computation of the estimate of φ_i. Thus, consider the following Lyapunov function for the sliding surface

$$V_\sigma = \frac{1}{2}\sum_{i=1}^{N}\left[\sigma_i^\top (\mathbf{C}_i \mathbf{P}_i \mathbf{C}_i^\top)^{-1}\sigma_i + \|\mathbf{F}_i\|(\varphi_i - \Psi_i)^2\right]. \tag{24}$$

Since $\varphi_i \geq \|(\mathbf{C}_i \mathbf{P}_i \mathbf{C}_i^\top)^{-1}\mathbf{F}_i \mathbf{C}_i \mathbf{M}_i\|\varpi_i + \theta_i > 0$ and $\varpi_i > 0$, $\varphi_i \geq \theta_i$ and using Assumption 6 and (20), the derivative of (24) along (23) is

$$\begin{aligned}\dot{V}_\sigma &= \sum_{i=1}^{N}\Big[\mathbf{e}_{yi}^\top(\mathbf{C}_i\mathbf{P}_i\mathbf{C}_i^\top)^{-1}\mathbf{C}_i(\mathbf{A}_i-\mathbf{L}_i\mathbf{C}_i)\mathbf{e}_i+\mathbf{e}_{yi}^\top\mathbf{F}_i^\top\mathbf{f}_i-\mathbf{e}_{yi}^\top\mathbf{v}_i\\ &\qquad +\mathbf{e}_{yi}^\top(\mathbf{C}_i\mathbf{P}_i\mathbf{C}_i^\top)^{-1}\mathbf{C}_i\mathbf{M}_i\mathbf{h}_i-\|\mathbf{F}_i\|(\varphi_i-\Psi_i)\|\mathbf{e}_{yi}\|\Big]\\ &\le \sum_{i=1}^{N}\Big[\|\mathbf{e}_{yi}\|\|(\mathbf{C}_i\mathbf{P}_i\mathbf{C}_i^\top)^{-1}\mathbf{C}_i\|\|(\mathbf{A}_i-\mathbf{L}_i\mathbf{C}_i)\mathbf{e}_i+\mathbf{M}_i\mathbf{h}_i\|\\ &\qquad +\|\mathbf{e}_{yi}\|(\|\mathbf{F}_i\|\theta_i-\|\mathbf{F}_i\|\Psi_i-\eta_i)-\|\mathbf{e}_{yi}\|\|\mathbf{F}_i\|(\varphi_i-\Psi_i)\Big]\\ &\le \|\mathbf{e}_{yi}\|\Big[\|(\mathbf{C}_i\mathbf{P}_i\mathbf{C}_i^\top)^{-1}\mathbf{C}_i\|(\|(\mathbf{A}_i-\mathbf{L}_i\mathbf{C}_i)\mathbf{e}_i\|+\|\mathbf{M}_i\mathbf{h}_i\|)-\eta_i\Big].\end{aligned}$$

From the above, define the gain η_i as

$$\eta_i \ge \|(\mathbf{C}_i\mathbf{P}_i\mathbf{C}_i^\top)^{-1}\mathbf{C}_i\|\Big[(\mathbf{A}_i-\mathbf{L}_i\mathbf{C}_i)(2\mu_{1i}\varpi_i/\mu_{0i}+\varepsilon_i)+\|\mathbf{M}_i\|\varpi_i\Big]+\eta_{0i}$$

with a positive scalar η_{0i}. Note that in finite time $\mathbf{e}_i \in \Omega_\varepsilon$, which implies that $\|\mathbf{e}_i\| < 2\mu_{1i}\varpi_i/\mu_{0i}+\varepsilon_i$. It thus follows that

$$\dot{V}_\sigma \le -\sum_{i=1}^{N}\eta_{0i}\|\sigma_i\| \le 0. \tag{25}$$

Integrating both sides of (25) from 0 to t yields

$$V_\sigma(0) \ge V_\sigma(t)+\int_0^t\Big(\sum_{i=1}^{N}\eta_{0i}\|\sigma_i(\tau)\|\Big)d\tau \ge \int_0^t\Big(\sum_{i=1}^{N}\eta_{0i}\|\sigma_i(\tau)\|\Big)d\tau. \tag{26}$$

Thus, for all $t \in [0,\infty)$, $\int_0^t\left(\sum_{i=1}^{N}\eta_{0i}\|\sigma_i(\tau)\|\right)d\tau$ is always less or equal to $V_\sigma(0)$. Moreover, since $V_\sigma(0)$ is positive and bounded, it follows from the Barbalat Lemma that

$$\lim_{t\to\infty}(\eta_{0i}\|\sigma_i(t)\|)=0,$$

i.e. the sliding surface function is reachable. The error system trajectory converges to the sliding manifold $\sigma_i = \mathbf{e}_{yi} = \mathbf{0}$. Furthermore, it should be noted that because of (26), $V_\sigma(t) \le V_\sigma(0)$, i.e. $V_\sigma(t)$ is bounded. This also implies that $\|\sigma_i\|$ and Ψ_i are bounded for all $t > 0$.

The boundedness of the adaptive gain Ψ_i leads to a contradiction, since the previous proof was achieved based on the Assumption that the adaptive gain Ψ_i is unbounded. However, it can be shown below that the previous proof is valid, i.e. the estimation error $\mathbf{e}$ is still ultimately bounded and the sliding surface function is reachable with bounded Ψ_i.

Assume that there is a semi-positive scalar such that $\Psi_{is} \geq 2\|\mathbf{C}_i\|(\|\mathbf{F}_i\|\theta_i - \|\mathbf{F}_i\|\Psi_i - \eta_i)$ and $\Psi_{is} = 0$ if the right-hand side of this inequality is less than zero. Then the inequality (22) can be rewritten as

$$\dot{V}_{\mathrm{e}} \leq \sum_{i=1}^{N} \left\{ \|\mathbf{e}_i\| \Big[-\lambda_{\min}(\mathbf{Q}_i)\|\mathbf{e}_i\| + 2\|\mathbf{P}_i\mathbf{M}_i\|\varpi_i + \Psi_{is} \Big] \right\}.$$

Note that $\|\mathbf{e}_i\|$ is still bounded and the set becomes

$$\bar{\Omega}_{\varepsilon} = \{\mathbf{e}_i : \|\mathbf{e}_i\| < (2\mu_{1i}\varpi_i + \Psi_{is})/\mu_{0i} + \varepsilon_i\}. \tag{27}$$

Following the same proof procedure from (24) to (25), if

$$\eta_i \geq \|(\mathbf{C}_i\mathbf{P}_i\mathbf{C}_i^{\top})^{-1}\mathbf{C}_i\| \Big[(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i)(2\mu_{1i}\varpi_i/\mu_{0i} + \Psi_{is} + \varepsilon_i) + \|\mathbf{M}_i\|\varpi_i \Big] + \eta_{0i}$$

is defined, then (25) still holds, leading to $\mathbf{e}_{\mathrm{y}i} = \mathbf{0}$ in finite time. This completes the proof.

Remark 2 From the proof above, it can be seen that the bounds for the interaction and the perturbation are not required to be known *a priori*. This highlights the benefit of using the adaptive mechanism.

Remark 3 The set (27) shows that the tracking accuracy is affected by (1). The eigenvalues of $\mathbf{Q}_i$, i.e. the choice of the linear gain matrix $\mathbf{L}$ and *s.p.d* matrix $\mathbf{P}$, and (2) the norm bound of the interactions ϖ_i and (3) the values of the sliding gain functions Ψ_i and the parameters η_i. Thus, one can adjust the design strategy based on these relationships.

3.3 Interaction Compensation

In Sect. 3.2, a decentralized Walcott–Żak observer was proposed for decentralized system whose overall system contains interactions not satisfying the matching condition, i.e. $\mathrm{rank}(\mathbf{C}_i\mathbf{M}_i) \neq \mathrm{rank}(\mathbf{M}_i)$. However, in order to use the maximum capability of the nonlinear term, a partition strategy is proposed in this section to divide the interactions into two parts: matched part and unmatched part. In this way, the nonlinear switching terms of the proposed decentralized Walcott–Żak observer can compensate both the perturbations and matched part of the interactions.

The partition strategy is as follows:

(1) Find matrices $\mathbf{G}_{\mathrm{m}i} \in \mathbb{R}^{n_i \times p_i}$ and $\bar{\mathbf{G}}_i \in \mathbb{R}^{p_i \times q_i}$, $i = 1, \ldots, N$, satisfying $\mathbf{G}_i = \mathbf{G}_{\mathrm{m}i}\mathbf{G}_i$ and $\mathrm{rank}(\mathbf{C}_i\mathbf{G}_{\mathrm{m}i}) = \mathrm{rank}(\mathbf{G}_{\mathrm{m}i}) = p_i$, such that for the ith subsystem, the invariant zeros of the triple $(\mathbf{A}_i, \mathbf{G}_{\mathrm{m}i}, \mathbf{C}_i)$ are in the left-hand complex plane.

(2) Find compatible matrices $\bar{\mathbf{M}}_{1i},\ \bar{\mathbf{M}}_{2i} \in \mathbb{R}^{p_i \times r_i}$ satisfying

$$\mathbf{M}_i = \mathbf{G}_{\mathrm{m}i}\bar{\mathbf{M}}_{1i} + \mathbf{G}_{\mathrm{m}i}\bar{\mathbf{M}}_{2i} = \mathbf{M}_{1i} + \mathbf{M}_{2i}.$$

According to [35], there always exist matrices $\mathbf{P}_i$ and $\mathbf{F}_i$ such that $\mathbf{G}_{\mathrm{m}i}^{\top}\mathbf{P}_i = \mathbf{F}_i\mathbf{C}_i$. Thus, building an SMO based on $\mathbf{P}_i$ and $\mathbf{F}_i$, it is easy to verify that the nonlinear terms of SMO can compensate both the perturbations and interactions, i.e. $\mathbf{G}_{\mathrm{m}i}\mathbf{f}_i(\mathbf{x}_i, \mathbf{u}_i, t)$ and $\mathbf{M}_{1i}\mathbf{h}_i(\mathbf{x}, \mathbf{u}, t)$. Therefore, the error domain is contracted because only parts of the interactions $\mathbf{M}_{2i}\mathbf{h}_i(\mathbf{x}, \mathbf{u}, t)$ need to be handled. However, it should be noted that there sometimes exists more than one partition results, which might lead to different obtained observer gains.

The switching nonlinear term of the decentralized Walcott–Żak observer can reject the perturbations and the matched part of the interactions in the error system. In this respect, if the switching nonlinear part, the perturbations, and the matched part of the interactions are taken off, then the observer design problem becomes an observer that can deal with the remaining unmatched part of the interactions, for which H_∞ optimization can be applied. Considering the overall system (15) and the observer (17), the observer gain matrix $\mathbf{L}$ can be obtained by solving the LMI.

$$\text{Minimize } \gamma, \text{ subject to } \mathbf{P} > 0$$

$$\begin{bmatrix} \mathbf{A}^{\top}\mathbf{P} + \mathbf{P}\mathbf{A} - \mathbf{W}\mathbf{C} - \mathbf{C}^{\top}\mathbf{W}^{\top} & \mathbf{P}\mathbf{M}_2 & \mathbf{I} \\ \star & -\gamma\mathbf{I} & 0 \\ \star & \star & -\gamma\mathbf{I} \end{bmatrix} < 0 \tag{28}$$

where $\mathbf{M}_2 = \mathrm{diag}(\mathbf{M}_{21}, \ldots, \mathbf{M}_{2N})$ and $\gamma > 0$ is a constant scalar. Then the observer gain is determined by $\mathbf{L} = \mathbf{P}^{-1}W$.

Note that the main problem of building Walcott–Żak observer is to find the matrices $\mathbf{P}$ and $\mathbf{F}$ satisfying $\mathbf{G}^{\top}\mathbf{P} = \mathbf{F}\mathbf{C}$. Define $\mathbf{P} = \tilde{\mathbf{G}}_{\mathrm{m}i}\mathbf{W}_1\tilde{\mathbf{G}}_{\mathrm{m}i}^{\top} + \mathbf{C}^{\top}\mathbf{W}_2\mathbf{C}$ with

$$\mathbf{W}_1 = \mathrm{diag}(\mathbf{W}_{11}, \ldots, \mathbf{W}_{1N}),\ \mathbf{W}_2 = \mathrm{diag}(\mathbf{W}_{21}, \ldots, \mathbf{W}_{2N}),\ \mathbf{W} = \mathrm{diag}(\mathbf{W}_1, \ldots, \mathbf{W}_N)$$

where $\mathbf{W}_{1i} \in \mathbb{R}^{(n_i - p_i)\times(n_i - p_i)}$ and $\mathbf{W}_{2i} \in \mathbb{R}^{p_i \times p_i}$ are *s.p.d.* matrices, and $\mathbf{W}_i \in \mathbb{R}^{n_i \times p_i}$. The observer can then be designed using (17) and (20), with the observer gain $\mathbf{L}$ obtained by solving the LMI (28). This approach can attenuate the worst case influence from interactions and perturbations acting in the state estimation error system, which can be proved using a similar procedure as given for Theorem 2.

3.4 Optimal Observer Design

It should be noted that the LMI (28) is solved by minimizing the constant scalar γ to a small value, which often leads to an observer gain $\mathbf{L}$ of large magnitude. For practical application of an SMO, it is more acceptable that the size of the norm of

the gain matrix $\mathbf{L}$ is minimized as much as possible, e.g. using an LMI formulation. Therefore, a trade-off needs to be made between the observer gain magnitude and the uncertainty attenuation level. To achieve this, the norms of $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}$ are restricted under some levels as follows.

Denote

$$\mathbf{W}^\top \mathbf{W} < \beta_{\mathrm{W}} \mathbf{I}, \quad \beta_{\mathrm{W}} > 0$$

and

$$\tilde{\mathbf{G}} \mathbf{W}_1 \tilde{\mathbf{G}}^\top + \mathbf{C}^\top \mathbf{W}_2 \mathbf{C} < \beta_{\mathrm{p}} \mathbf{I}, \quad \beta_{\mathrm{p}} > 0.$$

These two inequalities can be written in the following LMI form:

$$\begin{bmatrix} \beta_{\mathrm{W}} \mathbf{I} & \mathbf{W} \\ \star & \mathbf{I} \end{bmatrix} > 0 \text{ and } \begin{bmatrix} \beta_{\mathrm{p}} \mathbf{I} & \mathbf{I} \\ \star & \tilde{\mathbf{G}} \tilde{W}_1 \tilde{\tilde{G}}^\top + \mathbf{C}^\top \mathbf{W}_2 \mathbf{C} \end{bmatrix} > 0.$$

From these constraints, the desired bound could be given by

$$\mathbf{L}\mathbf{L}^\top = \mathbf{P}^{-1} \mathbf{W} \mathbf{W}^\top \mathbf{P}^{-1} < \beta_{\mathrm{W}} \beta_{\mathrm{p}}^2 \mathbf{I}.$$

With these modifications, the H_∞ optimization problem with respect to minimizing the effects (in the estimation error) of the unmatched perturbations and unmatched interactions in the SMO becomes

$$\begin{gathered} \text{Minimize } \gamma + \beta_{\mathrm{W}} + \beta_{\mathrm{p}}, \text{ subject to } \mathbf{P} = \tilde{\mathbf{G}} \mathbf{W}_1 \tilde{\mathbf{G}}^\top + \mathbf{C}^\top \mathbf{W}_2 \mathbf{C} > 0 \\ \begin{bmatrix} \beta_{\mathrm{W}} \mathbf{I} & \mathbf{W} \\ \star & \mathbf{I} \end{bmatrix} > 0, \quad \begin{bmatrix} \beta_{\mathrm{p}} \mathbf{I} & \mathbf{I} \\ \star & \tilde{\mathbf{G}} \mathbf{W}_1 \tilde{\mathbf{G}}^\top + \mathbf{C}^\top \mathbf{W}_2 \mathbf{C} \end{bmatrix} > 0 \\ \begin{bmatrix} \mathbf{A}^\top \mathbf{P} + \mathbf{P}\mathbf{A} - \mathbf{W}\mathbf{C} - \mathbf{C}^\top \mathbf{W}^\top & \mathbf{P}\mathbf{M}_2 & \mathbf{I} \\ \star & -\gamma \mathbf{I} & 0 \\ \star & \star & -\gamma \mathbf{I} \end{bmatrix} < 0. \end{gathered} \tag{29}$$

4 Decentralized Walcott–Żak Observer-Based Fault Estimation

The mathematical representation of the ith subsystem of the LSS subject to actuator and sensor faults is

$$\begin{aligned} \dot{\mathbf{x}}_i &= \mathbf{A}_i \mathbf{x}_i + \mathbf{B}_i \mathbf{u}_i + \mathbf{G}_i \mathbf{f}_{\mathrm{a}i} + \mathbf{M}_i \mathbf{h}_i(\mathbf{x}, \mathbf{u}, t) \\ \mathbf{y}_i &= \mathbf{C}_i \mathbf{x}_i + \mathbf{E}_i \mathbf{f}_{\mathrm{s}i} \end{aligned} \tag{30}$$

where $\mathbf{f}_{ai} \in \mathbb{R}^{q_i}$ and $\mathbf{f}_{si} \in \mathbb{R}^{l_i}$ denote the actuator and sensor faults, respectively. The other terms are defined the same as those in (14). Assume that $\|\mathbf{f}_{ai}\| \leq \theta_{ai} < \infty$ and $\|\mathbf{f}_{si}\| \leq \theta_{si} < \infty$ for some large enough unknown positive scalars θ_{ai} and θ_{si}.

In order to illustrate the fault estimation strategy better, it is assumed in the following design procedure that either an actuator or sensor fault is to be reconstructed whilst the other (sensor or actuator fault) is absent. However, it is shown at the end of this section that the proposed fault estimation strategy can estimate the actuator and sensor faults simultaneously.

4.1 Actuator Fault Estimation

If the systems (30) contain only actuator faults, i.e. $\mathbf{f}_{ai} \neq \mathbf{0}$ and $\mathbf{f}_{si} = \mathbf{0}$, the decentralized Walcott–Żak observer proposed in Sect. 2.2 can be used to reconstruct the fault signals effectively. According to the description in Sect. 2.2, the local SMOs are designed in the form of (16), and the local state estimation error system is given in the form of (18).

Following the observer design procedure in Sect. 2.2, the reachability of the decentralized Walcott–Żak observer is ensured and the state estimation error is bounded: (1) In finite time, the sliding surface is reached and maintained, i.e. $\mathbf{C}_i\mathbf{e}_i = \mathbf{0}$, and $\mathbf{C}_i\dot{\mathbf{e}}_i = \mathbf{0}$. (2) After reaching the sliding surface, the state estimation errors are bounded.

For simplicity, assume that after finite time, the state estimation error will enter the domain

$$\bar{\Omega}_\varepsilon = \{\mathbf{e}_i : \|\mathbf{e}_i\| < \|\mathbf{e}_{Bi}\|\}$$

where $\|\mathbf{e}_{Bi}\|$, $i = 1, 2, \ldots, N$, are the estimation boundary constants which can be calculated following the proof procedure in Sect. 2.2.

Since the expansion matrix $\mathbf{G}_{mi}$ for the ith subsystem is used, the rank condition is satisfied, i.e. $\operatorname{rank}(\mathbf{C}_i\mathbf{G}_{mi}) = \operatorname{rank}(\mathbf{G}_{mi}) = p_i$. Moreover, $\mathbf{G}_i = \mathbf{G}_{mi}\bar{\mathbf{G}}_i$ and $\mathbf{M}_i = \mathbf{G}_{mi}\bar{\mathbf{M}}_{1i} + \tilde{\mathbf{G}}_{mi} = \mathbf{M}_{1i} + \mathbf{M}_{2i}$ can be constructed. This means that all the information about the $\mathbf{f}_{ai}$ and some of the information about interactions, i.e. $\mathbf{M}_{1i}\mathbf{h}_i(\mathbf{x}, \mathbf{u}, t)$, are preserved in the output. In this case, using the equivalent output error injection signal, after the sliding motion takes place, yields

$$\mathbf{0} = \mathbf{C}_i(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i)\mathbf{e}_i + \mathbf{C}_i\mathbf{G}_i\mathbf{f}_{ai} - \mathbf{C}_i\mathbf{K}_{ni}\mathbf{v}_{eq,i} + \mathbf{C}_i\mathbf{M}_i\mathbf{h}_i \tag{31}$$

where $\mathbf{v}_{eq,i}$ is the equivalent output error injection signal which plays the same role as the "equivalent control" in sliding mode control [8]. The equivalent output injection signal is represented by the values of the nonlinear switching terms $\mathbf{v}_i$ defined by (20), which is necessary to maintain the sliding motion.

Note that

$$\mathbf{C}_i\mathbf{G}_i = \mathbf{C}_i\mathbf{P}_i^{-1}\mathbf{C}_i^\top\mathbf{F}_i^\top\bar{\mathbf{G}}_i, \quad \mathbf{C}_i\mathbf{K}_{ni} = \mathbf{C}_i\mathbf{P}_i^{-1}\mathbf{C}_i^\top$$

and

$$\mathbf{C}_i\mathbf{M}_i = \mathbf{C}_i\mathbf{P}_i^{-1}\mathbf{C}_i^{\top}\mathbf{F}_i^{\top}\bar{\mathbf{M}}_{1i} + \mathbf{C}_i\mathbf{M}_{2i}.$$

Multiplying each side of (31) by $(\mathbf{C}_i\mathbf{P}_i^{-1}\mathbf{C}_i^{\top})^{-1}$, it follows that

$$\mathbf{v}_{\mathrm{eq},i} = \mathbf{F}_i^{\top}(\bar{\mathbf{G}}_i\mathbf{f}_{\mathrm{a}i} + \bar{\mathbf{M}}_{1i}\mathbf{h}_i) + (\mathbf{C}_i\mathbf{P}_i^{-1}\mathbf{C}_i^{\top})^{-1}\mathbf{C}_i[(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i)\mathbf{e}_i + \mathbf{M}_{2i}\mathbf{h}_i].$$

Thus, the actuator fault is given by

$$\mathbf{f}_{\mathrm{a}i} = \bar{\mathbf{G}}_i^{\dagger}(\mathbf{F}_i^{\dagger})^{\top}\mathbf{v}_{\mathrm{eq},i} - \bar{\mathbf{G}}_i^{\dagger}\bar{\mathbf{M}}_{1i}\mathbf{h}_i - \chi_i[(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i)\mathbf{e}_i + \mathbf{M}_{2i}\mathbf{h}_i] \tag{32}$$

where $\chi_i = \bar{\mathbf{G}}_i^{\dagger}(\mathbf{F}_i^{\dagger})^{\top}(\mathbf{C}_i\mathbf{P}_i^{-1}\mathbf{C}_i^{\top})^{-1}\mathbf{C}_i$. Using the equivalent output injection, the actuator fault estimations are $\hat{\mathbf{f}}_{\mathrm{a}i} = \bar{\mathbf{G}}_i^{\dagger}(\mathbf{F}_i^{\dagger})^{\top}\mathbf{v}_{\mathrm{eq},i}$, $i = 1, 2, \ldots, N$. Then consider the following two cases:

Case 1. It holds that $\bar{\mathbf{G}}_i^{\dagger}\mathbf{M}_{1i} = \mathbf{0}$ and $\mathbf{M}_{2i} = \mathbf{0}$.

In this case, the interactions satisfy $\mathrm{rank}(\mathbf{C}_i\mathbf{M}_i) = \mathrm{rank}(\mathbf{M}_i)$ and they act within a different channel of the system compared with the faults. It is easy to verify that the switching nonlinear term of the observer can completely compensate the interactions. Thus, the state estimation error $\mathbf{e}$ is quadratically stable and (32) can then be rewritten as

$$\mathbf{f}_{\mathrm{a}i} = \bar{\mathbf{G}}_i^{\dagger}(\mathbf{F}_i^{\dagger})^{\top}\mathbf{v}_{\mathrm{eq},i} - \chi_i(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i)\mathbf{e}_i.$$

Since $\lim_{t\to\infty}\mathbf{e}_i(t) = \mathbf{0}$, it then follows that

$$\mathbf{f}_{\mathrm{a}i} = \bar{\mathbf{G}}_i^{\dagger}(\mathbf{F}_i^{\dagger})^{\top}\mathbf{v}_{\mathrm{eq},i} = \hat{\mathbf{f}}_{\mathrm{a}i}$$

which means that the faults can be estimated precisely.

Case 2. It holds that $\bar{\mathbf{G}}_i^{\dagger}\bar{\mathbf{M}}_{1i} \neq \mathbf{0}$ and $\mathbf{M}_{2i} \neq \mathbf{0}$.

It follows from (32) that

$$\mathbf{f}_{\mathrm{a}i} - \hat{\mathbf{f}}_{\mathrm{a}i} = -\bar{\mathbf{G}}_i^{\dagger}\bar{\mathbf{M}}_{1i}\mathbf{h}_i - \chi_i[(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i)\mathbf{e}_i + \mathbf{M}_{2i}\mathbf{h}_i].$$

Following a similar procedure to Theorem 2, the local state estimation error system reaches the domain recalled below in finite time

$$\bar{\Omega}_{\varepsilon} = \{\mathbf{e}_i : \|\mathbf{e}_i\| < \|\mathbf{e}_{\mathrm{B}i}\|\}.$$

Meanwhile, as assumed in Sect. 2.2, $\|\mathbf{h}_i\| \leq \varpi_i$, it then follows that

$$\|\mathbf{f}_{\mathrm{a}i} - \hat{\mathbf{f}}_{\mathrm{a}i}\| = (\|\bar{\mathbf{G}}_i^{\dagger}\bar{\mathbf{M}}_{1i}\| + \|\chi_i\mathbf{M}_{2i}\|)\varpi_i + \|\chi_i(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}_i)\|\|\mathbf{e}_{\mathrm{B}i}\|. \tag{33}$$

If the distribution matrices $\mathbf{M}_i$ of the interactions do not satisfy the rank condition, i.e. $\text{rank}(\mathbf{C}_i\mathbf{M}_i) \neq \text{rank}(\mathbf{M}_i)$, the reconstructed faults cannot track the real faults precisely. However, according to (33), it can be seen that the fault estimation errors are bounded.

Remark 4 It should be noted that the discontinuous part in (20), i.e. $\text{sign}(\mathbf{e}_y) = \mathbf{e}_y/\|\mathbf{e}_y\|$ is either 0 or 1. This leads to the long-known twisting phenomenon of the sliding mode theory. Edwards and Spurgeon [8] pointed out that by adding an appropriate small positive scalar in the denominator, the twisting can be decreased to an acceptable extent and thus a continuous fault estimation signal will result. Following this idea, the nonlinear switching part of the observer (20) is modified into

$$\mathbf{v}_i = (\|\mathbf{F}_i\|\Psi_i + \eta_i)\frac{\mathbf{e}_y}{\|\mathbf{e}_y\| + \Delta}, \ \dot{\Psi}_i = \|\mathbf{e}_{yi}\|, \ \Psi_i(0) \geq 0.$$

where $\eta_i > 0$ is a constant scalar and Δ is a small positive scalar.

The proposed actuator fault estimation strategy based on the decentralized Walcott–Żak observer for each subsystem can be outlined in Fig. 1.

Remark 5 The robust fault estimation given in the above analysis and design procedure can be applied to a control system in which the uncertainties and/or faults can be compensated within the feedback control structure [20, 21]. However, one powerful

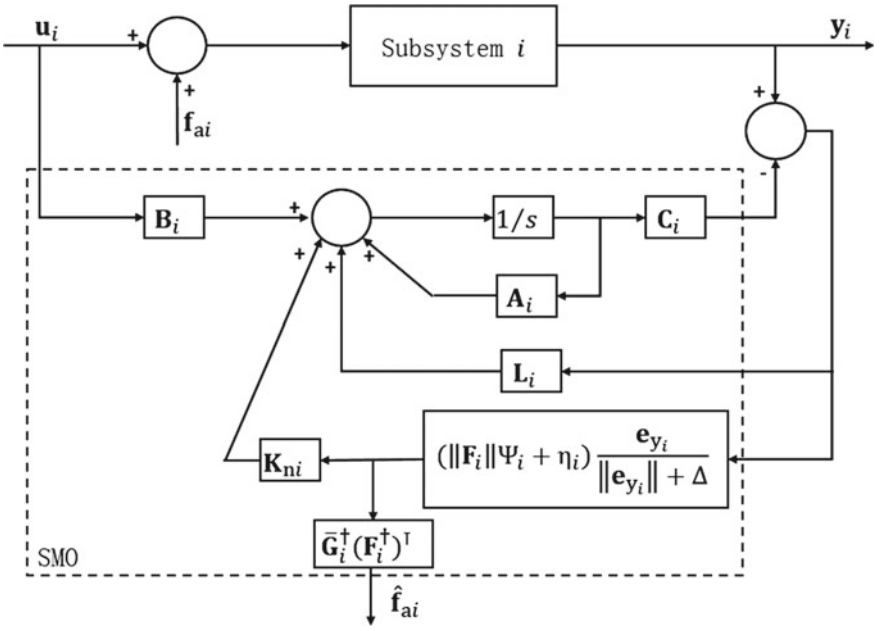


Fig. 1 Actuator fault estimation strategy for the ith subsystem

method of robust control, the sliding mode control (SMC) has a property capable of compensating matched perturbations (complete compensation in the ideal case) [12, 13] or matched faults [14]. Since most of the actuator faults are matched, if the system can tolerate the actuator faults with SMC, is it still valuable to use the proposed fault estimation method? It seems that the functions of the actuator fault tolerance using SMC and the actuator fault estimation using SMO overlap. However, looking at the problem from a practical application standpoint, the answer to this question is positive due to the following reasons:

(1) If the actuator fault is an actuator *outage* or the actuator is *stuck*, the actuator is no longer able to function. Thus, even with powerful SMC, the fault cannot be tolerated. The only solution to this problem is to replace the actuator. This is the reason why practical plants must use redundant actuators. This is the case, for example in aircraft where wing actuators are duplicated. Hence, the actuator fault estimation using SMO is necessary to determine when the fault is significant for the second actuator to be brought into action and to replace the assumed faulty one.

(2) If the actuator loses *effectiveness*, with SMC, the control input required by the system stabilization or tracking makes the actuator overload. Thus, although the actuator faults can be tolerated with SMC, it is still valuable to detect these faults so one can decide whether the actuator should be replaced.

(3) If the actuator faults are considered as external disturbances (uncertainties), and the compensation signal produced by SMC does not exceed the saturation of the actuator, then there is no need to estimate the disturbance.

From the above discussion, it can be seen that for the case of actuator faults, the best way is to detect (estimate) them instead of attempting to compensate them with SMC. This deduction leads to another discussion about the capability of the SMC. In practical problems, the actuator operates within practical bounds $\pm\omega$ where ω is a positive scalar. When applying SMC, since the choice of the gain ρ (the gain of the nonlinear component of the SMO defined in (6)) should be larger than the perturbation bound, i.e. $\rho > \|\mathbf{f}\|$, the actuator bounds $\pm\omega$ limit the capability of the SMC since $\rho < |\omega|$. Thus, if an actuator fault occurs and its magnitude exceeds the bounds $\pm\omega$, the SMC cannot tolerate it. In this case, a fault detection function should be considered to raise an alarm. When considering the combination of the SMC and SMO, the SMC is designed to tolerate the disturbances and uncertainties while the SMO is used to estimate the actuator faults in order to provide fault information to the engineer.

4.2 Sensor Fault Estimation

In contrast to actuator fault estimation, the equivalent output injection is not suitable to be used alone to reconstruct the sensor fault. The work by [8] proposed a case when the sensor fault varies slowly, i.e. $\dot{\mathbf{f}}_{si} \approx \mathbf{0}$, the equivalent output injection can still be applied to get the approximate fault estimation. However, this constraint is rather restrictive for sensor faults.

An output filter is proposed by [31] as a compensator to reconstruct the output signals based on use of the canonical form SMO observer. The idea of this strategy is to use a filter to filter the outputs so that the augmented system (combination of original system and filters) has reliable measurements. Thus, the sensor faults can be treated as actuator faults. This section proposes a sensor fault estimation algorithm using the decentralized Walcott–Żak observer as described in Sect. 2.2.

Consider the local subsystem state variables $\mathbf{z}_{\mathrm{f}i} \in \mathbb{R}^{p_i}$ for the filter

$$\dot{\mathbf{z}}_{\mathrm{f}i} = -\mathbf{A}_{\mathrm{f}i}\mathbf{z}_{\mathrm{f}i} + \mathbf{A}_{\mathrm{f}i}\mathbf{y}_i \tag{34}$$

where $(-\mathbf{A}_{\mathrm{f}i})$ is a non-singular Hurwitz matrix and generally selected as $\mathbf{A}_{\mathrm{f}i} = a\mathbf{I}_{\mathrm{p}_i}$ with a positive scalar a.

Substituting the system output Eq. (30) into the filter (34) yields

$$\dot{\mathbf{z}}_{\mathrm{f}i} = -\mathbf{A}_{\mathrm{f}i}\mathbf{z}_{\mathrm{f}i} + \mathbf{A}_{\mathrm{f}i}\mathbf{C}_i\mathbf{x}_i + \mathbf{A}_{\mathrm{f}i}\mathbf{E}_i\mathbf{f}_{si}.$$

Then combining the filter and the state equation gives the following augmented system:

$$\begin{aligned}\begin{bmatrix}\dot{\mathbf{x}}_i\\ \dot{\mathbf{z}}_{\mathrm{f}i}\end{bmatrix} &= \begin{bmatrix}\mathbf{A}_i & \mathbf{0}\\ \mathbf{A}_{\mathrm{f}i}\mathbf{C}_i & -\mathbf{A}_{\mathrm{f}i}\end{bmatrix}\begin{bmatrix}\mathbf{x}_i\\ \mathbf{z}_{\mathrm{f}i}\end{bmatrix} + \begin{bmatrix}\mathbf{B}_i\\ \mathbf{0}\end{bmatrix}\mathbf{u}_i + \begin{bmatrix}\mathbf{M}_i\\ \mathbf{0}\end{bmatrix}\mathbf{h}_i + \begin{bmatrix}\mathbf{0}\\ \mathbf{A}_{\mathrm{f}i}\mathbf{E}_i\end{bmatrix}\mathbf{f}_{si}\\ \mathbf{z}_{\mathrm{f}i} &= [\mathbf{0}\ \ \mathbf{I}_{\mathrm{p}_i}]\begin{bmatrix}\mathbf{x}_i\\ \mathbf{z}_{\mathrm{f}i}\end{bmatrix}.\end{aligned} \tag{35}$$

Denote $\mathbf{z}_i = [\mathbf{x}_i^\top\ \mathbf{z}_{\mathrm{f}i}^\top]^\top$, then the augmented system can be written in the following system containing the actuator fault:

$$\begin{aligned}\dot{\mathbf{z}}_i &= \mathbf{A}_{0i}\mathbf{z}_i + \mathbf{B}_{0i}\mathbf{u}_i + \mathbf{M}_{0i}\mathbf{h}_i + \mathbf{G}_{0i}\mathbf{f}_{si}\\ \mathbf{z}_{\mathrm{f}i} &= \mathbf{C}_{0i}\mathbf{z}_i\end{aligned} \tag{36}$$

where $\mathbf{A}_{0i} \in \mathbb{R}^{(n_i+p_i)\times(n_i+p_i)}$, $\mathbf{B}_{0i} \in \mathbb{R}^{(n_i+p_i)\times m_i}$, $\mathbf{M}_{0i} \in \mathbb{R}^{(n_i+p_i)\times n_i}$, $\mathbf{G}_{0i} \in \mathbb{R}^{(n_i+p_i)\times p_i}$, and $\mathbf{C}_{0i} \in \mathbb{R}^{p_i\times(n_i+p_i)}$.

The sensor faults become actuator faults in the augmented system (36). The same algorithm described in Sect. 2.2 can be used to reconstruct the sensor faults. However, considering this method, the following two crucial conditions still need to be satisfied:

(1) $\mathrm{rank}(\mathbf{C}_{0i}\mathbf{G}_{0i}) = \mathrm{rank}(\mathbf{G}_{0i})$;

(2) All the invariant zeros of $(\mathbf{A}_{0i}, \mathbf{G}_{0i}, \mathbf{C}_{0i})$ lie in the left-hand complex plane.

Inspired by [31], these two conditions are verified to be satisfied as follows. For the satisfaction of the condition (1), it is easy to verify since that

$$\operatorname{rank}(\mathbf{C}_{0i}\mathbf{G}_{0i}) = \operatorname{rank}\left([\mathbf{0}\ \ \mathbf{I}_{\mathrm{p}_i}]\begin{bmatrix}\mathbf{0}\\ \mathbf{A}_{\mathrm{f}i}\mathbf{E}_i\end{bmatrix}\right) = \operatorname{rank}\begin{bmatrix}\mathbf{0}\\ \mathbf{A}_{\mathrm{f}i}\mathbf{E}_i\end{bmatrix} = \operatorname{rank}(\mathbf{G}_{0i}).$$

For the condition (2), construct the Rosenbrock system matrix $\mathbf{P}(z)$ based on the matrices $(\mathbf{A}_{0i}, \mathbf{G}_{0i}, \mathbf{C}_{0i})$ as

$$\mathbf{P}(z) = \begin{bmatrix} z\mathbf{I} - \mathbf{A}_{0i} & \mathbf{G}_{0i} \\ \mathbf{C}_{0i} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} z\mathbf{I} - \mathbf{A}_{0i} & \mathbf{0} & \mathbf{0} \\ -\mathbf{A}_{\mathrm{f}i}\mathbf{C}_i & z\mathbf{I} + \mathbf{A}_{\mathrm{f}i} & \mathbf{A}_{\mathrm{f}i}\mathbf{E}_i \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix}.$$

Since the invariant zeros of $(\mathbf{A}_{0i}, \mathbf{G}_{0i}, \mathbf{C}_{0i})$ are the values of z which make the matrix $\mathbf{P}(z)$ loses rank. In this case, the matrix $P(z)$ loses rank if and only if $\tilde{\mathbf{P}}(z)$ loses rank

$$\tilde{\mathbf{P}}(z) = \begin{bmatrix} z\mathbf{I} - \mathbf{A}_i & \mathbf{0} \\ -\mathbf{A}_{\mathrm{f}i}\mathbf{C}_i & \mathbf{A}_{\mathrm{f}i}\mathbf{E}_i \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{\mathrm{f}i} \end{bmatrix} \begin{bmatrix} z\mathbf{I} - \mathbf{A}_i & \mathbf{0} \\ -\mathbf{C}_i & \mathbf{E}_i \end{bmatrix}.$$

Since $\mathbf{A}_{\mathrm{f}i}$ is invertible, the values of z which make $\tilde{\mathbf{P}}(z)$ lose rank are the eigenvalues of $\mathbf{A}_i$. As the observer is designed to estimate the faults, the original system is expected to be asymptotically stable. This means that if one can build a decentralized Walcott–Żak observer based on the original system, then a decentralized Walcott–Żak observer based on the augmented system can also be constructed.

It was stated in [31] that the choice of $\mathbf{A}_{\mathrm{f}i}$ does not affect the estimation signal but affects the observer gains. Thus, the choice of $\mathbf{A}_{\mathrm{f}i}$ affects the accuracy of the sensor fault estimation. As can be seen from the observer structure, the observer for the augmented system estimates the filter outputs. The filtered outputs may not retain a good sensitivity to the faults. If a low-pass filter is used in the presence of a high frequency sensor fault, then the observer cannot estimate the fault properly. This is because the fault effects appearing in the outputs of the filter are not the original sensor faults. For this first-order filter, larger diagonal elements of $\mathbf{A}_{\mathrm{f}i}$ give better estimation accuracy. However, large elements of $\mathbf{A}_{\mathrm{f}i}$ might lead to high observer gains and further lead to significant chattering effect on the observer state estimates. Thus, the value of the diagonal elements should be chosen carefully so that the numerical conditioning of the augmented system (35) can be maintained/established.

Remark 6 Note that since the sensor fault estimation can be converted into an actuator fault problem, the combination of the proposed designs in Sects. 2.2 and 4.2 offers a way to simultaneously reconstruct the actuator and sensor faults.

5 Tutorial Simulation Study

To illustrate the observer and fault estimation algorithms, an example system comprising two interconnected linear subsystems with nonlinear interconnections is given as follows:

$$
\begin{aligned}
\dot{\mathbf{x}}_1 &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x}_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_1 + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{h}_1(\mathbf{x}, t) + \mathbf{G}_1 f_{\mathrm{a}1} \\
y_1 &= [1\ 1]\mathbf{x}_1 + \mathbf{E}_1 f_{\mathrm{s}1} \\
\dot{\mathbf{x}}_2 &= \begin{bmatrix} -3 & 0 & 1 \\ 1 & 2 & 0 \\ 0 & 1 & -1 \end{bmatrix} \mathbf{x}_2 + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u_2 + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{h}_2(\mathbf{x}, t) + \mathbf{G}_2 f_{\mathrm{a}2} \\
\mathbf{y}_2 &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \mathbf{x}_2 + \mathbf{E}_2 f_{\mathrm{s}2}
\end{aligned} \tag{37}
$$

with the nonlinear interactions

$$
\mathbf{h}_1(\mathbf{x}, t) = \alpha_1 \cos(x_{22}) \mathbf{H}_1 \mathbf{x},\ \ \mathbf{h}_2(\mathbf{x}, t) = \alpha_2 \cos(x_{11}) \mathbf{H}_2 \mathbf{x}
$$

where $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2]$, $\alpha_1 = 0.1$, $\alpha_2 = 0.5$ and

$$
\mathbf{H}_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix},\ \mathbf{H}_2 = \frac{1}{\sqrt{13}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.
$$

Assume that the systems are already stabilized by the state feedback controller $u_i = \mathbf{K}_i \mathbf{x}_i$, $i = 1, 2$, whose gains are obtained as follows using pole placement:

$$
\mathbf{K}_1 = [-70\ \ -19],\ \ \mathbf{K}_2 = [-22\ \ -17\ \ -43].
$$

Consider the actuator and sensor faults using the fault estimation method proposed in Sect. 2.2.

5.1 Actuator Faults Estimation

Assume that the system (37) is only affected by actuator faults, i.e. $f_{\mathrm{s}i} = 0, i = 1, 2$. The actuator faults and their distribution matrices are assumed to be

$$
\mathbf{G}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ \mathbf{G}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},\ f_{\mathrm{a}1} = \begin{cases} 0, & 0 \le t \le 5 \\ \sin(2t), & t > 5 \end{cases},\ f_{\mathrm{a}2} = \begin{cases} 0, & 0 \le t \le 8 \\ 1, & t > 8 \end{cases}.
$$

Defining an expansion matrix for the aggregated system leads to

$$
\mathbf{G}_{\mathrm{m}} = \mathrm{diag}(\mathbf{G}_{\mathrm{m}1}, \mathbf{G}_{\mathrm{m}2}) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.
$$

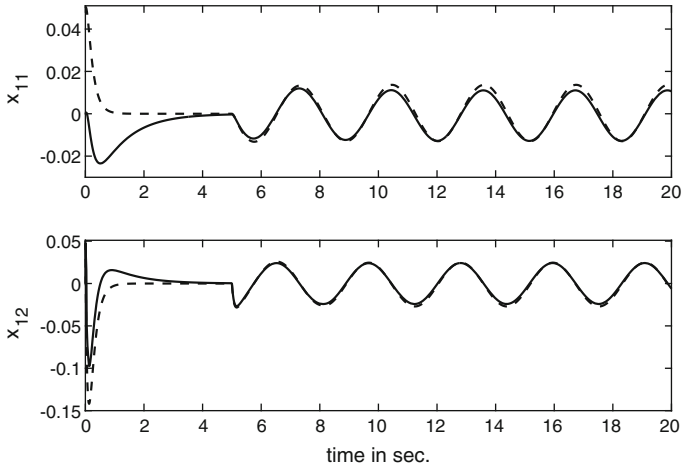


Fig. 2 System states (*dashed lines*) and their estimates (*solid lines*) of the Subsystem 1 with actuator fault

A solution to the LMIs in (29) with $\gamma = 2.1$ gives

$$\mathbf{L}_1 = \begin{bmatrix} -0.2666 \\ 1.7987 \end{bmatrix}, \quad \mathbf{L}_2 = \begin{bmatrix} 0.0956 & 0.0974 \\ 1.5978 & 1.6081 \\ 0.0689 & 0.1334 \end{bmatrix}.$$

The MATLAB simulation results for the system state and actuator faults estimation are shown in Figs. 3, 4, 5 and 6.

Figures 2 and 3 imply that the system states are estimated with good accuracy even in the presence of the actuator faults. Figures 4 and 5 show that the actuator faults are reconstructed with good accuracy.

5.2 Sensor Fault Estimation

Suppose there only exists sensor fault f_{s2} in the second subsystem. Following Sect. 4.2 with a filter matrix $\mathbf{A}_{f2} = 20\mathbf{I}_{p_2}$, the new augmented subsystem is

$$\begin{bmatrix} \dot{\mathbf{x}}_2 \\ \dot{\mathbf{z}}_{f2} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_2 & \mathbf{0} \\ \mathbf{A}_{f2}\mathbf{C}_2 & -\mathbf{A}_{f2} \end{bmatrix} \begin{bmatrix} \mathbf{x}_2 \\ \mathbf{z}_{f2} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_2 \\ \mathbf{0} \end{bmatrix} u_2 + \begin{bmatrix} \mathbf{M}_2 \\ \mathbf{0} \end{bmatrix} \mathbf{h}_2 + \begin{bmatrix} \mathbf{0} \\ \mathbf{A}_{f2}\mathbf{E}_2 \end{bmatrix} f_{s2}$$

$$\mathbf{z}_{f2} = [\mathbf{0}\ \mathbf{I}_{p_2}] \begin{bmatrix} \mathbf{x}_2 \\ \mathbf{z}_{f2} \end{bmatrix}$$

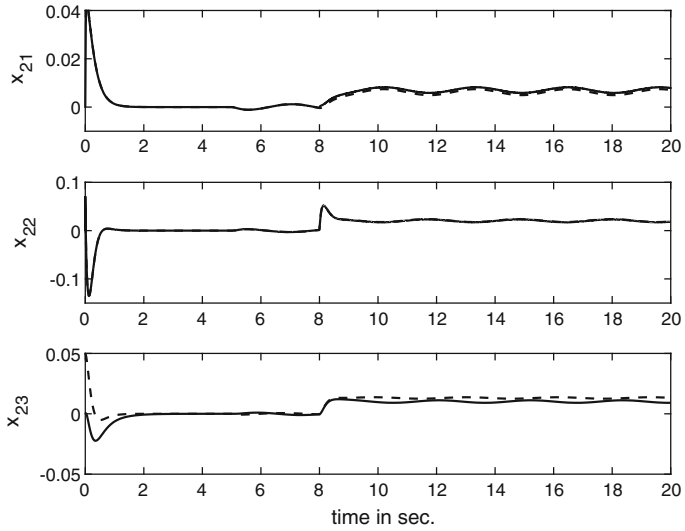


Fig. 3 System states (*dashed lines*) and their estimates (*solid lines*) of the Subsystem 2 with actuator fault

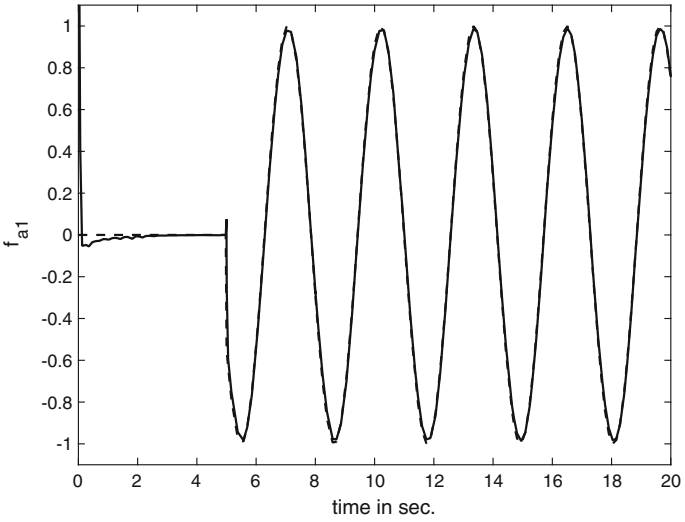


Fig. 4 Actuator fault (*dashed line*) and its estimate (*solid line*) of the Subsystem 1

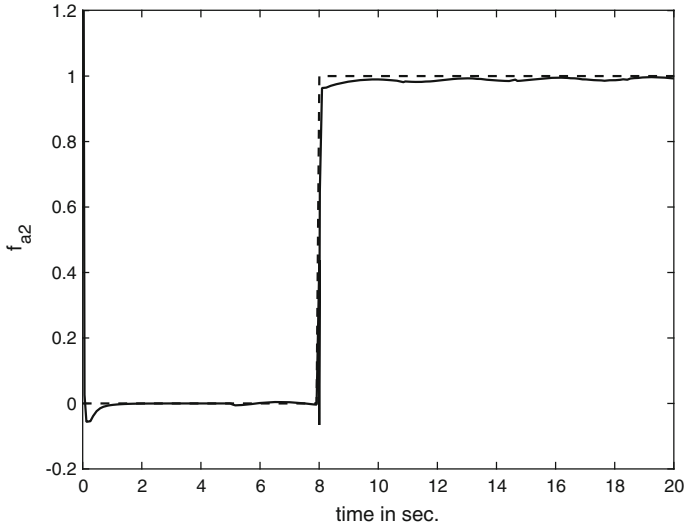


Fig. 5 Actuator fault (*dashed line*) and its estimate (*solid line*) of the Subsystem 2

where $\mathbf{E}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. The sensor fault f_{s2} is defined as

$$f_{s2} = \begin{cases} 0, & 0 \le t \le 2 \\ 0.05(2-t), & 2 \le t \le 7 \\ -0.1\sin(10(t-7)), & t > 7 \end{cases}.$$

The new matrix $\mathbf{A}_{02}$ for the augment system is given by

$$\begin{bmatrix} \mathbf{A}_2 & 0 \\ \mathbf{A}_{f2}\mathbf{C}_2 & -\mathbf{A}_{f2} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 2 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 20 & 0 & -20 & 0 \\ 20 & 20 & 0 & 0 & -20 \end{bmatrix}.$$

A feasible solution of the LMIs in (29) is

$$\mathbf{L}_1 = \begin{bmatrix} -0.1524 \\ 2.3168 \end{bmatrix}, \quad \mathbf{L}_2 = \begin{bmatrix} 0.3230 & 0.2860 \\ 2.0944 & 1.7770 \\ 0.4747 & 0.4023 \\ 2.5910 & 2.2790 \\ 3.1550 & 2.7050 \end{bmatrix}.$$

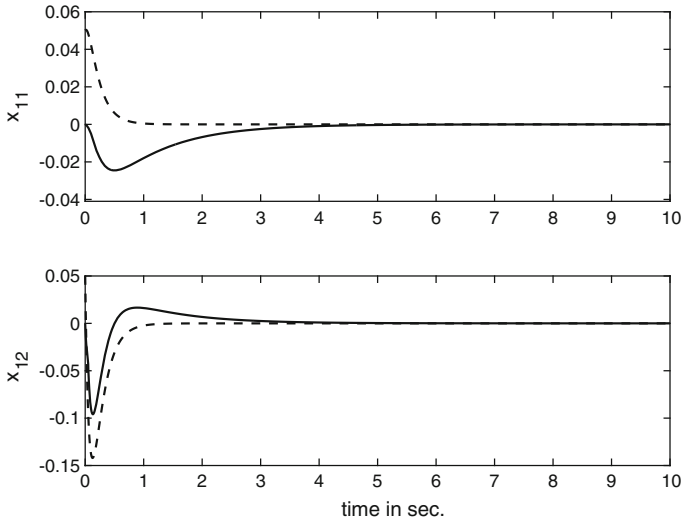


Fig. 6 System states (*dashed lines*) and their estimates (*solid lines*) of Subsystem 1 with sensor fault

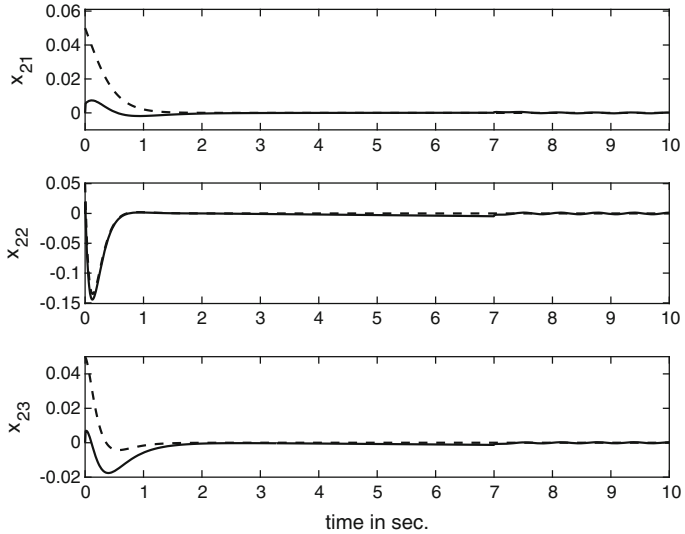


Fig. 7 System states (*dashed lines*) and their estimates (*solid lines*) of Subsystem 2 with sensor fault

The MATLAB simulation for the system states and actuator faults estimation is carried out. Figures 6 and 7 imply that the system states are estimated with good accuracy even in the presence of the actuator faults. Figure 8 shows that the proposed observer in Sect. 4.2 can robustly reconstruct the sensor fault.

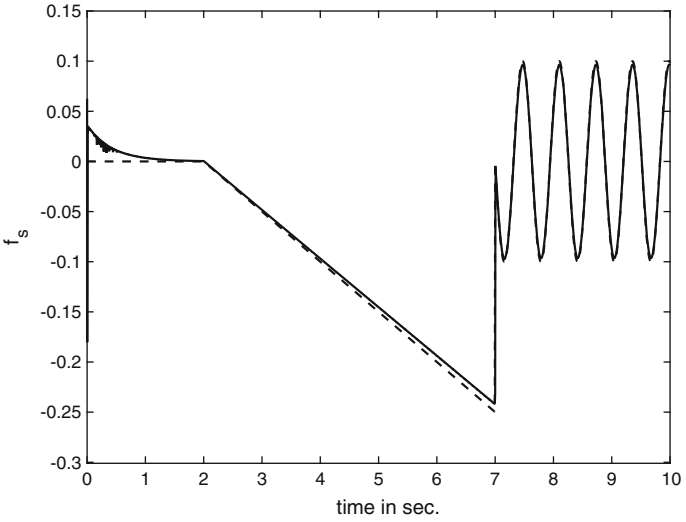


Fig. 8 Sensor fault (*dashed line*) and its estimate (*solid line*)

6 Application to Multimachine Power System

A multimachine power system has been widely used to illustrate the decentralized methods. The interactions of this system are nonlinear which makes it more challenge. Various research has been done based on this system recently, to list a few, [10, 18, 29, 32].

According to [32], an N-machine power system with steam valve control can be described by the interconnection of N subsystems. Denote the state vector of each machine as $\mathbf{x}_i = [\Delta\sigma_i(t)\ \omega_i(t)\ \Delta P_{\mathrm{m}i}(t)\ \Delta X_{\mathrm{e}i}(t)]^\top$, then the dynamics of the ith machine, $i = 1, \ldots, N$, can be represented as

$$\begin{aligned} \dot{\mathbf{x}}_i &= \mathbf{A}_i\mathbf{x}_i + \mathbf{B}_i u_i + \mathbf{h}_i(\mathbf{x}, t) \\ y_i &= \mathbf{C}_i\mathbf{x}_i \end{aligned} \tag{38}$$

where $\mathbf{h}_i(\mathbf{x}, t) = \sum_{j=1, j\neq i}^{N} p_{ij}\mathbf{G}_{ij}g_{ij}(\mathbf{x}_i, \mathbf{x}_j)$ is a nonlinear function vector characterizing the interaction between subsystems.

$$\mathbf{A}_i = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \frac{-D_i}{2H_i} & \frac{\omega_0}{2H_i} & 0 \\ 0 & 0 & \frac{-1}{T_{\mathrm{m}i}} & \frac{K_{\mathrm{m}i}}{T_{\mathrm{m}i}} \\ 0 & \frac{K_{\mathrm{e}i}}{T_{\mathrm{e}i}R_i\omega_0} & 0 & \frac{-1}{T_{\mathrm{e}i}} \end{bmatrix},\ \mathbf{B}_i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{T_{\mathrm{e}i}} \end{bmatrix},\ \mathbf{C}_i^\top = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},\ \mathbf{G}_{ij} = \begin{bmatrix} 0 \\ -\frac{\omega_0 E'_{\mathrm{q}i} E'_{\mathrm{q}j} B_{ij}}{2H_i} \\ 0 \\ 0 \end{bmatrix}$$

and $g_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \sin(\sigma_i - \sigma_j) - \sin(\sigma_{i0} - \sigma_{j0})$.

The physical meaning of the system parameters of (38) are defined by

- u_i is the control vector of the ith subsystem, $u_i = \Delta X_{ei}$;
- y_i is the output vector of the ith subsystem, $y_i = \Delta\sigma_i(t)$;
- σ_i is the rotor angle for the ith machine, in radians;
- ω_i is the relative speed for the ith machine, in radians;
- P_{mi} is the per unit (pu) mechanical power for the ith machine;
- X_{ei} is the pu steam valve aperture for the ith machine;
- p_{ij} is a constant of either 1 or 0 (0 means no connection between the ith and jth machines, while 1 means connection exists);
- H_i is the inertia constant for the ith machine, in seconds;
- D_i is the pu damping coefficient for the ith machine;
- T_{mi} is time constant for the ith machine's turbine, in seconds;
- K_{mi} is the gain of the ith machine's turbine;
- T_{ei} is the time constant for the ith machine's speed governor, in seconds;
- K_{ei} is the gain of the ith machine's speed governor;
- R_i is the pu regulation constant for the ith machine;
- B_{ij} is the pu Nodal susceptance between the ith and jth machines;
- ω_0 is the synchronous machine speed, $\omega_0 = 2\pi f_0$, in radians;
- E'_{qi} is the pu internal transient voltage for the ith machine, which is assumed to be constant;
- σ_{i0}, P_{mi0} and X_{ei0} are the nominal values of σ_i, P_{mi} and X_{ei}, respectively;
- $\Delta\sigma_i = \sigma_i - \sigma_{i0}$ is the deviation of the rotor angle;
- $\Delta P_{mi} = P_{mi} - P_{mi0}$ is the deviation of the mechanical power;
- $\Delta X_{ei} = X_{ei} - X_{ei0}$ is the deviation of the steam valve aperture.

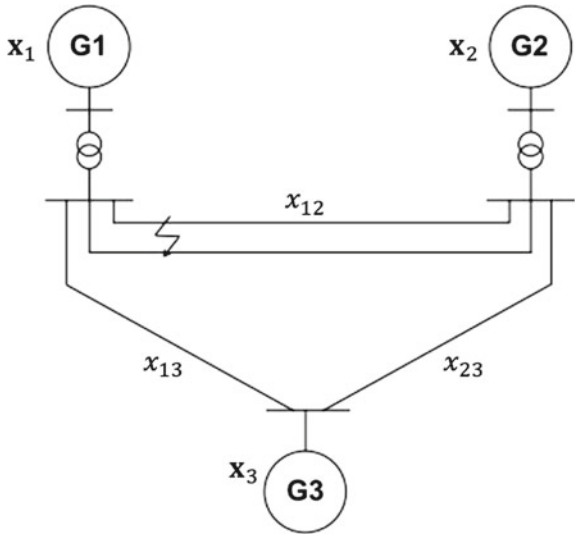


Fig. 9 3-machine power system

In this simulation, a 3-machine system with the structure shown in Fig. 9 is used.

Using the method proposed in [18], one can determine that the interactions are in the form of a quadratic constraint. The parameters of the 3-machine power system (38) can be found in [18, 32].

Assume that the systems have been stabilized by the state feedback controllers $u_i = \mathbf{K}_i \mathbf{x}_i$, $i = 1, 2, 3$, with the following gains obtained via pole placement:

$$\begin{aligned}
\mathbf{K}_1 &= [-224.7592 \quad -21.7583 \quad -59.7609 \quad -2.4899] \\
\mathbf{K}_2 &= [-217.7374 \quad -24.4425 \quad -52.6834 \quad -2.4268] \\
\mathbf{K}_3 &= [-217.7374 \quad -24.4425 \quad -52.6834 \quad -2.4268].
\end{aligned}$$

Suppose that the system (38) is only affected by actuator faults f_{ai}, $i = 1, 2, 3$ with distribution matrices $\mathbf{G}_i$, $i = 1, 2, 3$, respectively. The faults and distribution matrices are defined as

$$\mathbf{G}_1 = \mathbf{G}_2 = \mathbf{G}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad f_{a1} = \begin{cases} 0, & 0 \le t \le 5 \\ \sin(2t), & t > 5 \end{cases}$$

$$f_{a2} = \begin{cases} 0, & 0 \le t \le 3 \\ 0.5, & t > 3 \end{cases}, \quad f_{a3} = \begin{cases} 0, & 0 \le t \le 6 \\ 0.1, & 6 < t \le 8 \\ 1.5, & 8 < t \le 16 \\ 0.5, & t > 16 \end{cases}.$$

A feasible solution of the LMIs in (29) is $\gamma = 0.5$ with

$$\mathbf{L}_1 = \begin{bmatrix} 0.6251 \\ 98.9215 \\ -7.1268 \\ 3.0020 \end{bmatrix}, \quad \mathbf{L}_2 = \begin{bmatrix} 0.6251 \\ 86.9479 \\ -7.2829 \\ 1.1291 \end{bmatrix}, \quad \mathbf{L}_3 = \begin{bmatrix} 0.6251 \\ 86.9479 \\ -7.2829 \\ 1.1291 \end{bmatrix}.$$

The MATLAB simulation for the system state and actuator faults estimation is carried out. Figures 10, 11, 12, 13, 14 and 15 imply that the system states are estimated with good accuracy even in the presence of the actuator faults. Figure 16 shows that the actuator faults are reconstructed with good accuracy. It should also be noted that the estimations of the actuator faults affect each other.

7 Conclusion

The purpose of this chapter is to develop a suitable framework for observer design for decentralized systems based on the SMO. Both the Walcott–Żak and canonical form SMOs are reviewed and to avoid the complicated transformations the Walcott–Żak

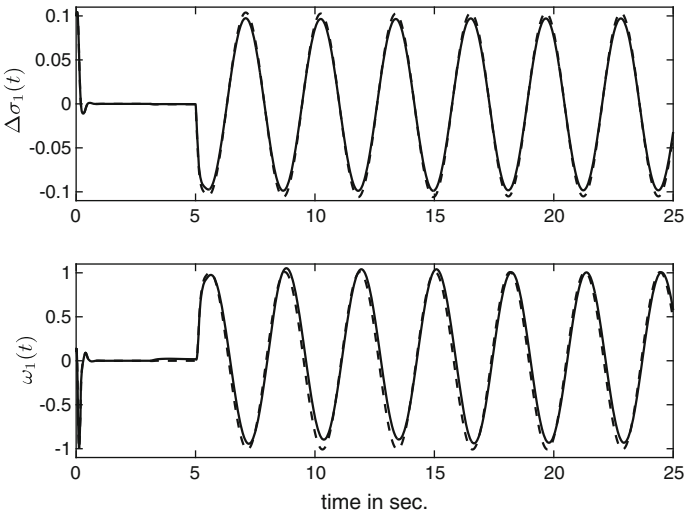


Fig. 10 System states (*dashed lines*) and their estimates (*solid lines*) of the Subsystem 1

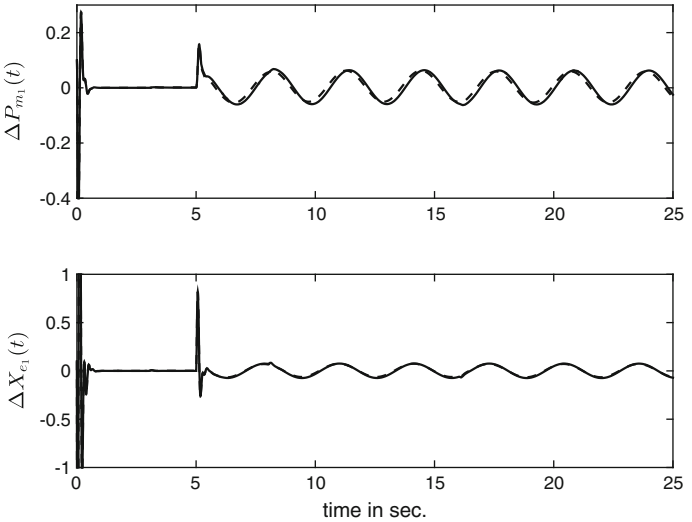


Fig. 11 System states (*dashed lines*) and their estimates (*solid lines*) of the Subsystem 1

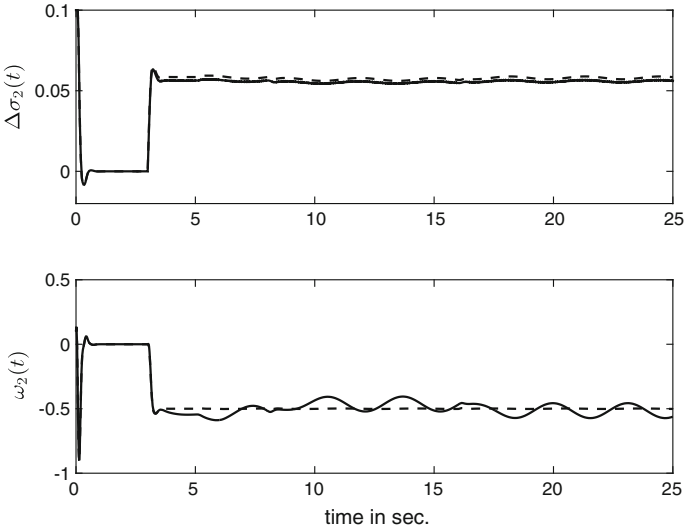


Fig. 12 System states (*dashed lines*) and their estimates (*solid lines*) of the Subsystem 2

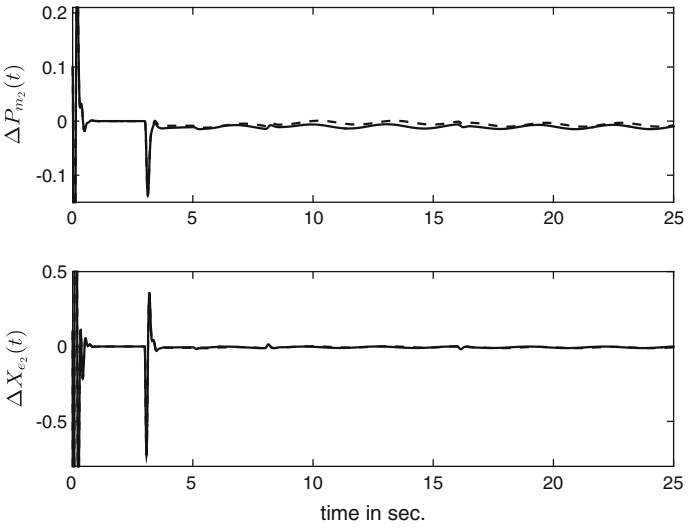


Fig. 13 System states (*dashed lines*) and their estimates (*solid lines*) of the Subsystem 2

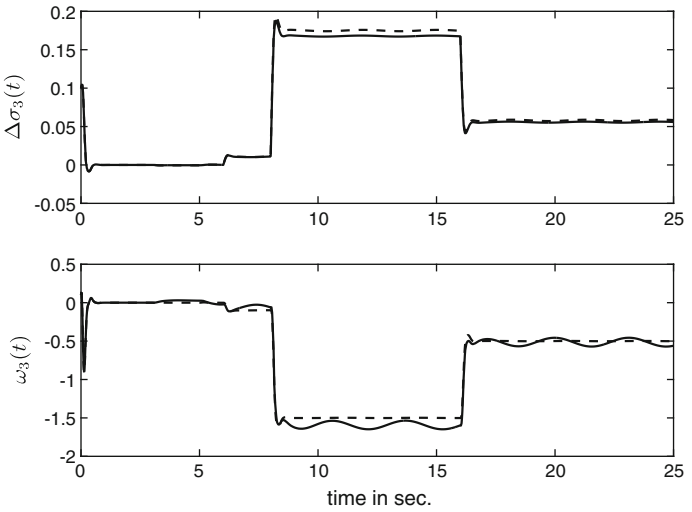


Fig. 14 System states (*dashed lines*) and their estimates (*solid lines*) of the Subsystem 3

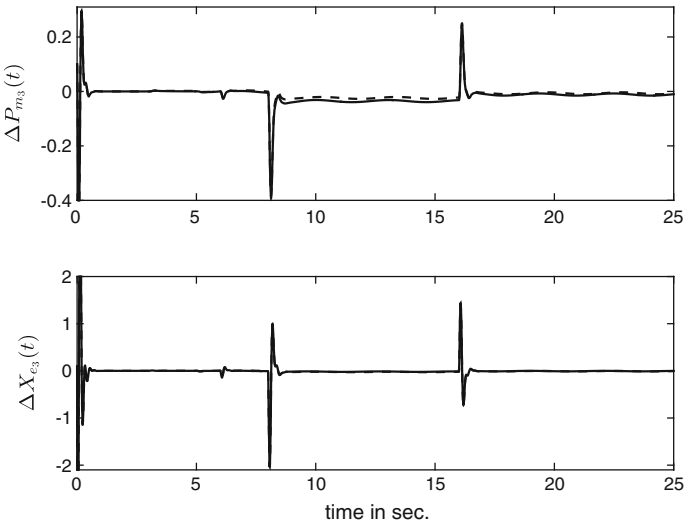


Fig. 15 System states (*dashed lines*) and their estimates (*solid lines*) of the Subsystem 3

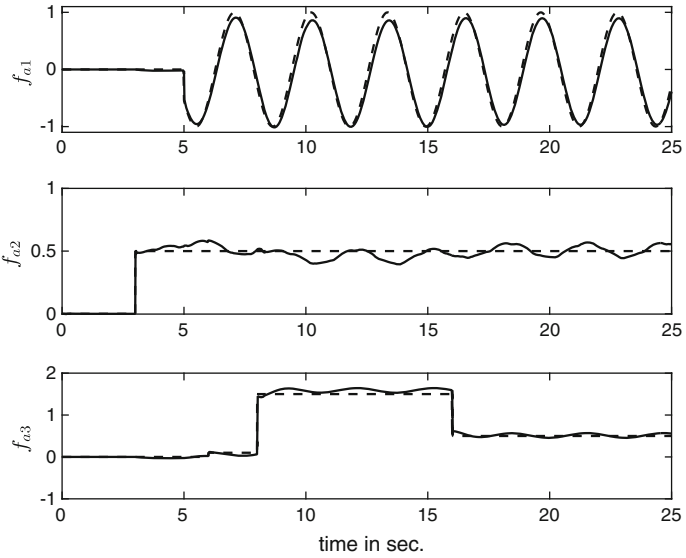


Fig. 16 Actuator faults (*dashed lines*) and their estimates (*solid lines*)

SMO has been selected as the most appropriate state and fault estimation approach for decentralized systems. The fact that state transformations are not required is a big advantage for a complex decentralized system with a potentially large number of interconnected subsystems. Furthermore, to simplify the design procedure, an LMI strategy is used to develop the Lyapunov formulation. Furthermore, as a further contribution the Walcott–Żak SMO is modified to avoid the inaccurate output estimation when the sliding surface has been reached.

According to the SMO output injection property, the feedback switching term can approximate the "matched" unwanted signal in the system. The so-called "equivalent output injection" can then be used to estimate the actuator fault. Although the canonical form observer has been researched extensively in the literature for the fault estimation problem, a main contribution of this chapter is a discussion of the suitability of the Walcott–Żak observer for state/fault estimation in decentralized systems. For the sensor fault estimation problem and to keep to the standard form of SMO (based on actuator faults), a filter is used to construct a suitable augmented system. The sensor fault estimation problem is thus transferred into an equivalent actuator fault estimation problem.

Both actuator and sensor fault estimation algorithms are applied in a tutorial example to illustrate the effectiveness of the proposed estimation strategy. An example of a three-machine system is then used to illustrate the feasibility of using this approach on a physical system problem.

References

1. Bakule L, Rodellar J (1996) Decentralised control design of uncertain nominally linear symmetric composite systems. IEE Proc Control Theory Appl 143:530–536
2. Bakule L, Rossel JM (2008) Overlapping controllers for uncertain delay continuous-time systems. Kybernetika 44(1):17–34
3. Benigni A, D'Antona G, Ghisla U, Monti A, Ponci F (2010) A decentralized observer for ship power system applications: implementation and experimental validation. IEEE Trans Instrum Meas 59(2):440–449
4. Choi HH, Ro KS (2005) LMI-based sliding-mode observer design method. IEE Proc Control Theory Appl 152(1):113–115
5. Chung WH, Speyer JL, Chen RH (2001) A decentralized fault detection filter. J Dyn Syst Meas Control 123(2):237–247
6. Corless M, Tu J (1998) State and input estimation for a class of uncertain systems. Automatica 34(6):757–764
7. Edwards C, Menon PP (2008) State reconstruction in complex networks using sliding mode observers. In: 47th IEEE conference on decision and control, IEEE, Cancun, Mexico, pp 2832–2837
8. Edwards C, Spurgeon SK (1998) Sliding mode control: theory and applications. CRC Press, New York
9. Edwards C, Spurgeon SK, Patton RJ (2000) Sliding mode observers for fault detection and isolation. Automatica 36(4):541–553
10. Guo Y, Hill DJ, Wang Y (2000) Nonlinear decentralized control of large-scale power systems. Automatica 36(9):1275–1289
11. Hassan MF, Sultan MA, Attia MS (1992) Fault detection in large-scale stochastic dynamic systems. In: IEE proceedings control theory and applications, IET, vol 139, pp 119–124
12. Huang Z, Patton RJ (2012) An adaptive sliding mode approach to decentralized control of uncertain systems. In: 2012 UKACC international conference on control, IEEE, Cardiff, UK, pp 70–75
13. Huang Z, Patton RJ (2012) Decentralized control of uncertain systems via adaptive sliding and overlapping decomposition. In: Proceedings of the 7th IFAC symposium on robust control design, IFAC, Aalborg, Denmark, vol 7, pp 784–789
14. Huang Z, Patton RJ (2015) Output feedback sliding mode ftc for a class of nonlinear interconnected systems. IFAC-PapersOnLine 48(21):1140–1145
15. Hui S, Żak SH (2005) Observer design for systems with unknown inputs. Int J Appl Math Comput Sci 15(4):431–446
16. Ikeda M (1989) Decentralized control of large scale systems. Three Decad Math Syst Theory 135:219–242
17. Kalsi K, Lian J, Żak SH (2010) Decentralized dynamic output feedback control of nonlinear interconnected systems. IEEE Trans Autom Control 55(8):1964–1970
18. Kalsi K, Lian J et al (2009) Decentralized control of multimachine power systems. In: Proceedings of the American control conference, IEEE, Missouri, United States, pp 2122–2127
19. Pagilla PR, Zhu Y (2004) A decentralized output feedback controller for a class of large-scale interconnected nonlinear systems. In: Proceedings of the American control conference, IEEE, Boston, United States, vol 4, pp 3711–3716
20. Patton RJ, Klinkhieo S (2009) Actuator fault estimation and compensation based on an augmented state observer approach. In: Proceedings of the 48th IEEE conference on decision and control, IEEE, pp 8482–8487
21. Patton RJ, Putra D, Klinkhieo S (2010) Friction compensation as a fault-tolerant control problem. Int J Syst Sci 41(8):987–1001
22. Sandell NR Jr, Varaiya P, Athans M, Safonov MG (1978) Survey of decentralized control methods for large scale systems. IEEE Trans Autom Control 23(2):108–128

23. Shafai B, Ghadami R, Saif M (2011) Robust decentralized PI observer for linear interconnected systems. In: 2011 IEEE international symposium on computer-aided control system design, IEEE, Denver, CO, United States, pp 650–655
24. Shankar S, Darbha S, Datta A (2002) Design of a decentralized detection of interacting LTI systems. Math Probl Eng 8(3):233–248
25. Šiljak DD (1991) Decentralized control of complex systems. Academic Press, Boston
26. Šiljak DD (1996) Decentralized control and computations: status and prospects. Annu Rev Control 20:131–141
27. Šiljak DD, Vukcevic MB (1976) Decentralization, stabilization, and estimation of large-scale linear systems. IEEE Trans Autom Control 21(3):363–366
28. Šiljak DD, Zečević AI (2005) Control of large-scale systems: beyond decentralized feedback. Annu Rev Control 29(2):169–179
29. Šiljak DD, Stipanovic DM, Zecevic AI et al (2002) Robust decentralized turbine/governor control using linear matrix inequalities. IEEE Trans Power Syst 17(3):715–722
30. Tan CP, Edwards C (2001) An LMI approach for designing sliding mode observers. Int J Control 74(16):1559–1568
31. Tan CP, Edwards C (2002) Sliding mode observers for detection and reconstruction of sensor faults. Automatica 38(10):1815–1821
32. Tlili AS, Braiek NB (2009) Decentralized observer based guaranteed cost control for nonlinear interconnected systems. Int J Control Autom 2(2):19–34
33. Utkin VI (1992) Sliding modes in control and optimization. Springer, Berlin
34. Walcott BL, Żak SH (1987) State observation of nonlinear uncertain dynamical systems. IEEE Trans Autom Control 32(2):166–170
35. Xiang J, Su H, Chu J (2005) On the design of Walcott-żak sliding mode observer. In: Proceedings of the American control conference, IEEE, Portland, OR, United States, pp 2451–2456
36. Yan X, Edwards C (2008) Robust decentralized actuator fault detection and estimation for large-scale systems using a sliding mode observer. Int J control 81(4):591–606
37. Zinober AS (1990) Deterministic control of uncertain systems. IET, London

Fault Diagnosis of Nonlinear Differential-Algebraic Systems Using Hybrid Estimation

Dirk Weidemann and Ilja Alkov

Abstract Modern technical systems often contain components capable for extensive autonomous actions. Thus, an integrated system supervision is essential in enabling an adequate reaction for compensation of unpredictable substantial variations. This is addressed by fault detection, isolation and identification techniques discussed in this article. Therefore, an overview is given about modelling of systems subject to faults, continuous state estimation utilizing an unscented Kalman filter and hybrid state estimation by the interacting multiple model approach. These methods are generalized for application to nonlinear differential-algebraic equations, i.e. DAE systems. DAE systems arise in such fields as discretization of partial differential equations or optimization problems. However, the appearance of DAE systems most often results from an object-oriented modelling (OOM) approach. Since OOM is probably the most relevant approach for modelling complex systems, the generalization and adaptation of supervision methods to DAE is the principal subject of this contribution. Finally, the proposed fault identification approach is applied to a hydraulic system, and the related results are discussed in detail.

1 Introduction

As modern technical systems become more complex, often the likelihood that faults occur increases. An appropriate way to compensate for effects of faults is to use redundant hardware. For cost reasons, however, the deployment of redundant systems is uncommon if the faults do not lead to life-threatening situations. Rather, the aim is to compensate for the effect of faults by diagnosing them and using an appropriate fault-tolerant control strategy. Thus, model-based fault diagnosis and (active)

D. Weidemann (✉) · I. Alkov
Institute of System Dynamics and Mechatronics, University of Applied Sciences Bielefeld, Interaktion 1, 33619 Bielefeld, Germany
e-mail: dirk.weidemann@fh-bielefeld.de

I. Alkov
e-mail: ilja.alkov@uni-bielefeld.de

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*, Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_10

fault-tolerant control play an increasingly important role for complex systems and are active research areas.

According to [6], active fault-tolerant control approaches usually require not only to locate the fault, but also to estimate its type and magnitude, which is known as fault identification. A significant amount of model-based approaches are proposed for fault detection and fault isolation (cf. [6, 10, 12, 19]). However, there are fewer approaches for fault identification. One of the first fault identification approaches was proposed in [18] for linear systems and is based on the assumption that faults lead to a change of the system parameters. Therefore, the faults can be identified using online parameter estimation approaches. In contrast, in [29] a sliding mode observer is proposed to identify actuator and sensor faults. This approach is also only applicable to linear systems. Most systems, however, are characterized by a significant nonlinear behaviour. An extension of the approach proposed in [29] to a specific class of nonlinear systems is given in [9]. Though, the latter approach is only applicable to identify actuator faults.

An alternative that offers the possibility to identify sensor, actuator and system faults is a combined state and fault estimation in which the faults are modelled by augmented state variables. A typical example of this kind of fault modelling is an abrupt fault that is modelled as a constant parameter using a single augmented state variable. Combined state and parameter estimation with a two-stage adaptive Kalman filter is proposed in [31] as a basis for a fault-tolerant control scheme. If the system is only subject to one fault, it is sufficient to use only a single model. However, complex systems typically are subject to several faults such that not only a model for the fault-free system but also an additional model for each fault has to be used. In order to find which of these models fit best the actual system behaviour, a mode estimation scheme is appropriate. As indicated in [17], mode estimation can be done in the sense of hybrid estimation.

Common hybrid estimation schemes assume that a model is given in state-space form. In the context of complex systems, object-oriented modelling (OOM) has gained increasing attention within science and industry, since it reduces significantly the effort needed to model technical systems of high complexity. This advantage is obtained by hierarchical modelling, inheritance and the ability to interconnect component models. As a consequence of connecting component models, the resulting mathematical model describing the behaviour of complex systems consists not only of ordinary differential equations but also of algebraic equations. Such sets of equations are called differential-algebraic equations (DAEs), also known as descriptor, implicit, or singular systems. Note, however, that even though DAE systems are the typical model class in the context of OOM, these systems also arise in such fields as discretization of partial differential equations, optimization problems, or approximation the fast dynamics of stiff ODE systems by algebraic equations.

In this article, we propose a hybrid estimation scheme for nonlinear DAE systems for fault identification purposes. Hence, the estimates consist of discrete- and continuous-valued states. A hybrid modelling approach is given in which the discrete-valued states represent the modes of the system, such that the discrete dynamics include phenomena such as jumps in the states or switching of the vector field. The

continuous-valued states represent the corresponding continuous dynamics of each mode described by a nonlinear DAE system.

2 Differential-Algebraic Systems

Even though modelling of complex systems generally becomes easier if an object-oriented modelling approach is used, this modelling approach typically leads to a set of differential-algebraic equations. The class of DAE systems contains systems given in the state-space form as a special case, but in general the properties of DAE systems are more complex. Therefore, analysis of DAE systems and also synthesis of controllers, observers and filters for DAE systems are more challenging than for systems given in state-space form. This section discusses the difference between DAE and ODE systems, giving a formal definition of a DAE system.

The most general form of a DAE system is a system of fully implicit first-order DAEs

$$\mathbf{f}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}, t) = \mathbf{0} \tag{1}$$

where

$$\det\left(\frac{\partial \mathbf{f}}{\partial \dot{\mathbf{x}}}\right) = 0 \tag{2}$$

and $\mathbf{f} : \mathbb{R}^{n_\mathrm{x}} \times \mathbb{R}^{n_\mathrm{x}} \times \mathbb{R}^{n_\mathrm{u}} \times \mathbb{R} \to \mathbb{R}^{n_\mathrm{x}}$ is a sufficiently differentiable vector-valued function, implicitly dependent on the generalized states $\mathbf{x} \in \mathbb{R}^{n_\mathrm{x}}$ of the DAE system, the corresponding time derivative of the generalized states $\dot{\mathbf{x}} \in \mathbb{R}^{n_\mathrm{x}}$, time-dependent input variables $\mathbf{u} \in \mathbb{R}^{n_\mathrm{u}}$ and the time $t \in \mathbb{R}$. The dimensions of the generalized states and the input variables are given by n_x and n_u, respectively.

If (2) holds, according to the implicit function theorem, the fully implicit system (1), which contains differential and algebraic equations, cannot be solved for $\dot{\mathbf{x}}$ algebraically. Furthermore, the condition (2) may be satisfied only locally but not globally, and the structural properties of the system may change significantly. For the sake of simplicity, only DAE systems with invariable structural properties are considered in the following.

2.1 Differentiation Index of DAE Systems

The core of the hybrid estimation presented in Sect. 5 consists of Kalman filters for nonlinear DAE systems. Most Kalman filter approaches to nonlinear DAE systems are applicable only to index-1 systems, whereas few approaches for higher index problems exist (cf. [2]). Therefore, the concept of index is explained briefly. Several concepts for the index of DAE systems are established in the literature. A detailed overview of the concepts of differentiation, perturbation, geometric, strangeness,

structural and tractability indices is given in [25]. However, the differentiation index is the most widely used concept.

The differentiation index (cf. [8]) of a DAE system (1) is the smallest number of differentiations $\mu \in \mathbb{N}$ required to solve

$$\begin{bmatrix} \mathbf{f}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}, t) \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{f}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}, t) \\ \vdots \\ \frac{\mathrm{d}^{\mu}}{\mathrm{d}t^{\mu}}\mathbf{f}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}, t) \end{bmatrix} = \mathbf{0} \tag{3}$$

algebraically for $\dot{\mathbf{x}}$

$$\dot{\mathbf{x}} = \phi\left(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, \ldots, \mathbf{u}^{(\mu)}, t\right) \tag{4}$$

as a function of the generalized states, the input variables, time derivatives of the input variables and the time. The equation ϕ is referred to as the underlying ODE corresponding to the original DAE system and may be an implicit equation algebraically solvable for $\dot{\mathbf{x}}$. A frequently indicated disadvantage of the differentiation index concept is the requirement of unique solvability making it unsuitable for over- and underdetermined systems. In this contribution, it is assumed that the system is uniquely solvable, which holds for most engineering problems.

2.2 Solution of DAE Systems

Solving nonlinear differential-algebraic systems is still a research subject, however, several approaches are established in the literature. A number of direct integration methods are applicable to specific classes of DAE systems (cf. [7, 11]). These methods often require a low differentiation index and/or a specific structure of the DAE system. Probably, the most widely used solution method is a combination of an index reduction technique (e.g. [24, 27]) and a direct numerical integration of an index-reduced DAE formulation. This approach is implemented in popular commercial (e.g. Dymola or MapleSim™) and open source (e.g. OpenModelica) simulation environments. Furthermore, solver implementations based on an equivalent strangeness-free formulation of the DAE (cf. [22]), on Taylor series expansion (cf. [26]) and on nonlinear programming (cf. [15]) are relevant for scientific as well as industrial applications. The application of hybrid state estimation to DAE systems requires a description of a solution concept, while computational aspects and the algorithmic implementation of solution methods are less relevant. Therefore, a reformulation concept for a general implicit nonlinear DAE system to the underlying ODE and algebraic constraints, i.e. the restrictive manifold is given subsequently.

Assume (1) is regular, i.e. uniquely solvable. The Jacobian $\partial \mathbf{f}/\partial \dot{\mathbf{x}}$ is assumed to be singular, and therefore two permutation matrices $\mathbf{P}_{\mathrm{x},0}$ and $\mathbf{P}_{\mathrm{f},0}$ exist such that

$$\mathbf{P}_{\mathrm{x},0}\mathbf{x} = \left[\mathbf{x}_0^{\mathrm{T}}, \mathbf{z}_0^{\mathrm{T}}\right]^{\mathrm{T}}, \quad \mathbf{P}_{\mathrm{f},0}\mathbf{f} = \left[\mathbf{f}_0^{\mathrm{T}}, \mathbf{g}_0^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{5}$$

$$\mathbf{P}_{\mathrm{f},0}\frac{\partial \mathbf{f}}{\partial \dot{\mathbf{x}}}\mathbf{P}_{\mathrm{x},0}^{\mathrm{T}} = \begin{bmatrix} \frac{\partial \mathbf{f}_0}{\partial \dot{\mathbf{x}}_0} & \frac{\partial \mathbf{f}_0}{\partial \dot{\mathbf{z}}_0} \\ \frac{\partial \mathbf{g}_0}{\partial \dot{\mathbf{x}}_0} & \frac{\partial \mathbf{g}_0}{\partial \dot{\mathbf{z}}_0} \end{bmatrix} \wedge \det\left(\frac{\partial \mathbf{f}_0}{\partial \dot{\mathbf{x}}_0}\right) \neq 0 \tag{6}$$

hold and the DAE system can be decomposed to

$$\mathbf{f}_0(\mathbf{x}_0, \dot{\mathbf{x}}_0, \mathbf{z}_0, \dot{\mathbf{z}}_0, \mathbf{u}, t) = \mathbf{0}, \tag{7}$$

$$\mathbf{g}_0(\mathbf{x}_0, \dot{\mathbf{x}}_0, \mathbf{z}_0, \dot{\mathbf{z}}_0, \mathbf{u}, t) = \mathbf{0}. \tag{8}$$

Since $\partial \mathbf{f}_0/\partial \dot{\mathbf{x}}_0$ is regular, $\mathbf{f}_0$ can be solved locally for

$$\dot{\mathbf{x}}_0 = \phi_0(\mathbf{x}_0, \mathbf{z}_0, \dot{\mathbf{z}}_0, \mathbf{u}, t). \tag{9}$$

Furthermore, $\dot{\mathbf{x}}_0$ is substituted in $\mathbf{g}_0$ resulting in

$$\mathbf{g}_0(\mathbf{x}_0, \mathbf{z}_0, \mathbf{u}, t) = \mathbf{0}, \tag{10}$$

which does not depend on $\dot{\mathbf{z}}_0$, otherwise the previous decomposition would result in additional components in $\mathbf{x}_0$.

Thereafter, $\mathbf{g}_0$ is differentiated with respect to time yielding

$$\frac{\partial \mathbf{g}_0}{\partial \mathbf{x}_0}\dot{\mathbf{x}}_0 + \frac{\partial \mathbf{g}_0}{\partial \mathbf{z}_0}\dot{\mathbf{z}}_0 + \frac{\partial \mathbf{g}_0}{\partial \mathbf{u}}\dot{\mathbf{u}} + \frac{\partial \mathbf{g}_0}{\partial t} = \mathbf{0}. \tag{11}$$

The Eq. (11) is an implicit DAE if $\partial(\mathrm{d}\,\mathbf{g}_0/\,\mathrm{d}\,t)/\partial \dot{\mathbf{z}}_0$ is singular.

Again, proceeding with an equivalent decomposition, two permutation matrices $\mathbf{P}_{\mathrm{x},1}$ and $\mathbf{P}_{\mathrm{f},1}$ exist such that

$$\mathbf{P}_{\mathrm{x},1}\mathbf{z}_0 = \left[\mathbf{x}_1^{\mathrm{T}}, \mathbf{z}_1^{\mathrm{T}}\right]^{\mathrm{T}}, \quad \mathbf{P}_{\mathrm{f},1}\frac{\mathrm{d}\,\mathbf{g}_0}{\mathrm{d}\,t} = \left[\mathbf{f}_1^{\mathrm{T}}, \mathbf{g}_1^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{12}$$

$$\mathbf{P}_{\mathrm{f},1}\frac{\partial}{\partial \dot{\mathbf{z}}_0}\frac{\mathrm{d}\,\mathbf{g}_0}{\mathrm{d}\,t}\mathbf{P}_{\mathrm{x},1}^{\mathrm{T}} = \begin{bmatrix} \frac{\partial \mathbf{f}_1}{\partial \dot{\mathbf{x}}_1} & \frac{\partial \mathbf{f}_1}{\partial \dot{\mathbf{z}}_1} \\ \frac{\partial \mathbf{g}_1}{\partial \dot{\mathbf{x}}_1} & \frac{\partial \mathbf{g}_1}{\partial \dot{\mathbf{z}}_1} \end{bmatrix} \wedge \det\left(\frac{\partial \mathbf{f}_1}{\partial \dot{\mathbf{x}}_1}\right) \neq 0 \tag{13}$$

hold with

$$\mathbf{f}_1(\mathbf{x}_0, \mathbf{x}_1, \dot{\mathbf{x}}_0, \dot{\mathbf{x}}_1, \mathbf{z}_1, \dot{\mathbf{z}}_1, \mathbf{u}, \dot{\mathbf{u}}, t) = \mathbf{0}, \tag{14}$$

$$\mathbf{g}_1(\mathbf{x}_0, \mathbf{x}_1, \dot{\mathbf{x}}_0, \dot{\mathbf{x}}_1, \mathbf{z}_1, \dot{\mathbf{z}}_1, \mathbf{u}, \dot{\mathbf{u}}, t) = \mathbf{0}. \tag{15}$$

Since $\partial \mathbf{f}_1 / \partial \dot{\mathbf{x}}_1$ is regular, there exists a local solution

$$\dot{\mathbf{x}}_1 = \phi_1(\mathbf{x}_0, \mathbf{x}_1, \dot{\mathbf{x}}_0, \mathbf{z}_1, \dot{\mathbf{z}}_1, \mathbf{u}, \dot{\mathbf{u}}, t). \tag{16}$$

Moreover, $\dot{\mathbf{x}}_0$ and $\dot{\mathbf{x}}_1$ are substituted in $\mathbf{g}_1$ yielding

$$\mathbf{g}_1(\mathbf{x}_0, \mathbf{x}_1, \mathbf{z}_1, \mathbf{u}, \dot{\mathbf{u}}, t) = \mathbf{0}, \tag{17}$$

which is a function that does not depend on $\dot{\mathbf{z}}_1$, otherwise the previous decomposition would result in additional components in $\mathbf{x}_1$.

Repeating this procedure with decomposition and differentiation μ times, where μ is the differentiation index of the DAE, results in

$$\mathbf{f}_m\left(\mathbf{x}_0, \ldots, \mathbf{x}_m, \dot{\mathbf{x}}_0, \ldots, \dot{\mathbf{x}}_m, \mathbf{z}_m, \dot{\mathbf{z}}_m, \mathbf{u}, \ldots, \mathbf{u}^{(m)}, t\right) = \mathbf{0} \wedge \det\left(\frac{\partial \mathbf{f}_m}{\partial \dot{\mathbf{x}}_m}\right) \neq 0, \tag{18}$$

$$\dot{\mathbf{x}}_m = \phi_m\left(\mathbf{x}_0, \ldots, \mathbf{x}_m, \dot{\mathbf{x}}_0, \ldots, \dot{\mathbf{x}}_{m-1}, \mathbf{z}_m, \dot{\mathbf{z}}_m, \mathbf{u}, \ldots, \mathbf{u}^{(m)}, t\right), \tag{19}$$

$$\mathbf{g}_m\left(\mathbf{x}_0, \ldots, \mathbf{x}_m, \mathbf{z}_m, \mathbf{u}, \ldots, \mathbf{u}^{(m)}, t\right) = \mathbf{0}, \tag{20}$$

$$\dim\left(\mathbf{g}_\mu\right) = 0 \wedge \dim\left(\mathbf{z}_\mu\right) = 0 \tag{21}$$

for all $m \in \{0, 1, \ldots, \mu\}$. Furthermore, repeated substitution

$$\mathbf{z}_{m-1} = \mathbf{P}_{\mathrm{x},m}^{\mathrm{T}}\left[\mathbf{x}_m^{\mathrm{T}}, \mathbf{z}_m^{\mathrm{T}}\right]^{\mathrm{T}} \quad \forall\, m \in \{1, 2, \ldots, \mu\}, \tag{22}$$

$$\dot{\mathbf{x}}_m = \phi_m \quad \forall\, m \in \{0, 1, \ldots, \mu\} \tag{23}$$

in (18)–(20) and reconstruction

$$\mathbf{x} = \prod_{m=0}^{\mu} \begin{bmatrix} \mathbf{I}_{n_\mathrm{x} - \dim(\mathbf{x}_m) - \dim(\mathbf{z}_m)} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{\mathrm{x},m}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_\mu \end{bmatrix} \tag{24}$$

yields

$$\dot{\mathbf{x}} = \phi\left(\mathbf{x}, \mathbf{u}_\mu, t\right), \tag{25}$$

$$\mathbf{g}\left(\mathbf{x}, \mathbf{u}_\mu, t\right) = \mathbf{0}, \tag{26}$$

$$\mathbf{u}_\mu(t) = \{\mathbf{u}, \ldots, \mathbf{u}^{(\mu)}\} \tag{27}$$

with

$$\phi = \prod_{m=0}^{\mu} \begin{bmatrix} \mathbf{I}_{n_\mathrm{x} - \dim(\mathbf{x}_m) - \dim(\mathbf{z}_m)} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{\mathrm{x},m}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \phi_0 \\ \vdots \\ \phi_\mu \end{bmatrix}, \tag{28}$$

$$\mathbf{g} = \left[\mathbf{g}_0^{\mathrm{T}}, \ldots, \mathbf{g}_\mu^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{29}$$

The presented reformulation determines the underlying ODE (25) and the restrictive manifold (26) corresponding to the DAE system given by (1). Since the restrictive manifold imposes an algebraic constraint on the generalized states, the system obtained is obviously overdetermined. Consequently, only a subset of the generalized states can be initialized arbitrarily. According to the representation of the DAE system as an underlying ODE and the associated restrictive manifold, the generalized states $\mathbf{x}$ can be decomposed in the differential states $\mathbf{x}_\mathrm{d} \in \mathbb{R}^{n_{\mathrm{x,d}}}$ and the algebraic states $\mathbf{x}_\mathrm{a} \in \mathbb{R}^{n_{\mathrm{x,a}}}$ with corresponding dimensions $n_{\mathrm{x,d}} \in \mathbb{N}$ and $n_{\mathrm{x,a}} \in \mathbb{N}$. Hence, a common form used for direct numerical integration is a semi-explicit index-1 DAE given as

$$\dot{\mathbf{x}}_\mathrm{d} = \boldsymbol{\phi}_\mathrm{d}\left(\mathbf{x}_\mathrm{d}, \mathbf{x}_\mathrm{a}, \mathbf{u}_\mu, t\right), \tag{30}$$

$$\mathbf{g}_\mathrm{a}\left(\mathbf{x}_\mathrm{d}, \mathbf{x}_\mathrm{a}, \mathbf{u}_\mu, t\right) = \mathbf{0} \ \wedge \ \det\left(\frac{\partial \mathbf{g}_\mathrm{a}}{\partial \mathbf{x}_\mathrm{a}}\right) \neq 0. \tag{31}$$

This decomposition is not unique and is addressed by the index reduction procedure (e.g. [13, 24, 27]). In this formulation the differential states can be initialized arbitrarily, whereas the consistent algebraic states are calculated in dependence of the constraints. Thus, the corresponding state-transition function can be given directly as

$$\mathbf{x}_{\mathrm{d},k} = \mathbf{x}_{\mathrm{d},k-1} + \int_{t_{k-1}}^{t_k} \boldsymbol{\phi}_\mathrm{d}\left(\mathbf{x}_\mathrm{d}, \mathbf{x}_\mathrm{a}, \mathbf{u}_\mu, t\right) \mathrm{d}\, t = \boldsymbol{\xi}_{\mathrm{d},k}\left(\mathbf{x}_{\mathrm{d},k-1}, \mathbf{x}_{\mathrm{a},k-1}, \mathbf{u}_\mu(\cdot), t_{k-1}, t_k\right), \tag{32}$$

$$\mathbf{x}_{\mathrm{a},k} = \boldsymbol{\xi}_{\mathrm{a},k}\left(\mathbf{x}_{\mathrm{d},k}, \mathbf{u}_{\mu,k}, t_k\right), \tag{33}$$

where $t_{\{k-1,k\}} \in \mathbb{R}$ with $k \in \mathbb{N}$ are the time instants, $\mathbf{x}_{\mathrm{d},\{k-1,k\}} \in \mathbb{R}^{n_{\mathrm{x,d}}}$ are the differential and $\mathbf{x}_{\mathrm{a},\{k-1,k\}} \in \mathbb{R}^{n_{\mathrm{x,a}}}$ the algebraic states, $\boldsymbol{\xi}_{\mathrm{d},\{k-1,k\}}$ is the differential state-transition function and $\boldsymbol{\xi}_{\mathrm{a},\{k-1,k\}}$ the algebraic state-transition function, at these time instants. The deduced index-1 DAE formulation and the given state-transition functions are utilized subsequently for estimation purposes.

3 Modelling of Systems Subject to Faults

Concerning modelling of systems subject to faults, it is necessary to distinguish between the locations, where faults can occur, and between different fault characteristics. Since faults can affect sensors, actuators and the system itself, these faults commonly are referred to as sensor-, actuator- and system faults. Even though the fault magnitude can posses an arbitrary characteristic with respect to time, it is often modelled either by an abruptly changing signal (abrupt fault) or by a constantly increasing signal (incipient fault).

To identify the fault magnitude of an abrupt fault or the constant slope of an incipient fault, it is appropriate to augment the differential states of the fault-free system which is given by a nonlinear index-1 DAE system

$$\dot{\mathbf{x}}_{\mathrm{d}}^{(\mathrm{f})} = \boldsymbol{\phi}_{\mathrm{d}}^{(\mathrm{f})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{f})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right), \tag{34}$$

$$\mathbf{g}_{\mathrm{a}}^{(\mathrm{f})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{f})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right) = \mathbf{0} \ \wedge \ \det\left(\frac{\partial \mathbf{g}_{\mathrm{a}}^{(\mathrm{f})}}{\partial \mathbf{x}_{\mathrm{a}}}\right) \neq 0, \tag{35}$$

$$\mathbf{y} = \mathbf{h}^{(\mathrm{f})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{f})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right), \tag{36}$$

where $\mathbf{x}_{\mathrm{d}}^{(\mathrm{f})}$ denotes the differential states of the fault-free system and $\boldsymbol{\phi}_{\mathrm{d}}^{(\mathrm{f})}$ the corresponding differential equations, $\mathbf{g}_{\mathrm{a}}^{(\mathrm{f})}$ the algebraic equations and $\mathbf{h}^{(\mathrm{f})}$ the output mapping.

Let $\mathbf{x}_{\mathrm{d}}^{(\mathrm{ab})} = [(\mathbf{x}_{\mathrm{d}}^{(\mathrm{f})})^{\mathrm{T}}, x_{\mathrm{d,aug}}^{(\mathrm{ab})}]^{\mathrm{T}}$ as well as $\mathbf{x}_{\mathrm{d}}^{(\mathrm{in})} = [(\mathbf{x}_{\mathrm{d}}^{(f)})^{\mathrm{T}}, x_{\mathrm{d,aug}}^{(\mathrm{in})}]^{\mathrm{T}}$ denote the augmented state vector for a single abrupt and incipient fault, respectively. Then, modelling an abrupt fault by $\dot{x}_{\mathrm{d,aug}}^{(\mathrm{ab})} = 0$ and the dynamics of an incipient fault by $\dot{x}_{\mathrm{d,aug}}^{(\mathrm{in})} = s$, with s characterizing the slope, yields the augmented models

$$\dot{\mathbf{x}}_{\mathrm{d}}^{(\mathrm{ab})} = \boldsymbol{\phi}_{\mathrm{d}}^{(\mathrm{ab})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{ab})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right) = \begin{bmatrix} \boldsymbol{\phi}_{\mathrm{d}}^{(\mathrm{f})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{ab})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right) \\ 0 \end{bmatrix}, \tag{37}$$

$$\mathbf{g}_{\mathrm{a}}^{(\mathrm{ab})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{ab})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right) = \mathbf{0} \ \wedge \ \det\left(\frac{\partial \mathbf{g}_{\mathrm{a}}^{(\mathrm{ab})}}{\partial \mathbf{x}_{\mathrm{a}}}\right) \neq 0, \tag{38}$$

$$\mathbf{y} = \mathbf{h}^{(\mathrm{ab})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{ab})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right), \tag{39}$$

and

$$\dot{\mathbf{x}}_{\mathrm{d}}^{(\mathrm{in})} = \boldsymbol{\phi}_{\mathrm{d}}^{(\mathrm{in})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{in})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right) = \begin{bmatrix} \boldsymbol{\phi}_{\mathrm{d}}^{(\mathrm{f})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{in})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right) \\ s \end{bmatrix}, \tag{40}$$

$$\mathbf{g}_{\mathrm{a}}^{(\mathrm{in})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{in})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right) = \mathbf{0} \ \wedge \ \det\left(\frac{\partial \mathbf{g}_{\mathrm{a}}^{(\mathrm{in})}}{\partial \mathbf{x}_{\mathrm{a}}}\right) \neq 0, \tag{41}$$

$$\mathbf{y} = \mathbf{h}^{(\mathrm{in})}\left(\mathbf{x}_{\mathrm{d}}^{(\mathrm{in})}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right) \tag{42}$$

for the fault cases. As indicated by the superscripts $(\cdot)^{(\{\mathrm{f,ab,in}\})}$, each of these models characterizes the behaviour of the system in the corresponding mode. Note that the approach of state augmentation is not restricted to a single abrupt or incipient fault, but can also be applied to represent more complex fault characteristics. In contrast to simple fault characteristics, such as a single abrupt or incipient fault that can be modelled by a single augmented state variable, complex fault characteristics typically can only be modelled by more than one augmented state variable. The abrupt and incipient faults are selected to illustrate the modelling of faults. Moreover, under

the assumption that an appropriate estimation approach is used, it is possible to distinguish between a large number of different faults if each fault model describes only a single or a few faults. In contrast, if one fault model describes all faults, e.g. the fault-free system is augmented by all faults, the augmented system model might be unobservable. In general, the more faults are described by one model the more likely it is that observability is lost.

Depending on whether a fault occurs or not, the system operates in different modes. Let an operation mode $m^{(\cdot)}$ be an element of the set $\mathscr{M}$ which contains all fault modes as well as the fault-free mode. As the transitions between modes are probabilistic but neither conditioned on the continuous states nor the continuous inputs, the discrete behaviour is modelled by a Markov chain (cf. Sect. 5.1). The continuous-time dynamics is integrated by associating a DAE system with each state of the Markov chain. This yields a so-called probabilistic hybrid automaton. A complete definition of a probabilistic hybrid automaton is omitted for brevity but can be found in [16]. The state of the hybrid system is described by the triple $\mathbf{x}_\mathrm{h} = (z, \mathbf{x}_\mathrm{d}, \mathbf{x}_\mathrm{a})$ with $z \in \mathscr{M}$, $\mathbf{x}_\mathrm{d} \in \mathbb{R}^{n_{\mathrm{x,d}}}$, $\mathbf{x}_\mathrm{a} \in \mathbb{R}^{n_{\mathrm{x,a}}}$ and corresponding dimensions $n_{\mathrm{x,d}}, n_{\mathrm{x,a}} \in \mathbb{N}$. Assume, for the sake of simplicity, that the system is subject to one abrupt and one incipient fault. Then the related hybrid automaton is depicted in Fig. 1 with $\mathscr{M} = \{m^{(\mathrm{f})}, m^{(\mathrm{ab})}, m^{(\mathrm{in})}\}$, where the labels of the arcs specify the transition probabilities. Here $m^{(\mathrm{f})}$ stands for the behavioural mode of the fault-free system, whereas $m^{(\mathrm{ab})}$ and $m^{(\mathrm{in})}$ indicate the modes if the system is subject to an abrupt and an incipient fault, respectively. The continuous dynamics in each mode (or state) is governed by a nonlinear index-1 DAE system.

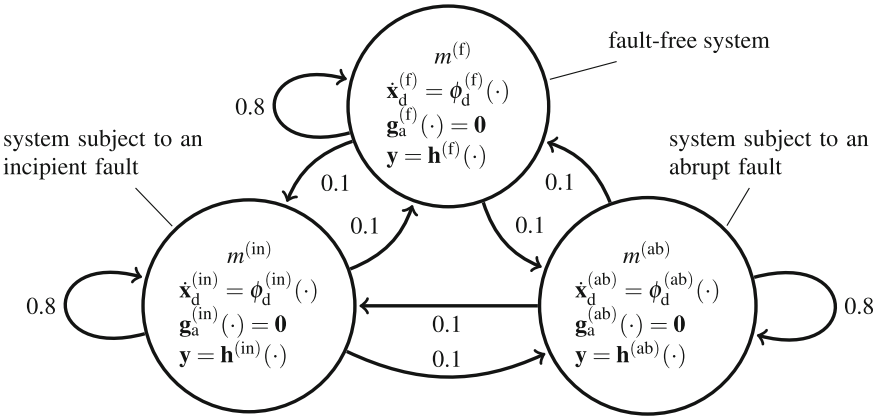


Fig. 1 Hybrid automaton for a system subject to two faults

4 Unscented Kalman Filter for Index-1 DAE Systems

The Unscented Kalman Filter (UKF) is an extension of the original Kalman filter algorithm for the application to nonlinear systems, where the eponymous technique is the Unscented Transformation (UT). The UT introduced in [21] is utilized for the calculation of the mean and covariance of stochastic variables determined by a nonlinear system (cf. also [20]). Therefore, deterministic sets of model-based state predictions and output variables are evaluated. The number of elements in the evaluated sets is relatively small in comparison to a particle filter with a comparable accuracy. Thus, the key advantage of the UKF is the calculation method of the involved stochastic moments avoiding a determination of any Jacobian of the state equations. As a consequence, the estimation of highly nonlinear and stiff systems is typically more accurate and reliable than by Kalman filter variants involving partial derivatives. The computational complexity of the UKF is $\mathscr{O}(L^3)$, where L is the dimension of the vector of the estimated variables. However, the Extended Kalman Filter (EKF), probably the most frequently used and most simple Kalman filter variant for nonlinear systems, has the same order of complexity (c.f. [30]).

The application and parameterization of model-based estimation methods to DAE systems is substantially more complex than the corresponding techniques for ODE systems. A dynamic state observer algorithm for nonlinear DAE systems is developed in [3]. Moving horizon methods are utilized for state estimation to DAEs in [28] and a variant of the EKF for DAE systems is described in [5]. Moreover, an approach for the application of a particle filter to nonlinear index-1 DAEs is given in [14] and an UKF variant for nonlinear DAE systems with index-1 is proposed in [23]. In the following, a revised version of the latter filter, which has already been used for fault detection (c.f. [1]), is presented.

The DAE formulation given by

$$\begin{aligned}\mathbf{x}_{\mathrm{d},k} &= \mathbf{x}_{\mathrm{d},k-1} + \int_{t_{k-1}}^{t_k} \phi_{\mathrm{d}}\left(\mathbf{x}_{\mathrm{d}}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right) \mathrm{d}\, t + \mathbf{v}_{\mathrm{x,d},k-1} \\ &= \xi_{\mathrm{d},k}\left(\mathbf{x}_{\mathrm{d},k-1}, \mathbf{x}_{\mathrm{a},k-1}, \mathbf{u}_{\mu}(\cdot), t_{k-1}, t_k\right) + \mathbf{v}_{\mathrm{x,d},k-1}, \end{aligned} \tag{43}$$

$$\mathbf{g}_{\mathrm{a}}\left(\mathbf{x}_{\mathrm{d}}, \mathbf{x}_{\mathrm{a}}, \mathbf{u}_{\mu}, t\right) = \mathbf{0} \ \wedge \ \det\left(\frac{\partial \mathbf{g}_{\mathrm{a}}}{\partial \mathbf{x}_{\mathrm{a}}}\right) \neq 0, \tag{44}$$

$$\mathbf{x}_{\mathrm{a},k} = \xi_{\mathrm{a},k}\left(\mathbf{x}_{\mathrm{d},k}, \mathbf{u}_{\mu,k}, t_k\right), \tag{45}$$

$$\mathbf{y}_k = \mathbf{h}_k\left(\mathbf{x}_{\mathrm{d},k}, \mathbf{x}_{\mathrm{a},k}, \mathbf{u}_{\mu,k}, t_k\right) + \mathbf{v}_{\mathrm{y},k} \tag{46}$$

is an index-1 system as discussed in Sect. 2.2, where $t_{\{k-1,k\}} \in \mathbb{R}$ with $k \in \mathbb{N}$ are time instants, $\mathbf{x}_{\mathrm{d},\{k-1,k\}} \in \mathbb{R}^{n_{\mathrm{x,d}}}$ are differential states, $\mathbf{x}_{\mathrm{a},\{k-1,k\}} \in \mathbb{R}^{n_{\mathrm{x,a}}}$ are algebraic states, $\mathbf{u}_{\mu,\{k-1,k\}} \subset \mathbb{R}^{n_{\mathrm{u}}}$ are input variables and corresponding time derivatives, $\mathbf{y}_{\{k-1,k\}} \in \mathbb{R}^{n_{\mathrm{y}}}$ are output variables, $\xi_{\mathrm{d},\{k-1,k\}}$ is the differential state-transition function, $\xi_{\mathrm{a},\{k-1,k\}}$ is the algebraic state-transition function, $\mathbf{h}_{\{k-1,k\}}$ is the output function, $\mathbf{v}_{\mathrm{x,d},\{k-1,k\}} \in \mathbb{R}^{n_{\mathrm{x,d}}}$ are stochastic differential state noise signals and $\mathbf{v}_{\mathrm{y},\{k-1,k\}} \in \mathbb{R}^{n_{\mathrm{y}}}$ are stochastic output noise signals at these time instants. Furthermore, ϕ_{d} is the ODE part of

the index-1 DAE system and $\mathbf{g}_\mathrm{a}$ is the restrictive manifold with the local solution ξ_a for algebraic states. The stochastic processes $\mathbf{v}_{\mathrm{x,d},\{k-1,k\}} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{C}_{\mathrm{v_{x,d}v_{x,d}},\{k-1,k\}}\right)$ and $\mathbf{v}_{\mathrm{y},\{k-1,k\}} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{C}_{\mathrm{v_y v_y},\{k-1,k\}}\right)$ are assumed to be normally distributed white noise processes independent of each other and other system quantities with zero mean and known covariances at the corresponding time instants. Completing, $n_{\mathrm{x,d}}, n_{\mathrm{x,a}}, n_\mathrm{u}, n_\mathrm{y} \in \mathbb{N}$ are corresponding dimensions and $\mu \in \mathbb{N}$ is the differentiation index of the original higher index DAE reduced to the given index-1 formulation. Thus, $\mathbf{x}_{\mathrm{d},\{k-1,k\}}, \mathbf{y}_{\{k-1,k\}}$ are stochastic quantities with $\mathrm{E}\left[\mathbf{x}_{\mathrm{d},\{k-1,k\}}\right], \mathrm{E}\left[\mathbf{y}_{\{k-1,k\}}\right]$ denoting the mean and $\mathbf{C}_{\mathrm{x_d x_d},\{k-1,k\}} = \mathrm{Cov}\left(\mathbf{x}_{\mathrm{d},\{k-1,k\}}, \mathbf{x}_{\mathrm{d},\{k-1,k\}}\right)$ $\mathbf{C}_{\mathrm{y\,y},\{k-1,k\}} = \mathrm{Cov}\left(\mathbf{y}_{\{k-1,k\}}, \mathbf{y}_{\{k-1,k\}}\right)$ and $\mathbf{C}_{\mathrm{x_d\,y},\{k-1,k\}} = \mathrm{Cov}\left(\mathbf{x}_{\mathrm{d},\{k-1,k\}}, \mathbf{y}_{\{k-1,k\}}\right)$ the covariances.

As typical for Kalman filters, dual-time indices $k-1|k-1$, $k|k-1$ and $k|k$ are introduced, denoting quantities that are determined considering model information up to the time instant indicated by the left and actual system information, up to the time instant indicated by the right index component. Furthermore, a function $\mathrm{srm}\,(\cdot)$ is introduced, denoting the matrix square root operation applicable to square positive definite matrices as

$$\mathrm{srm}\,(\mathbf{A})\,\mathrm{srm}\,(\mathbf{A})^\mathrm{T} = \mathbf{A}. \tag{47}$$

The UKF for DAE systems is represented using matrix operations instead of conventional set operations. Therefore, application of a function defined for a vector argument to a matrix argument with equal number of rows means that this function is applied to each column of the matrix. The result is a matrix with the same number of columns as the input matrix. If several matrix arguments are used instead of vector arguments, the column numbers of the matrices have to be equal. The estimation algorithm is structured as usual into the steps *Initialization*, *Prediction* and *Correction* as given below. Thus, the initialization is executed at the beginning of the estimation, whereas prediction and correction are repeated sequentially for each taken measurement.

The weights for the scaled UT are defined by

$$\mathbf{w}_\mathrm{m} = \frac{\left[2\alpha^2(n_{\mathrm{x,d}} + \gamma) - 2n_{\mathrm{x,d}},\ \mathbf{1}_{1\times 2n_{\mathrm{x,d}}}\right]^\mathrm{T}}{2\alpha^2(n_{\mathrm{x,d}} + \gamma)} \tag{48}$$

and

$$\mathbf{w}_\mathrm{c} = \mathbf{w}_\mathrm{m} + \left[(1 - \alpha^2 + \beta),\ \mathbf{0}_{1\times 2n_{\mathrm{x,d}}}\right]^\mathrm{T}, \tag{49}$$

$$\Psi = \mathbf{I}_{2n_{\mathrm{x,d}}+1} - \mathbf{w}_\mathrm{m}\mathbf{1}_{1\times 2n_{\mathrm{x,d}}+1}, \tag{50}$$

$$\mathbf{W}_\mathrm{c} = \Psi\,\mathrm{diag}\,(\mathbf{w}_\mathrm{c})\,\Psi^\mathrm{T}, \tag{51}$$

where $0 < \alpha \leq 1, 0 \leq \beta, 0 \leq \gamma$ are scalar-valued set-up parameters of the UT, $\mathbf{w}_\mathrm{m} \in \mathbb{R}^{2n_{\mathrm{x,d}}+1}$ are the weights for the calculation of the mean and $\mathbf{W}_\mathrm{c} \in \mathbb{R}^{2n_{\mathrm{x,d}}+1\times 2n_{\mathrm{x,d}}+1}$ is the matrix-valued weight for calculation of the covariance. Further, $\mathbf{I}_{(\cdot)}$, $\mathbf{0}_{(\cdot)}$ and

$\mathbf{1}_{(\cdot)}$ denote identity, zeros and ones matrices of dimension specified in the index. Remaining quantities are locally defined auxiliaries.

Initialization :

$$k-1=k_0 \tag{52}$$

$$\mathrm{E}\left[\mathbf{x}_{\mathrm{d},k-1|k-1}\right]=\mathbf{x}_{\mathrm{d},0} \tag{53}$$

$$\mathbf{C}_{\mathrm{x_d x_d},k-1|k-1}=\mathbf{C}_{\mathrm{x_d x_d},0} \tag{54}$$

At the initial time instant t_{k_0}, estimates of the initial differential states $\mathbf{x}_{\mathrm{d},0}$ and of the corresponding covariance $\mathbf{C}_{\mathrm{x_d x_d},0}$ are assumed.

Prediction :

$$\mathbf{D}_{\mathrm{x_d},k-1|k-1}=\left[\mathbf{0}_{n_{\mathrm{x,d}}\times 1},-\operatorname{srm}\left(\mathbf{C}_{\mathrm{x_d x_d},k-1|k-1}\right),\operatorname{srm}\left(\mathbf{C}_{\mathrm{x_d x_d},k-1|k-1}\right)\right] \tag{55}$$

$$\mathbf{S}_{\mathrm{x_d},k-1|k-1}=\mathrm{E}\left[\mathbf{x}_{\mathrm{d},k-1|k-1}\right]\mathbf{1}_{1\times 2n_{\mathrm{x,d}}+1}+\alpha\sqrt{n_{\mathrm{x,d}}+\gamma}\,\mathbf{D}_{\mathrm{x_d},k-1|k-1} \tag{56}$$

$$\mathbf{S}_{\mathrm{x_a},k-1|k-1}=\xi_{\mathrm{a},k-1}\left(\mathbf{S}_{\mathrm{x_d},k-1|k-1},\mathbf{u}_{\mu,k-1},t_{k-1}\right) \tag{57}$$

$$\mathbf{S}^{*}_{\mathrm{x_d},k|k-1}=\xi_{\mathrm{d},k}\left(\mathbf{S}_{\mathrm{x_d},k-1|k-1},\mathbf{S}_{\mathrm{x_a},k-1|k-1},\mathbf{u}_{\mu}(\cdot),t_{k-1},t_k\right)+\mathrm{E}\left[\mathbf{v}_{\mathrm{x,d},k-1}\right]\mathbf{1}_{1\times 2n_{\mathrm{x,d}}+1} \tag{58}$$

$$\mathrm{E}\left[\mathbf{x}_{\mathrm{d},k|k-1}\right]=\mathbf{S}^{*}_{\mathrm{x_d},k|k-1}\mathbf{w}_{\mathrm{m}} \tag{59}$$

$$\mathbf{C}_{\mathrm{x_d x_d},k|k-1}=\mathbf{S}^{*}_{\mathrm{x_d},k|k-1}\mathbf{W}_{\mathrm{c}}\mathbf{S}^{*T}_{\mathrm{x_d},k|k-1}+\mathbf{C}_{\mathrm{v_{x,d}v_{x,d}},k} \tag{60}$$

$$\mathbf{D}_{\mathrm{x_d},k|k-1}=\left[\mathbf{0}_{n_{\mathrm{x,d}}\times 1},-\operatorname{srm}\left(\mathbf{C}_{\mathrm{x_d x_d},k|k-1}\right),\operatorname{srm}\left(\mathbf{C}_{\mathrm{x_d x_d},k|k-1}\right)\right] \tag{61}$$

$$\mathbf{S}_{\mathrm{x_d},k|k-1}=\mathrm{E}\left[\mathbf{x}_{\mathrm{d},k|k-1}\right]\mathbf{1}_{1\times 2n_{\mathrm{x,d}}+1}+\alpha\sqrt{n_{\mathrm{x,d}}+\gamma}\,\mathbf{D}_{\mathrm{x_d},k|k-1} \tag{62}$$

$$\mathbf{S}_{\mathrm{x_a},k|k-1}=\xi_{\mathrm{a},k}\left(\mathbf{S}_{\mathrm{x_d},k|k-1},\mathbf{u}_{\mu,k},t_k\right) \tag{63}$$

$$\mathbf{S}_{\mathrm{y},k|k-1}=\mathbf{h}_{k}\left(\mathbf{S}_{\mathrm{x_d},k|k-1},\mathbf{S}_{\mathrm{x_a},k|k-1},\mathbf{u}_{\mu,k},t_k\right)+\mathrm{E}\left[\mathbf{v}_{\mathrm{y},k}\right]\mathbf{1}_{1\times 2n_{\mathrm{x,d}}+1} \tag{64}$$

$$\mathrm{E}\left[\mathbf{y}_{k|k-1}\right]=\mathbf{S}_{\mathrm{y},k|k-1}\mathbf{w}_{\mathrm{m}} \tag{65}$$

$$\mathbf{C}_{\mathrm{y\,y},k|k-1}=\mathbf{S}_{\mathrm{y},k|k-1}\mathbf{W}_{\mathrm{c}}\mathbf{S}^{T}_{\mathrm{y},k|k-1}+\mathbf{C}_{\mathrm{v_y v_y},k} \tag{66}$$

$$\mathbf{C}_{\mathrm{x_d\,y},k|k-1}=\mathbf{S}_{\mathrm{x_d},k|k-1}\mathbf{W}_{\mathrm{c}}\mathbf{S}^{T}_{\mathrm{y},k|k-1} \tag{67}$$

In the calculation step (55)–(67), the means $\mathrm{E}\left[\mathbf{x}_{\mathrm{d},k|k-1}\right]$, $\mathrm{E}\left[\mathbf{y}_{k|k-1}\right]$ as well as the covariances $\mathbf{C}_{\mathrm{x_d x_d},k|k-1}$, $\mathbf{C}_{\mathrm{y\,y},k|k-1}$, $\mathbf{C}_{\mathrm{x_d\,y},k|k-1}$ are determined using the model and the UT. The columns of the matrices composed of state and output values $\mathbf{S}_{\mathrm{x_d},k|k-1}$, $\mathbf{S}_{\mathrm{x_a},k|k-1}$, $\mathbf{S}_{\mathrm{y},k|k-1}$ are commonly referred to as sigma points. Remaining intermediate variables are auxiliaries enabling a clear representation of the algorithm.

Correction :

$$\mathbf{K}_{\mathrm{x_d},k}=\mathbf{C}_{\mathrm{x_d\,y},k|k-1}\mathbf{C}^{-1}_{\mathrm{y\,y},k|k-1} \tag{68}$$

$$\mathrm{E}\left[\mathbf{x}_{\mathrm{d},k|k}\right]=\mathrm{E}\left[\mathbf{x}_{\mathrm{d},k|k-1}\right]+\mathbf{K}_{\mathrm{x_d},k}\left(\mathbf{y}_{\mathrm{m},k}-\mathrm{E}\left[\mathbf{y}_{k|k-1}\right]\right) \tag{69}$$

$$\mathbf{C}_{\mathrm{x_d x_d},k|k}=\mathbf{C}_{\mathrm{x_d x_d},k|k-1}-\mathbf{K}_{\mathrm{x_d},k}\mathbf{C}_{\mathrm{y\,y},k|k-1}\mathbf{K}^{\mathrm{T}}_{\mathrm{x_d},k} \tag{70}$$

In this final calculation step, the Kalman gain $\mathbf{K}_{\mathrm{x_d},k}$ is determined. Thus, the mean $\mathrm{E}\left[\mathbf{x}_{\mathrm{d},k|k}\right]$ and the covariance $\mathbf{C}_{\mathrm{x_d x_d},k|k}$ of the estimated differential states can be obtained accordingly, considering the actual system measurements $\mathbf{y}_{\mathrm{m},k}$.

The substantial difference between the UKF variants for ODE and DAE systems is the determination of the algebraic states $\mathbf{S}_{\mathrm{x_a},k-1|k-1}$ and $\mathbf{S}_{\mathrm{x_a},k|k-1}$ within the DAE variant. This refers, first, to the consistent initialization of the differential state-transition function for prediction in each considered discrete-time instant and second, to the calculation of modelled outputs. Thus, the algebraic constraint imposed on the generalized DAE states is satisfied by the estimation algorithm. Furthermore, the modelled system measurements may depend on both the differential as well as the algebraic states.

5 Hybrid Estimation for DAE Systems

As described in Sect. 3, it is appropriate to model systems subject to faults as hybrid systems. Thus, in order to solve the fault identification task, both the discrete- and the continuous-valued states have to be determined, which can be done using a hybrid state estimation approach. Discussing the drawbacks of both the simplest and the optimal hybrid estimation approach indicates why neither of them is applicable to fault identification purposes. A suitable alternative is a suboptimal hybrid estimation scheme such as the Interacting Multiple Model (IMM) approach. This approach originated in the context of target tracking, but has also shown to be effective for fault diagnosis if the continuous dynamics of each mode in the related hybrid automaton is modelled by a set of ODEs, i.e. by an index-0 DAE system. Since the IMM approach assumes that the discrete dynamics is modelled by a Markov chain of order one, this will be introduced before proceeding with the aforementioned hybrid estimation schemes for DAE systems.

5.1 Discrete-Time Markov Chains

A discrete-time Markov chain is a model for representing stochastic processes, whereby state changes occur at discrete-time instants $k \in \mathbb{N}$ and the state-space is either a finite or countable set. The characteristic property of a Markov chain is that the value of the discrete state z_k is only influenced by the value of the state z_{k-1}, but not by the values of the states z_i with $i \in \{0, 1, \ldots, k-2\}$. In mathematical terms, the Markov property is given by

$$\mathbf{P}\left(z_k = m^{(i)} | z_{k-1} = m^{(j)}, \ldots, z_0 = m^{(\ell)}\right) = \mathbf{P}\left(z_k = m^{(i)} | z_{k-1} = m^{(j)}\right) \tag{71}$$

with $m^{(\cdot)} \in \mathscr{M}$. The probability of z_k being in the discrete state i or, if it comes to fault diagnosis, in mode $m^{(i)}$ given that the z_{k-1} is in mode $m^{(j)}$ is called the state-transition probability and denoted by

$$P_{ij} = \mathbf{P}\left(z_k = m^{(i)} | z_{k-1} = m^{(j)}\right). \tag{72}$$

Assume that the number of discrete states $n_z = |\mathscr{M}|$ is finite. Then the state-transition probabilities can be arranged in a transition probability matrix $\mathbf{T} \in \mathbb{R}^{n_z \times n_z}$, where each element $t_{ij} = P_{ij}$ assigns the probability for a state-transition form $z_{k-1} = m^{(j)}$ to $z_k = m^{(i)}$ for all $k \in \mathbb{N}$. Thus, the probability transition matrix possess the properties

$$t_{ij} \geq 0 \ \wedge \ \sum_{i=1}^{n_z} t_{ij} = 1 \tag{73}$$

for all $i, j \in \{1, \ldots, n_z\}$. Assume that the continuous dynamics in each state of the automaton depicted in Fig. 1 are removed, then that automaton immediately reduces to a Markov chain.

5.2 Hybrid Estimation Schemes

Non-switching hybrid estimation. The simplest hybrid estimation scheme is the non-switching approach, which is also known as fixed model approach or static case. In this approach, it is assumed that the system stays in one mode for the entire time, i.e. the system does not switch to another mode during the estimation process. Furthermore, a bank of $q = |\mathscr{M}|$ Kalman filters is implemented to perform the estimation of the mode-conditioned mean

$$\hat{\mathbf{x}}_{k|k}^{(i)} = \mathrm{E}\left[\mathbf{x}_k | Y_k, m^{(i)}\right] \tag{74}$$

and covariance

$$\mathbf{P}_{k|k}^{(i)} = \mathrm{E}\left[\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^{(i)}\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^{(i)}\right)^{\mathrm{T}} | Y_k, m^{(i)}\right] \tag{75}$$

of each mode $m^{(i)} \in \mathscr{M}$, where $\mathbf{x}_k = [\mathbf{x}_{k,\mathrm{d}}^{\mathrm{T}}, \mathbf{x}_{k,\mathrm{a}}^{\mathrm{T}}]^{\mathrm{T}}$ denotes the concatenation of differential and algebraic states and $Y_k = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ is the sequence of measurements up to time instant k.

In order to estimate the discrete-valued variable, i.e. the mode the system is operating in, the most likely mode has to be determined. Applying Bayes' theorem, the probability of each mode at time instant k can be calculated recursively by

$$\mu_k^{(i)} = \mathbf{P}\left(m^{(i)} | Y_k\right) = \mathbf{P}\left(m^{(i)} | \mathbf{y}_k, Y_{k-1}\right) = \frac{p\left(\mathbf{y}_k | Y_{k-1}, m^{(i)}\right) \mathbf{P}\left(m^{(i)} | Y_{k-1}\right)}{p\left(\mathbf{y}_k | Y_{k-1}\right)}$$

$$= \frac{p\left(\mathbf{y}_k | Y_{k-1}, m^{(i)}\right) \mu_{k-1}^{(i)}}{\sum_{j=1}^{q} p\left(\mathbf{y}_k | Y_{k-1}, m^{(j)}\right) \mu_{k-1}^{(j)}}. \tag{76}$$

The conditional probability density function $p(\mathbf{y}_k|Y_{k-1}, m^{(i)})$ in (76) can, according to [16], be replaced by the likelihood function

$$\Lambda_k^{(i)} = \frac{\mathrm{e}^{-\frac{1}{2}(\mathbf{r}_k^{(i)})^{\mathrm{T}} \mathbf{S}_k^{-1} \mathbf{r}_k^{(i)}}}{\sqrt{(2\pi)^{n_y} |\mathbf{S}_k^{(i)}|}}, \tag{77}$$

where $\mathbf{r}_k^{(i)} = \mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}^{(i)}$ is the residual of the i-th Kalman filter and $\mathbf{S}_k^{(i)}$ is the corresponding residual covariance matrix. Thus, fault identification can be done by exploiting the current hybrid estimate $\hat{\mathbf{x}}_{\mathrm{h},k} = (\hat{z}_k, \hat{\mathbf{x}}_{k|k}^{(i)})$, where $\hat{z}_k = m^{(i)}$ is the most likely mode with $\mu_k^{(i)} \geq \mu_k^{(j)}$ for all $j \in \{1, 2, \ldots, q\}$ and $\hat{\mathbf{x}}_{k|k}^{(i)}$ is according to (74) the mode-conditioned mean.

Moreover, the mode probability $\mu_k^{(i)}$ is also used to determine the combined mean

$$\hat{\mathbf{x}}_k = \mathrm{E}\left[\mathbf{x}_k | Y_k\right] = \sum_{i=1}^{q} \mathbf{P}\left(m^{(i)} | Y_k\right) \mathrm{E}\left[\mathbf{x}_k | Y_k, m^{(i)}\right] = \sum_{i=1}^{q} \mu_k^{(i)} \hat{\mathbf{x}}_{k|k}^{(i)} \tag{78}$$

and covariance matrix

$$\begin{aligned}
\mathbf{P}_k &= \mathrm{E}\left[\left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right)^{\mathrm{T}} | Y_k\right] \\
&= \sum_{i=1}^{q} \mathbf{P}\left(m^{(i)} | Y_k\right) \mathrm{E}\left[\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^{(i)}\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^{(i)}\right)^{\mathrm{T}} | Y_k, m^{(i)}\right] \\
&\quad + \sum_{i=1}^{q} \mathbf{P}\left(m^{(i)} | Y_k\right)\left(\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_k\right)\left(\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_k\right)^{\mathrm{T}} \\
&= \sum_{i=1}^{q} \mu_k^{(i)} \mathbf{P}_{k|k}^{(i)} + \sum_{i=1}^{q} \mu_k^{(i)}\left(\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_k\right)\left(\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_k\right)^{\mathrm{T}}
\end{aligned} \tag{79}$$

of the continuous-valued states.

Optimal hybrid estimation. Concerning the assumption of the non-switching approach that the system has to stay in one mode, it is obvious that this approach is only appropriate for faults, which occurred before the estimation process has been started. In order to cope with faults, which occur during the estimation process, the optimal hybrid estimation approach can be used. The switching between different modes, or rather between different states of the hybrid automation, is modelled by a Markov chain with known state-transition probabilities

$$P_{ij} = \mathbf{P}\left(m_k^{(i)} | m_{k-1}^{(j)}\right), \tag{80}$$

where $m_k^{(i)}, m_{k-1}^{(j)} \in \mathscr{M}$ denote that the system is in mode $m^{(i)}$ at time k and in mode $m^{(j)}$ at time $k-1$, respectively.

Due to the switching, it is necessary to track all q^k possible mode sequences

$$M_k^{(i)} = \left\{ m_0^{(i_0)}, m_1^{(i_1)}, \ldots, m_k^{(i_k)} \right\} \tag{81}$$

with $i \in \{1, 2, \ldots, q^k\}$, $m_j^{(i_j)} \in \mathscr{M}$ being the mode of the i-th mode sequence at time instant $j \in \{0, 1, \ldots, k\}$. In contrast to the non-switching case, the optimal hybrid estimation does not require one Kalman filter for each mode $m^{(i)} \in \mathscr{M}$ but a filter bank with one filter for each possible mode sequence $M_k^{(i)} \in \{M_k^{(1)}, M_k^{(2)}, \ldots, M_k^{(q^k)}\}$. Note that in this context the number of possible mode sequences increases exponentially over time. These Kalman filters yield

$$\hat{\mathbf{x}}_{k|k}^{(i)} = \mathrm{E}\left[\mathbf{x}_k | Y_k, M_k^{(i)} \right] \tag{82}$$

and

$$\mathbf{P}_{k|k}^{(i)} = \mathrm{E}\left[\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^{(i)} \right) \left(\mathbf{x}_k - \hat{\mathbf{x}}_k^{(i)} \right)^{\mathrm{T}} | Y_k, M_k^{(i)} \right] \tag{83}$$

for each mode sequence.

The estimation of the discrete-valued state is done by first calculating the probabilities of the mode sequences as follows

$$\begin{aligned} \upsilon_k^{(i)} &= \mathbf{P}\left(M_k^{(i)} | Y_k \right) = \mathbf{P}\left(M_k^{(i)} | \mathbf{y}_k, Y_{k-1} \right) = \frac{1}{c} p\left(\mathbf{y}_k | Y_{k-1}, M_k^{(i)} \right) \mathbf{P}\left(M_k^{(i)} | Y_{k-1} \right) \\ &= \frac{1}{c} p\left(\mathbf{y}_k | Y_{k-1}, M_k^{(i)} \right) \mathbf{P}\left(m_k^{(i_k)}, M_{k-1}^{(\ell)} | Y_{k-1} \right) \\ &= \frac{1}{c} p\left(\mathbf{y}_k | Y_{k-1}, M_k^{(i)} \right) \mathbf{P}\left(m_k^{(i_k)} | Y_{k-1}, M_{k-1}^{(\ell)} \right) \mathbf{P}\left(M_{k-1}^{(\ell)} | Y_{k-1} \right) \end{aligned} \tag{84}$$

with the normalisation constant $c = \sum_{j=1}^{q^k} p(\mathbf{y}_k | Y_{k-1}, M_k^{(j)}) \mathbf{P}(M_k^{(j)} | Y_k)$. Since the discrete dynamics are modelled by a Markov chain, the probability of the current mode depends only on the previous mode. Hence, (84) simplifies to

$$\begin{aligned} \upsilon_k^{(i)} &= \frac{1}{c} p\left(\mathbf{y}_k | Y_{k-1}, M_k^{(i)} \right) \mathbf{P}\left(m_k^{(i_k)} | m_{k-1}^{(\ell_{k-1})} \right) \mathbf{P}\left(M_{k-1}^{(\ell)} | Y_{k-1} \right) \\ &= \frac{1}{c} p\left(\mathbf{y}_k | Y_{k-1}, M_k^{(i)} \right) P_{i\ell}\, \upsilon_{k-1}^{(\ell)}. \end{aligned} \tag{85}$$

Note that even though the Markov property has been exploited, the probability density function is still conditioned on the entire mode sequence. The discrete-valued state at time k is then given by the current mode $m_k^{(i_k)}$ of the most likely mode sequence

$\upsilon_k^{(i)} \geq \upsilon_k^{(j)}$ with $j \in \{1, 2, \ldots, q^k\}$, such that the hybrid state is given by $\hat{\mathbf{x}}_{\mathrm{h},k} = (\hat{z}_k = m_k^{(i_k)}, \hat{\mathbf{x}}_{k|k}^{(i)})$.

Finally, the combined mean and covariance matrix are given by

$$\hat{\mathbf{x}}_k = \mathrm{E}\left[x_k | Y_k\right] = \sum_{i=1}^{q^k} \mathbf{P}\left(M_k^{(i)} | Y_k\right) \mathrm{E}\left[\mathbf{x}_k | Y_k, M_k^{(i)}\right] = \sum_{i=1}^{q^k} \upsilon_k^{(i)} \hat{\mathbf{x}}_{k|k}^{(i)} \tag{86}$$

and

$$\mathbf{P}_k = \mathrm{E}\left[\left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right)^{\mathrm{T}} | Y_k\right] = \sum_{i=1}^{q^k} \upsilon_k^{(i)} \mathbf{P}_{k|k}^{(i)} + \sum_{i=1}^{q^k} \upsilon_k^{(i)} \left(\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_k\right)\left(\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_k\right)^{\mathrm{T}}. \tag{87}$$

Hybrid estimation with interacting multiple model approach. Even though it is possible to detect faults, which occur during the estimation process, a major drawback of the optimal approach is the exponential grow of possible mode sequences. Thus, the optimal approach cannot be realized in practice. An appropriate alternative is a suboptimal hybrid estimation approach, such as the IMM approach that merges mode sequences. As in the non-switching case, the IMM approach requires only a filter bank with one Kalman filter for each mode. However, in contrast to the non-switching approach, the Kalman filters share information with each other. For index-1 systems, this is done by calculating the initial mean and covariance of each mode $i \in \{1, 2, \ldots, q\}$ according to the mixing step

$$\begin{bmatrix} \tilde{\mathbf{x}}_{\mathrm{d},k-1}^{(i)} \\ \tilde{\mathbf{x}}_{\mathrm{a},k-1}^{(i)} \end{bmatrix} = \sum_{j=1}^{q} \lambda_{k-1}^{(ij)} \begin{bmatrix} \hat{\mathbf{x}}_{\mathrm{d},k-1|k-1}^{(j)} \\ \hat{\mathbf{x}}_{\mathrm{a},k-1|k-1}^{(j)} \end{bmatrix} \tag{88}$$

and

$$\begin{aligned} \tilde{\mathbf{P}}_{k-1}^{(i)} &= \begin{bmatrix} \tilde{\mathbf{P}}_{x_\mathrm{d}x_\mathrm{d},k-1}^{(i)} & \tilde{\mathbf{P}}_{x_\mathrm{d}x_\mathrm{a},k-1}^{(i)} \\ \tilde{\mathbf{P}}_{x_\mathrm{a}x_\mathrm{d},k-1}^{(i)} & \tilde{\mathbf{P}}_{x_\mathrm{a}x_\mathrm{a},k-1}^{(i)} \end{bmatrix} \\ &= \sum_{j=1}^{q} \lambda_{k-1}^{(ij)} \begin{bmatrix} \mathbf{P}_{x_\mathrm{d}x_\mathrm{d},k-1|k-1}^{(j)} & \mathbf{P}_{x_\mathrm{d}x_\mathrm{a},k-1|k-1}^{(j)} \\ \mathbf{P}_{x_\mathrm{a}x_\mathrm{d},k-1|k-1}^{(j)} & \mathbf{P}_{x_\mathrm{a}x_\mathrm{a},k-1|k-1}^{(j)} \end{bmatrix} + \mu_{ij} \begin{bmatrix} \Delta\mathbf{x}_{\mathrm{d},ji}\Delta\mathbf{x}_{\mathrm{d},ji}^{\mathrm{T}} & \Delta\mathbf{x}_{\mathrm{d},ji}\Delta\mathbf{x}_{\mathrm{a},ji}^{\mathrm{T}} \\ \Delta\mathbf{x}_{\mathrm{a},ji}\Delta\mathbf{x}_{\mathrm{d},ji}^{\mathrm{T}} & \Delta\mathbf{x}_{\mathrm{a},ji}\Delta\mathbf{x}_{\mathrm{a},ji}^{\mathrm{T}} \end{bmatrix} \end{aligned} \tag{89}$$

at time instant $k-1$ with $\Delta\mathbf{x}_{\mathrm{d},ji} = \hat{\mathbf{x}}_{\mathrm{d},k-1|k-1}^{(j)} - \tilde{\mathbf{x}}_{\mathrm{d},k-1}^{(i)}$, $\Delta\mathbf{x}_{\mathrm{a},ji} = \hat{\mathbf{x}}_{\mathrm{a},k-1|k-1}^{(j)} - \tilde{\mathbf{x}}_{\mathrm{a},k-1}^{(i)}$ and the mixing probability

$$\lambda_{k-1}^{(ij)} = \mathbf{P}\left(m_{k-1}^{(i)} | m_k^{(j)}, Y_{k-1}\right) = \frac{P_{ji}\mu_{k-1}^{(i)}}{\sum_{i=1}^{q} P_{ji}\mu_{k-1}^{(i)}} \tag{90}$$

with $i, j \in \{1, 2, \ldots, q\}$. As indicated by (88) and (89), neither the mixed mean nor the mixed covariance matrix of the differential states depend on the algebraic states. Note that the Kalman filter for index-1 systems only estimates the mean and covariance matrix of the differential states, whereas the algebraic states are calculated such that the algebraic constraints are met. Thus, the IMM algorithm only has to be set up for the differential states and has the same structure as for systems given in state-space form. Taking this into account, it seems reasonable to sketch briefly the steps of the IMM algorithm and to illustrate one cycle of this algorithm in Fig. 2. A detailed discussion can be found in [4].

1. Calculate all mixing probabilities $\lambda_{k-1}^{(ij)}$ with $i, j \in \{1, 2, \ldots, q\}$ according to (90).
2. Calculate the mixed initial mean $\tilde{\mathbf{x}}_{\mathrm{d},k-1}^{(i)}$ and covariance $\tilde{\mathbf{P}}_{x_\mathrm{d}x_\mathrm{d},k-1}^{(i)}$ of the differential states of each mode $i \in \{1, 2, \ldots, q\}$ according to the corresponding parts of (88) and (89).
3. Perform Kalman filtering according to the algorithm given in Sect. 4 to calculate $\hat{\mathbf{x}}_{\mathrm{d},k|k}^{(i)}$, $\mathbf{P}_{x_\mathrm{d}x_\mathrm{d},k|k}^{(i)}$ and the likelihood function $\Lambda_k^{(i)}$ of each mode.
4. Calculate the mode probabilities

$$\mu_k^{(i)} = \mathbf{P}\left(m_k^{(i)} | Y_k\right) = \frac{\Lambda_k^{(i)} \sum_{j=1}^{q} P_{ij} \mu_{k-1}^{(j)}}{\sum_{i=1}^{q} \Lambda_k^{(i)} \left(\sum_{j=1}^{q} P_{ij} \mu_{k-1}^{(j)}\right)}. \tag{91}$$

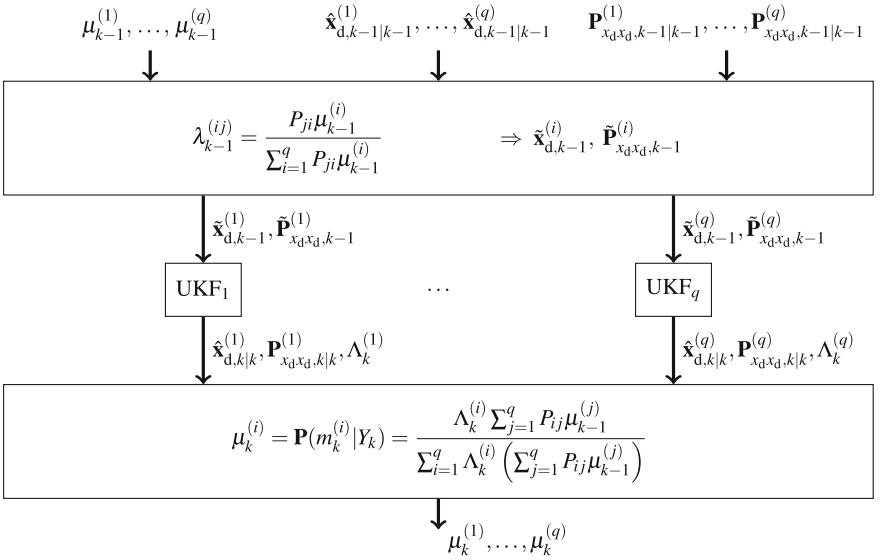


Fig. 2 One cycle of the recursive IMM algorithm

5. Calculate the combined mean and covariance matrix according to (78) and (79), respectively, with the mode-conditioned mean values and covariance matrices that have been computed in step 3 and the mode probabilities from step 4.

6 Numerical Results

The concept presented in the previous sections is particularly applicable to fault identification of physical systems modelled as general DAEs which are representable as an index-1 DAE system. Therefore, the faults considered for the supervision are modelled in compliance with the described formalism. The essential substructure of the supervision procedure is the appropriate physical state estimation performed by a set of UKFs. In this section, the efficiency of the proposed approach is illustrated with reference to a realistic simulated system.

The system considered is a hydraulic cylinder depicted in Fig. 3. The working cycle is controlled by pressures imposed at the hydraulic ports pt_1 and pt_2 in order to execute mechanical work at port pt_3. Further, 5 energy storage elements are modelled. These are compliances of the hydraulic compartments c_1 and c_2, an inertia of moved masses aggregated to i_1, a compliance of displacement limiting elements represented by se_2 and a compliance included in the mechanical load model. The transformation from hydraulic to mechanical power takes place at the piston surfaces assigned by tf_1 and tf_2. The dissipative elements in the hydraulic and mechanical domain are involved by r_1–r_4 and r_5–r_6, respectively. Moreover, the ambient pressure is represented by the element se_1. The pressures in the hydraulic compartments and the positions of the piston and mechanical port are measurable output variables (sample time $T_S = 10\,\text{ms}$). These measurements are superposed additively by an appropriate discrete-time white Gaussian noise process.

The nonlinear characteristics of the model are induced by equations describing the mechanical load, hydraulic resistances as well as position limiting forces. Any switching behaviour is avoided by the introduction of steep sigmoid functions. Furthermore, the time characteristics of the differential states are significantly different. As a consequence, the model exhibits a highly nonlinear and stiff behaviour.

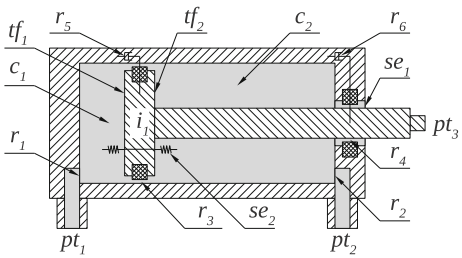


Fig. 3 Simulated system: Hydraulic cylinder

The index-1 DAE model includes 5 differential and 41 algebraic equations. Hence, the model is not presented completely here for sake of space but can be provided by the authors on request. Subsequently, only the most essential equations are discussed in detail. These are the differential equations as well as equations describing mechanical load, position limiting forces and a typical implemented hydraulic resistance.

The energy storage elements mentioned before comply with the corresponding differential constitutive equations

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\begin{bmatrix} p_{\mathrm{c},1} \\ p_{\mathrm{c},2} \\ v_{\mathrm{i},1} \\ d_{\mathrm{se},2} \\ d_{\mathrm{pt},3} \end{bmatrix} = \operatorname{diag}\begin{bmatrix} a_{\mathrm{c},1} \\ a_{\mathrm{c},2} \\ a_{\mathrm{i},1} \\ 1 \\ 1 \end{bmatrix}^{-1}\begin{bmatrix} \dot{V}_{\mathrm{c},1} \\ \dot{V}_{\mathrm{c},2} \\ F_{\mathrm{i},1} \\ v_{\mathrm{se},2} \\ v_{\mathrm{pt},3} \end{bmatrix} \tag{92}$$

giving the model of the differential states. In (92) $p_{\mathrm{c},1}$, $p_{\mathrm{c},2}$ indicate the pressures in the hydraulic compartments, $\dot{V}_{\mathrm{c},1}$, $\dot{V}_{\mathrm{c},2}$ the corresponding volume flow rates and $a_{\mathrm{c},1}$, $a_{\mathrm{c},2}$ the associated compliance coefficients. Moreover, $v_{\mathrm{i},1}$ is the velocity of the aggregated mass, $F_{\mathrm{i},1}$ the force applied to the mass and $a_{\mathrm{i},1}$ the mass. Further, $d_{\mathrm{se},2}$, $d_{\mathrm{pt},3}$ are the displacements of the piston and the mechanical port calculated as time integrals over corresponding velocities $v_{\mathrm{se},2}$, $v_{\mathrm{pt},3}$.

The implemented mechanical load has a viscoelastic characteristic. Thus, the load force $F_{\mathrm{pt},3}$ at the mechanical port pt_3 is

$$F_{\mathrm{pt},3} = \frac{\dfrac{a_{\mathrm{pt},3,3}\,v_{\mathrm{pt},3}}{1+e^{-a_{\mathrm{pt},3,1}\,v_{\mathrm{pt},3}}}}{1+e^{-a_{\mathrm{pt},3,1}\,(d_{\mathrm{pt},3}-a_{\mathrm{pt},3,4})}} + \frac{\dfrac{d_{\mathrm{pt},3}-a_{\mathrm{pt},3,5}}{a_{\mathrm{pt},3,2}}}{1+e^{-a_{\mathrm{pt},3,1}\,(d_{\mathrm{pt},3}-a_{\mathrm{pt},3,5})}}, \tag{93}$$

where the previously introduced velocity $v_{\mathrm{pt},3}$ and displacement $d_{\mathrm{pt},3}$ of the port are involved. Appropriate setting of parameters $a_{\mathrm{pt},3,\{1,2,3,4,5\}}$ results in a realistic load function. The piston position limiting force $F_{\mathrm{se},2}$ is modeled rather simple by

$$F_{\mathrm{se},2} = \frac{\dfrac{d_{\mathrm{se},2}-a_{\mathrm{se},2,3}}{a_{\mathrm{se},2,2}}}{1+e^{a_{\mathrm{se},2,1}\,(d_{\mathrm{se},2}-a_{\mathrm{se},2,3})}} + \frac{\dfrac{d_{\mathrm{se},2}-a_{\mathrm{se},2,4}}{a_{\mathrm{se},2,2}}}{1+e^{-a_{\mathrm{se},2,1}\,(d_{\mathrm{se},2}-a_{\mathrm{se},2,4})}}, \tag{94}$$

i.e. it depends only on the displacement of the piston $d_{\mathrm{se},2}$ but dampers at the end positions are not considered. Again, the parameters $a_{\mathrm{se},2,\{1,2,3,4\}}$ are set in reference to the technical system considered here. Both equations imply sigmoidal parts achieved by exponential functions, which effect smooth activation and deactivation of the forces according to the set parameter values.

Since all hydraulic resistances involved in the system are described by equivalent models, the equations

$$\Delta p_{\mathrm{r},1} = q_{\mathrm{r},1,2}\,\dot{V}_{\mathrm{r},1}, \tag{95}$$

$$q_{\mathrm{r},1,1} = \tanh(a_{\mathrm{r},1,1}\Delta p_{\mathrm{r},1})\Delta p_{\mathrm{r},1} + a_{\mathrm{r},1,3}\left(1 - \tanh\left(\frac{\tanh(a_{\mathrm{r},1,1}\Delta p_{\mathrm{r},1})\Delta p_{\mathrm{r},1}}{a_{\mathrm{r},1,3}}\right)\right), \tag{96}$$

$$q_{\mathrm{r},1,2} = a_{\mathrm{r},1,2}\sqrt{q_{\mathrm{r},1,1}} \tag{97}$$

are given only for the component r_1. Here, $\Delta p_{\mathrm{r},1}$ is the pressure drop across the resistance, $\dot{V}_{\mathrm{r},1}$ the corresponding volume flow rate, $q_{\mathrm{r},1,\{1,2\}}$ are the auxiliary quantities and $a_{\mathrm{r},1,\{1,2,3\}}$ the associated parameters. This model assure a numerically stable zero crossing of the pressure drop when the flow direction changes, which is realized by an approximation of the absolute value of the pressure drop in the calculation of the hydraulic resistance coefficient using the tanh-function.

All remaining components are modelled object-oriented. According to the system structure, a set of algebraic connection equations describes the power exchange within the model. The set of non-differential constitutive and connection equations are solvable for the algebraic states such that the hydraulic cylinder is modelled by an index-1 DAE system.

The working cycle can be conveniently discussed considering the pressures at the hydraulic ports u_1, u_2, the pressures in the hydraulic compartments $p_{\mathrm{c},1}$, $p_{\mathrm{c},2}$ and the displacement of the mechanical port $d_{\mathrm{pt},3}$. A corresponding plot is given in Fig. 4 where altering backgrounds indicate 4 distinctive phases of the working cycle. During the first phase, the actuator moves forward until the mechanical load effects a rapid deceleration. Then, in the second phase, the actuator executes effective work. After an input pressure change, the actuator moves backwards. In the last phase, the system is at rest in the initial position. Between these phases, characteristic transient processes occur with significant high-frequency oscillations. These oscillations are not apparent due to the scaling used in Fig. 4.

Three operation modes $m^{(i)}$ with $i \in \{1, 2, 3\}$ are considered in the following: $m^{(1)}$ stands for the fault-free operation, $m^{(2)}$ for the operation with an increased internal leakage and $m^{(3)}$ for the operation with an increased external leakage. The internal leakage can increase, if the piston sealing of the hydraulic cylinder is worn or defective. If the piston rod sealing is affected, the system can lose hydraulic fluid and operates with increased external leakage. The operation modes with increased leakage are introduced in the simulated model by a modification of the resistance

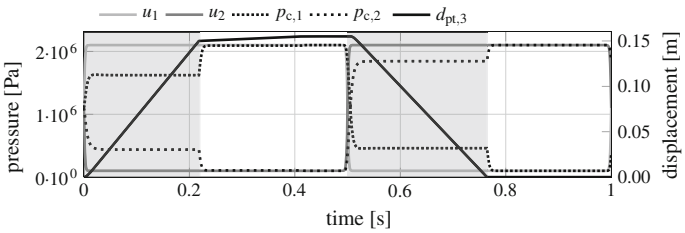


Fig. 4 Simulation results: Behaviour of the fault-free system

coefficients of elements r_3 and r_4 by a factor $10^{s_{\mathrm{a},1}}$ and $10^{s_{\mathrm{a},2}}$, respectively. Within the hybrid state estimation concept, the quantities $s_{\mathrm{a},1}$ and $s_{\mathrm{a},2}$ are estimated by mode allocated Kalman filters calculating the corresponding estimates $x_{\mathrm{s,a},1}$ and $x_{\mathrm{s,a},2}$. Thus, fault identification is implemented by calculation of the quantities $x_{\mathrm{s,a},1}$ and $x_{\mathrm{s,a},2}$.

In order to rate how much the measurable system outputs are influenced by each fault, measurable residua $e_{\mathrm{y},i}$ with $i \in \{1, 2, 3\}$ are defined for each mode $m^{(i)}$ of the noise-free system by

$$e_{\mathrm{y},i} = \frac{\left(\mathbf{y}_{\mathrm{s,inv}}\right)^{\mathrm{T}} \mathrm{abs_c}\left(\mathbf{y}_i - \mathbf{y}_1\right)}{\left(\mathbf{y}_{\mathrm{s,inv}}\right)^{\mathrm{T}} \mathrm{abs_c}\left(\mathbf{y}_1\right)}, \qquad \mathbf{y}_{\mathrm{s}} = \frac{1}{\tau_{\mathrm{wc}}} \int_{\tau}^{\tau+\tau_{\mathrm{wc}}} \mathrm{abs_c}\left(\mathbf{y}_1\right) \mathrm{d}\, t. \tag{98}$$

Here $\mathbf{y}_i \in \mathbb{R}^{n_{\mathrm{y}}}$ with $n_{\mathrm{y}} = 4$ are the output variables of each mode, $\tau_{\mathrm{wc}} = 1\mathrm{s}$ denotes the working cycle time range, $\mathbf{y}_{\mathrm{s,inv}} = [(y_{\mathrm{s},1})^{-1}, \ldots, (y_{\mathrm{s},4})^{-1}]^{\mathrm{T}}$ is a component-wise inverted output scaling factor with $y_{\mathrm{s},i} \neq 0$ for $i \in \{1, 2, 3, 4\}$ and $\mathrm{abs_c}(\cdot)$ the component-wise absolute value function such that $(\mathbf{y}_{\mathrm{s,inv}})^{\mathrm{T}} \mathrm{abs_c}(\mathbf{y}_1) \neq 0$ for all t. Figure 5 gives the residua for the particular model and fault parameterisation. According to its definition $e_{\mathrm{y},1} = 0$ whereas $e_{\mathrm{y},2}$ and $e_{\mathrm{y},3}$ vary significantly within the working cycle time range.

A hybrid state estimation is set up for the simulated system utilizing the IMM approach. The simulation results are depicted in Fig. 6 for eight working cycles. Altering operation modes (see Fig. 6a) are imposed to the simulated system in different working cycle phases by abruptly changing the fault magnitudes (see Fig. 7a). Note that the performance of the algorithm strongly correlates with the previously described residua. Or, worded differently, the larger the magnitude of $e_{\mathrm{y},i}$, the better is the ability to identify the related fault. Thus, the operation with increased external leakage can be detected less accurate as the corresponding measurable residuum is significantly smaller. The influence of the stochastic noise introduced in the measurement is apparent when comparing the calculated mode probabilities in Fig. 6b, c. In both cases the short response time of the estimation process enables operation mode identification within approximately a fourth of the working cycle. The accuracy of the mode estimation results from the efficiency of the implemented Kalman filters and the small estimation error of the primary differential states $\mathbf{x}_{\mathrm{d}}^{(\cdot)}$.

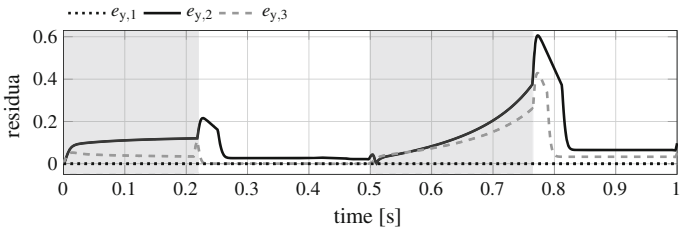


Fig. 5 Simulation results: Measurable residua $e_{\mathrm{y},1}$, $e_{\mathrm{y},2}$, $e_{\mathrm{y},3}$

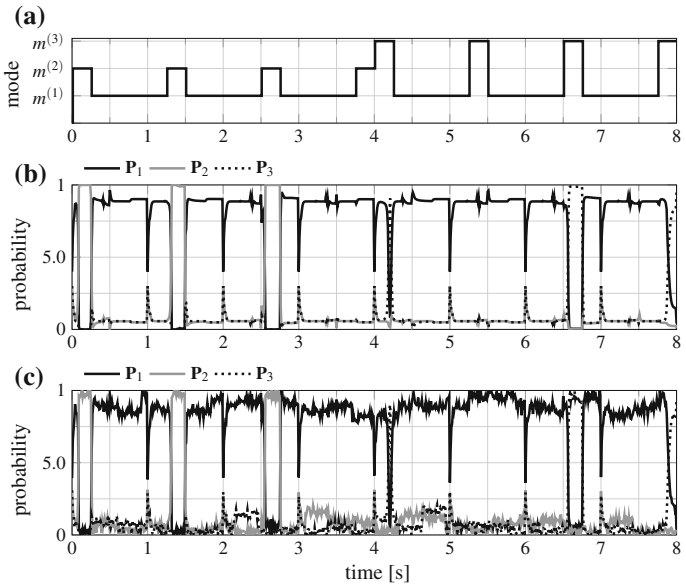


Fig. 6 Mode estimation with the IMM approach: **a** mode changes, **b** mode probabilities of the noise-free system, **c** mode probabilities of the system subject to measurement noise

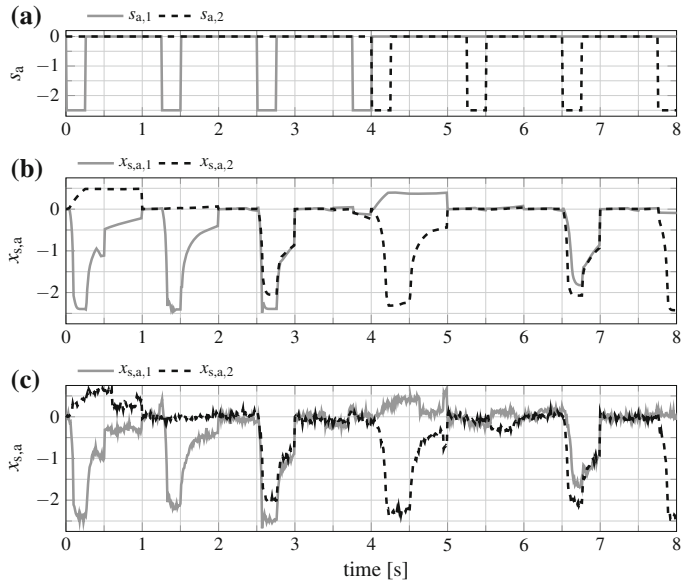


Fig. 7 Magnitude estimation with the IMM approach: **a** modification of the fault magnitude, **b** estimated fault magnitude of the noise-free system, **c** estimated fault magnitude of the system subject to measurement noise

Fault identification, however, does not only require to estimate the mode but also to estimate the magnitude of the fault that is depicted in Fig. 7a. As shown in Fig. 7b, c, a similar characteristic as for the mode estimation arise for the estimation of the fault magnitude. In case of the noise-free system as well as in case of the system subject to measurement noise, the ability to estimate precisely the fault magnitude correlates with the measurable residua. Hence, the estimation of magnitude $s_{\mathrm{a},1}$ is more accurate than that of $s_{\mathrm{a},2}$.

7 Conclusion

This article concerns advanced state estimation and fault identification methods. Particularly, the UKF and the IMM approach, which are well-known for systems given in state-space form, are generalized to nonlinear differential-algebraic equations. Therefore, properties of DAE systems and modelling of systems subject to faults are discussed in detail. The key issue of the state estimation of DAE systems with an UKF is the consideration of the model of the algebraic states within the estimation algorithm. A suitable concept for index-1 DAE systems is the exclusion of the algebraic states from the estimation procedure and its consistent calculation when required. Augmenting this scheme, system supervision including fault identification can be implemented utilizing the IMM approach. The applicability of the proposed approach is illustrated by a numerical example that considers a working cycle of a hydraulic cylinder. Even though the model properties of that system are challenging and the induced faults have varying degrees of impact on the measured system outputs, the mode estimation as well as the magnitude estimation gives reliable results.

References

1. Alkov I, Weidemann D (2013) Fault detection with unscented Kalman filter applied to nonlinear differential-algebraic systems. In: Proceedings of the 18th international conference on methods and models in automation and robotics (MMAR), Miedzyzdroje, Poland, pp 166–171
2. Alkov I, Weidemann D (2014) Unscented Kalman filter for higher index nonlinear differential-algebraic equations. In: Proceedings of the 19th international conference on methods and models in automation and robotics (MMAR), Miedzyzdroje, Poland, pp 88–93
3. Åslund J, Frisk E (2006) An observer for non-linear differential-algebraic systems. Automatica 42(6):959–965
4. Bar-Shalom Y, Li X-R, Kirubarajan T (2001) Estimation with applications to tracking and navigation. Wiley, New York
5. Becerra VM, Roberts PD, Griffiths GW (1999) Dynamic data reconciliation for a class of nonlinear differential-algebraic equation models using the extended Kalman filter. In: Proceedings of the 14th IFAC world congress, Oxford, UK, pp 303–308
6. Blanke M, Kinnaert M, Lunze J, Staroswiecki M (2006) Diagnosis and fault-tolerant control. Springer, Berlin
7. Brenan KE, Petzold LR (1989) The numerical solution of higher index differential/algebraic equations by implicit methods. SIAM J Numer Anal 26(4):976–996

8. Campbell SL, Gear CW (1995) The index of general nonlinear DAEs. Numerische Mathematik 72(2):173–196
9. Chen DW, Saif M (2007) Observer-based strategies for actuator fault detection, isolation and estimation for certain class of nonlinear systems. IET Control Theory Appl 1(6):1672–1680
10. Chen J, Patton RJ (2005) Robust model-based fault diagnosis for dynamic systems. Kluwer Academic Publisher, Boston
11. Deuflhard P, Hairer E, Zugck J (1987) One-step and extrapolation methods for differential-algebraic systems. Numerische Mathematik 51(5):501–516
12. Ding SX (2008) Model-based fault diagnosis techniques. Springer, Berlin
13. Gear C (1988) Differential-algebraic equation index transformations. SIAM J Sci Stat Comput 9(1):39–47
14. Gerdin M, Sjoberg J (2006) Nonlinear stochastic differential-algebraic equations with application to particle filtering. In: Proceedings of the 45th IEEE conference on decision and control (CDC), San Diego, CA, USA, pp 6630–6635
15. Hedengren JD, Shishavan RA, Powell KM, Edgar TF (2014) Nonlinear modeling, estimation and predictive control in APMonitor. Comput Chem Eng 70:133–148
16. Hofbaur MW (2005) Hybrid estimation of complex systems. Springer, Berlin
17. Hofbaur MW, Williams BC (2002) Mode estimation of probabilistic hybrid systems. In: Proceedings on the international conference on hybrid systems: computation and control, Springer, pp 253–266
18. Isermann R (1993) Fault diagnosis of machines via parameter estimation and knowledge processing—tutorial paper. Automatica 29(4):815–835
19. Isermann R (2006) Fault-diagnosis systems. Springer, Berlin
20. Julier S J (2002) The scaled unscented transformation. In: Proceedings of the American control conference (ACC), vol 6, Anchorage, AK, USA, pp 4555–4559
21. Julier S J, Uhlmann J K, Durrant-Whyte H F (1995) A new approach for filtering nonlinear systems. In: Proceedings of the American control conference (ACC), vol 3, Seattle, WA, USA, pp 1628–1632
22. Kunkel P, Mehrmann V, Seufer I (2002) GENDA: A software package for the solution of general nonlinear differential-algebraic equations. Technical report 730-02, Institut für Mathematik, Technische Universität Berlin, 2002
23. Mandela RK, Rengaswamy R, Narasimhan S (2009) Nonlinear state estimation of differential-algebraic systems. Adv Control Chem Process 7:792–797
24. Mattsson SE, Söderlind G (1993) Index reduction in differential-algebraic equations using dummy derivatives. SIAM J Sci Stat Comput 14(3):677–692
25. Mehrmann V (2012) Index concepts for differential-algebraic equations. Technical report 2012-03, Institut für Mathematik, Technische Universität Berlin, 2012
26. Nedialkov NS, Pryce JD (2005) Solving differential-algebraic equations by Taylor series (i): computing Taylor coefficients. BIT Numer Math 45(3):561–591
27. Pantelides CC (1988) The consistent initialization of differential-algebraic systems. SIAM J Sci Stat Comput 9(2):213–231
28. Rao C, Rawlings J, Mayne D (2003) Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. IEEE Trans Autom Control 48:246–258
29. Tan CP, Edwards C (2003) Sliding mode observers for reconstruction of simultaneously actuator and sensor faults. In: Proceedings of the IEEE conference on decision and control, pp 1455–1460
30. van der Merwe R, Wan E A (2001) The square-root unscented Kalman filter for state and parameter-estimation. In: Proceedings of the IEEE international conference on acoustics, speech, and signal processing, vol 6, Salt Lake City, UT, USA, pp 3461–3464
31. Zhang YM, Jiang J (2002) Active fault-tolerant control system against partial actuator failures. IEE Proc Control Theory Appl 149(1):95–104

Towards Robust Fault-Tolerant Model Predictive Control with Constraints for Takagi–Sugeno Systems

Piotr Witczak and Marcin Witczak

Abstract This chapter deals with the problem of a robust fault-tolerant model predictive control (RFT-MPC) for discrete-time nonlinear systems described by Takagi-Sugeno models. The RFT-MPC is a mixture of the $\mathscr{H}_\infty$-based parallel distributed controllers and the fast model Predictive Control. The approach proposed in the paper is based on a series of offline and online computations. For the given Takagi-Sugeno system, PDC is designed without considering input and state constraints. Moreover, the idea of robust invariant sets is employed to deal with both the input and state constraints. This also provides an efficient way to introduce the MPC algorithm. Therefore, enhancing the invariant set enlarges the domain of attraction. As the robustness is achieved offline, the MPC is not employed until large enough faults occur. Otherwise, it serves as a fault-tolerant control distributing any compensation actions between actuators to avoid their saturation if possible. Finally, an illustrative example is provided, proving the efficiency and quality of the proposed multi-stage RFT-MPC.

1 Introduction

A continuous growth of the complexity, efficiency and reliability of industrial systems implies a continuous and rapid development in control and fault diagnosis. An interconnection of these two paradigms is intensively studied under the name of fault-tolerant control (FTC). The fault-tolerant control (FTC) systems are classified into two distinct classes [32, 39]: passive and active. The passive FTC controllers [5] are designed to be robust against a finite set of predefined faults and fault diagnosis is needless. Such a design usually degrades the system performance. Contrary to the passive ones, active FTC [9] schemes react to faults actively by reconfiguring control

P. Witczak (✉) · M. Witczak
Institute of Control and Computation Engineering, University of Zielona Góra, Zielona Góra, Poland
e-mail: p.witczak@weit.uz.zgora.pl

M. Witczak
e-mail: m.witczak@issi.uz.zgora.pl

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*, Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_11

actions in such a way that the system stability and acceptable performance are maintained. To this end, the control system relies on the Fault Detection and Isolation (FDI) [3, 19, 25, 33]. Most of the recent works treat the FDI and FTC problems separately. Unfortunately, a perfect FDI is impossible and hence there always is an inaccuracy related to this process. Thus, there is a need for integrated FDI and FTC schemes for both linear and nonlinear systems [36].

A number of works have been published in the last decade on the FTC phenomena. The book [17], which is mainly devoted to fault diagnosis and its applications, provides some general rules for the hardware-redundancy-based FTC. In contrast, the work [24] introduces the concepts of the active and passive FTC. It also goes through the problem of performance and stability of the FTC under imperfect fault diagnosis. The authors consider (under a chain of some, not necessary easy to satisfy assumptions) the effect of a delayed fault detection and an imperfect fault identification, but the fault diagnosis [20, 34] scheme is treated separately during the design and no practical integration of the fault diagnosis and the FTC is proposed. The FTC and FDI are also treated in the work [32, 37] where the author treats both problems simultaneously under one framework. Hence, in this work a similar integrated approach is proposed.

The proposed approach deals with the above-mentioned difficulties by providing an elegant way of incorporating fault diagnosis (particularly fault identification) into the fault-tolerant control framework. A multi-stage procedure is proposed, which begins with fault estimation, then the fault is compensated with a robust controller. The robust controller is designed without taking into account the input constraints related to the actuator saturation. Hence, the robust invariant set is employed, which takes into account the input constraints. If the robust invariant set does not cover the current state, then suitable predictive control actions are performed in order to enhance the invariant set. This appealing property makes it possible to enlarge the domain of attraction, making the proposed approach an efficient solution. Indeed, the presented solution can be perceived as an extension of the recent developments in this area [35], which shows a fault estimation and compensation strategy for nonlinear systems. The novelty of the scheme consists of:

- introduction of robustness to disturbances, through the $\mathscr{H}_\infty$ approach,
- introduction of the multi-stage procedure: fault estimation, fault compensation with robust controller and predictive control enhancing the applicability of the approach,
- extension of the work of [22] to the case with exogenous disturbances,
- incorporating a robust invariant set to cover $\mathscr{H}_\infty$ controller design imperfections.

To fully present the reasoning path, the chapter is organized as follows. Section 2 presents preliminaries regarding the problem taken under consideration. Robust fault estimation and control approach is proposed in Sect. 3. Furthermore, Sect. 4 presents the idea of a robust invariant set while Sect. 5 depicts an efficient robust predictive fault-tolerant control strategy. The final part of the chapter contains the illustrative example. For the readers convenience in a symbol list appears in the appendix.

2 A General Description of the Fault-Tolerant Scheme

A nonlinear dynamic system can be described in a simple way by a Takagi-Sugeno (TS) fuzzy model, which uses the sequence of linear models approximating nonlinear behaviour in predefined regions, a parameter identification of an a-priori given structure or transformation of a nonlinear model using the nonlinear sector approach (see, e.g. [19, 27, 28]). According to this model, a nonlinear dynamic systems can be linearized around a number of operating points. The fuzzy fusion of all linear model outputs describes the global system behaviour. A Takagi-Sugeno model is described by fuzzy IF-THEN rules. The presented structure may represent a nonlinear system with a control-affine state equation. It has a rule base of M rules, each having p antecedents, where ith rule is expressed as

$$\begin{aligned} R^i &: \text{ IF } s_k^1 \text{ is } F_1^i \text{ and } \ldots \text{ and } s_k^p \text{ is } F_p^i, \\ \text{THEN } & \boldsymbol{x}_{f,k+1} = \boldsymbol{A}^i \boldsymbol{x}_{f,k} + \boldsymbol{B}^i \boldsymbol{u}_{f,k} + \boldsymbol{B}^i \boldsymbol{f}_k + \boldsymbol{W}^i \boldsymbol{w}_k, \end{aligned} \tag{1}$$

where $\boldsymbol{x}_{f,k} \in \mathbb{R}^n$ stands for the state and $\boldsymbol{u}_{f,k} \in \mathbb{R}^r$ denotes the nominal control input, $\boldsymbol{f}_k \in \mathbb{R}^r$ is the actuator fault, $i = 1, \ldots, M$, F_j^i $(j = 1, \ldots, p)$ are fuzzy sets and $\boldsymbol{s}_k = [s_k^1, s_k^2, \ldots, s_k^p]$ is a known vector of premise variables [19, 27]. It is of course possible to employ the approach that can be dedicated for the case when some of the premise variables are unmeasurable [2, 15]. Similarly, the approach can be extended to the case when only some state variables are available, but this leads to the need for employing observer-based schemes [35]. Although useful, for the sake of presentation simplicity, the above-mentioned strategies are not included the scope of this work; also, this can lead to more conservative solutions. Additionally, $wk \in l_2$ is an exogenous disturbance vector, while:

$$l_2 = \left\{ \mathbf{w} \in \mathbb{R}^n | \ \|\mathbf{w}\|_{l_2} < +\infty \right\}, \quad \|\mathbf{w}\|_{l_2} = \left(\sum_{k=0}^{\infty} \|\boldsymbol{w}_k\|^2 \right)^{\frac{1}{2}}. \tag{2}$$

It is also assumed that the control limits are given in form:

$$-\bar{\boldsymbol{u}}_i \leq \boldsymbol{u}_{i,k} \leq \bar{\boldsymbol{u}}_i, \quad i = 1, \ldots, r. \tag{3}$$

where $\bar{\boldsymbol{u}}_i > 0, i = 1, \ldots, r$ are particular control limits. To maintain the simplicity of further deliberations, these limits are symmetrical around zero but with an appropriate scaling it is relatively easy to introduce non-symmetrical ones.

Given $(\boldsymbol{s}_k, \boldsymbol{u}_k)$ and a product inference engine, the final output of the normalized T-S fuzzy model can be inferred as:

$$\begin{aligned} \boldsymbol{x}_{f,k+1} &= \sum_{i=1}^{M} h_i(\boldsymbol{s}_k)[\boldsymbol{A}^i \boldsymbol{x}_{f,k} + \boldsymbol{B}^i \boldsymbol{u}_k + \boldsymbol{B}^i \boldsymbol{f}_k + \boldsymbol{W}^i \boldsymbol{w}_k], \\ &= \boldsymbol{A}(h_k)\boldsymbol{x}_{f,k} + \boldsymbol{B}(h_k)\boldsymbol{u}_k + \boldsymbol{B}(h_k)\boldsymbol{f}_k + \boldsymbol{W}(h_k)\boldsymbol{w}_k, \end{aligned} \tag{4}$$

where $h_i(\boldsymbol{s}_k)$ are normalized rule firing strengths defined as

$$h_i(\boldsymbol{s}_k) = \frac{\mathscr{T}_{j=1}^{p} \mu_{F_j^i}(s_k^j)}{\sum_{i=1}^{M} (\mathscr{T}_{j=1}^{p} \mu_{F_j^i}(s_k^j))}, \tag{5}$$

and $\mathscr{T}$ denotes a t-norm (e.g. product). The term $\mu_{F_j^i}(s_k^j)$ is the grade of membership of the premise variable s_k^j. Additionally, the rule firing strengths $h_i(\boldsymbol{s}_k)$ $(i = 1, \ldots, M)$ satisfy the following constraints

$$\begin{cases} \sum_{i=1}^{M} h_i(\boldsymbol{s}_k) = 1, \\ 0 \leqslant h_i(\boldsymbol{s}_k) \leqslant 1, \qquad \forall i = 1, \ldots, M. \end{cases} \tag{6}$$

The main objective of the subsequent part of this contribution is to design the control strategy in such a way that the system (4) will converge to the origin irrespective of the presence of the fault $\boldsymbol{f}_k$. The proposed control scheme is as follows:

$$\boldsymbol{u}_{f,k} = -\boldsymbol{K}(h_k)\boldsymbol{x}_k - \hat{\boldsymbol{f}}_{k-1} + \boldsymbol{c}_k, \tag{7}$$

while the predicted future input is described by:

$$\boldsymbol{u}_{f,j} = \begin{cases} -\boldsymbol{K}(h_j)\boldsymbol{x}_j - \hat{\boldsymbol{f}}_{k-1} + \boldsymbol{c}_j, & j = k, \ldots, k + n_c - 1, \\ -\boldsymbol{K}(h_j)\boldsymbol{x}_j - \hat{\boldsymbol{f}}_{k-1}, & j \geq k + n_c. \end{cases} \tag{8}$$

where n_c stands for the prediction horizon, $\boldsymbol{K}(h_k) = \sum_{i=1}^{M} h_i(\boldsymbol{s}_k)\boldsymbol{K}^i$ is the $\mathscr{H}_\infty$ Parallel Distributed Compensation (PDC) controller designed to achieve robustness with respect to exogenous disturbances $\boldsymbol{w}_k$, $\hat{\boldsymbol{f}}_{k-1}$ is the fault estimate, which compensates the effect of a fault, $\boldsymbol{c}_j$ is a vector introducing additional design freedom, which should be exploited when the fault compensation does not provide the expected results due to the actuator saturation.

Alternatively in terms of online and offline layers of the proposed approach, the procedure can be described as

Offline computation:

Step 1. for a predefined disturbance attenuation level $\mu > 0$, design a robust controller $\boldsymbol{K}(h_k)$ by solving (28),

Step 2. determine the robust invariant set by solving (70) under the constraints (61) and (69).

Online computation: for each k,

Step 1. compute the fault estimate $\hat{\boldsymbol{f}}_{k-1}$ with (11),
Step 2. solve the optimisation problem (73),
Step 3. implement the first element of $\boldsymbol{\delta}_k$, i.e. $\boldsymbol{c}_k$.

An outline of the proposed scheme can be therefore depicted as in Fig. 1.

Note that beyond the prediction horizon, n_c, $\boldsymbol{c}_j$ is set to zero, which denotes the feasibility of the $\mathscr{H}_\infty$ control. Thus, the design of the proposed control strategy boils down to solving the following problems:

- to design a robust PDC controller $\boldsymbol{K}(h_k)$ in such way that a prescribed disturbance attenuation level is achieved with respect to $\boldsymbol{x}_{f,k}$ while guaranteeing its convergence to the origin,
- to estimate the fault $\boldsymbol{f}_k$,
- to determine a set of states for which the robust PDC controller along with the fault compensation (under the control constraints) is feasible,
- to determine $\boldsymbol{c}_j$ in such a way as to enhance a set of states and, hence making the control problem feasible.

Since the general scheme is given, the remaining part of this chapter is devoted to solving the above-mentioned design problems.

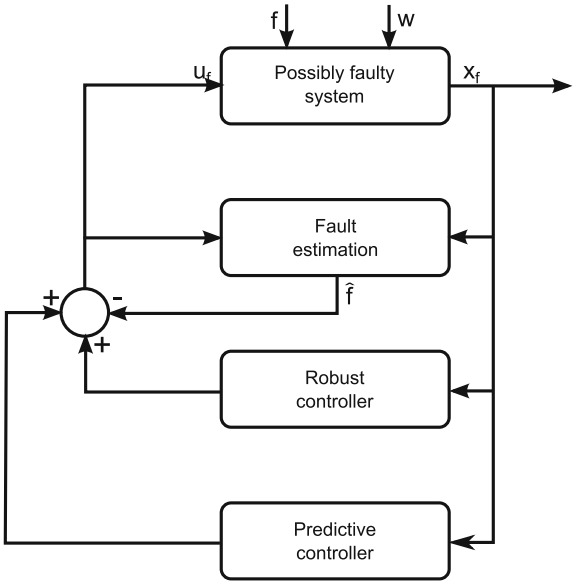


Fig. 1 Proposed robust predictive FTC scheme outline

3 Fault Estimation and Robust Control

In this section, a fault estimation technique will be proposed, which along with the robust PDC controller $\boldsymbol{K}(h_k)$ will be used to compensate the effect of a fault and feed the system in such a way that the state $\boldsymbol{x}_{f,k}$ goes to the origin. Note that the designs of the fault estimator and the robust PDC controller are realized for the unconstrained case. Moreover, the free control parameter $\boldsymbol{c}_j$ (cf. (8)) is set to zero. Following the seminal paper [22] along with further developments, the constraints are introduced during the development of the set of states, for which the robust PDC controller along with the fault compensation is feasible as well as during the computation of $\boldsymbol{c}_k$, which enhance a set of states, hence making the control problem feasible.

Thus, following [13, 31], by computing

$$\boldsymbol{H}(h_k) = \boldsymbol{B}(h_k)^{+} = \left[\boldsymbol{B}(h_k)^T \boldsymbol{B}(h_k)\right]^{-1} \boldsymbol{B}(h_k)^T, \tag{9}$$

and then multiplying (4) by $\boldsymbol{H}(h_k)$ along with extracting $\boldsymbol{f}_k$, it can be shown that:

$$\boldsymbol{f}_k = \boldsymbol{H}(h_k)\boldsymbol{x}_{f,k+1} - \boldsymbol{H}(h_k)\boldsymbol{A}(h_k)\boldsymbol{x}_{f,k} - \boldsymbol{u}_{f,k} - \boldsymbol{H}(h_k)\boldsymbol{W}(h_k)\boldsymbol{w}_k, \tag{10}$$

while its estimate can be given as:

$$\hat{\boldsymbol{f}}_k = \boldsymbol{H}(h_k)\boldsymbol{x}_{f,k+1} - \boldsymbol{H}(h_k)\boldsymbol{A}(h_k)\boldsymbol{x}_{f,k} - \boldsymbol{u}_{f,k}, \tag{11}$$

with the associated fault estimation error

$$\boldsymbol{\varepsilon}_{f,k} = \boldsymbol{f}_k - \hat{\boldsymbol{f}}_k = -\boldsymbol{H}(h_k)\boldsymbol{W}(h_k)\boldsymbol{w}_k. \tag{12}$$

The main difficulty of the above approach is related to the existence of $\boldsymbol{H}(h_k)$, which boils down to checking the full rank property of all convex combinations of $\boldsymbol{B}^i$, $i = 1, \ldots, M$. Clearly, in a general case, $\boldsymbol{H}(h_k)$ does not fulfil this condition. Thus, to settle this problem within the framework of this chapter, the following idea is employed [18] and it is assumed that the matrix $\boldsymbol{B}^M$ has full rank. Note that it is enough that for some i, $1 \leq i \leq M$, $\operatorname{rank}(\boldsymbol{B}^i) = r$ and then the order of the matrices can be rearranged to satisfy $\operatorname{rank}(\boldsymbol{B}^M) = r$. Let us define

$$\boldsymbol{Q}_{p,p} = (\boldsymbol{B}^p)^T \boldsymbol{B}^p, \qquad p = 1, \ldots M \tag{13}$$

$$\boldsymbol{Q}_{p,a} = (\boldsymbol{B}^p)^T \boldsymbol{B}^a + (\boldsymbol{B}^a)^T \boldsymbol{B}^p - (\boldsymbol{B}^a)^T \boldsymbol{B}^a - (\boldsymbol{B}^p)^T \boldsymbol{B}^p \qquad \text{for } p < a \tag{14}$$

$$\boldsymbol{R}_p = \boldsymbol{R}_{a,b}^p = \begin{cases} \boldsymbol{Q}_{p,p} & \text{if } (a,b) = (1,1) \\ \boldsymbol{Q}_{b-1,p} & \text{if } a = 1 \wedge b = 2, \ldots, p \\ \boldsymbol{I} & \text{if } a = b \wedge 1 < b < p \\ -\boldsymbol{I} & \text{if } b = 1 \wedge a = p+1 \\ \boldsymbol{0} & \text{otherwise} \end{cases} \tag{15}$$

Theorem 3.1 *The following statements are equivalent*

(a) All convex combinations of $\boldsymbol{B}^1, \ldots, \boldsymbol{B}^M$ *have full rank.*
(b) $\boldsymbol{B}^M$ *has full row rank and the* $(M-1)Mn$*-by-*$(M-1)Mn$ *matrix*

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{R}_1\boldsymbol{R}_M^{-1} & (\boldsymbol{R}_2-\boldsymbol{R}_1)\boldsymbol{R}_M^{-1} & (\boldsymbol{R}_3-\boldsymbol{R}_2)\boldsymbol{R}_M^{-1} & \ldots & (\boldsymbol{R}_{M-1}-\boldsymbol{R}_{M-2})\boldsymbol{R}_M^{-1} \\ -\boldsymbol{I}_{Mn} & \boldsymbol{I}_{Mn} & \boldsymbol{0}_{Mn} & \ldots & \boldsymbol{0}_{Mn} \\ \boldsymbol{0}_{Mn} & -\boldsymbol{I}_{Mn} & \boldsymbol{I}_{Mn} & \ldots & \boldsymbol{0}_{Mn} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \boldsymbol{0}_{Mn} & \ldots & \boldsymbol{0}_{Mn} & -\boldsymbol{I}_{Mn} & \boldsymbol{I}_{Mn} \end{bmatrix}, \tag{16}$$

is a block P-matrix [18] with respect to the partition $\{\boldsymbol{F}_1, \ldots, \boldsymbol{F}_{M-1}\}$ *of* $\{1, \ldots, (M-1)Mn\}$*, with* $\boldsymbol{F}_i = \{(M-1)Mn+1, \ldots, iMn\}$*,* $i = 1, \ldots, M-1$*.*

Proof The proof can be derived by a direct application of *Theorem 2* in [18].

Remark 1 Following [18], the matrix $\boldsymbol{V}$ is a block P-matrix with respect to any partition if all its principal minors are positive. This feature makes it possible to easily check the condition of Theorem 3.1.

Since the general framework for computing the fault estimate (11) is given, its computational feasibility can be verified. Indeed, to obtain $\hat{\boldsymbol{f}}_k$ it is necessary to have $\boldsymbol{x}_{f,k+1}$. Thus, the only choice to compensate $\boldsymbol{f}_k$ in (4) is to use $\hat{\boldsymbol{f}}_{k-1}$. This determines the above-proposed control strategy

$$\boldsymbol{u}_{f,k} = -\hat{\boldsymbol{f}}_{k-1} - \boldsymbol{K}(h_k)\boldsymbol{x}_{f,k}. \tag{17}$$

Note that this strategy is derived by setting $\boldsymbol{c}_k = \boldsymbol{0}$ in (7). Bearing in mind that in any physical system $\boldsymbol{f}_k$ is bounded, without a loss of generality, it is possible to write

$$\hat{\boldsymbol{f}}_k = \hat{\boldsymbol{f}}_{k-1} + v_k, \quad v_k \in l_2. \tag{18}$$

For the purpose of further analysis, it is additionally assumed that the above bounds have the following form

$$v_k \in \mathscr{E}_v, \quad \mathscr{E}_v = \{v \in \mathbb{R}^r \,|\, v^T \boldsymbol{Q} v \le 1\}, \quad \boldsymbol{Q} \succ 0. \tag{19}$$

Thus, (17) can be written in an equivalent form, which will be used for further deliberations

$$\boldsymbol{u}_{f,k} = -\hat{\boldsymbol{f}}_k + v_k - \boldsymbol{K}(h_k)\boldsymbol{x}_{f,k}. \tag{20}$$

Substituting (20) into (4) gives

$$\boldsymbol{x}_{f,k+1} = \boldsymbol{A}_1(h_k)\boldsymbol{x}_{f,k} + [\boldsymbol{I} - \boldsymbol{B}(h_k)\boldsymbol{H}(h_k)]\,\boldsymbol{W}(h_k)\boldsymbol{w}_k + \boldsymbol{B}(h_k)v_k, \tag{21}$$

with $\boldsymbol{A}_1(h_k) = \sum_{i=1}^{M}\sum_{j=1}^{M} h_i(\boldsymbol{s}_k)h_j(\boldsymbol{s}_k)(\boldsymbol{A}^i - \boldsymbol{B}^i\boldsymbol{K}^j)$.

Further analysis of (21), and in particular

$$\left[\boldsymbol{I} - \boldsymbol{B}(h_k)\left[\boldsymbol{B}(h_k)^T\boldsymbol{B}(h_k)\right]^{-1}\boldsymbol{B}(h_k)^T\right]\boldsymbol{W}(h_k)\boldsymbol{w}_k = \boldsymbol{W}(h_k)\boldsymbol{w}_k - \boldsymbol{B}(h_k)\left[\boldsymbol{B}(h_k)^T\boldsymbol{B}(h_k)\right]^{-1}\boldsymbol{B}(h_k)^T\boldsymbol{W}(h_k)\boldsymbol{w}_k, \tag{22}$$

along with the fact that any vector $\boldsymbol{W}(h_k)\boldsymbol{w}_k \in \mathrm{col}(\boldsymbol{B}(h_k))$, where $\mathrm{col}(\boldsymbol{B}(h_k)) = \{\boldsymbol{\alpha} \in \mathbb{R}^n : \boldsymbol{\alpha} = \boldsymbol{B}(h_k)\boldsymbol{\beta}\}$ for some $\boldsymbol{\beta} \in \mathbb{R}^r$ can be written as $\boldsymbol{W}(h_k)\boldsymbol{w}_k = \boldsymbol{B}(h_k)\bar{\boldsymbol{w}}_k$ for some non-zero $\bar{\boldsymbol{w}}(k)$, leads (22) to

$$\boldsymbol{B}(h_k)\bar{\boldsymbol{w}}_k - \boldsymbol{B}(h_k)\left[\boldsymbol{B}(h_k)^T\boldsymbol{B}(h_k)\right]^{-1}\boldsymbol{B}(h_k)^T\boldsymbol{B}(h_k)\bar{\boldsymbol{w}}_k = 0 \tag{23}$$

This significant simplification of (21) yields its new form:

$$\boldsymbol{x}_{f,k+1} = \boldsymbol{A}_1(h_k)\boldsymbol{x}_{f,k} + \boldsymbol{B}(h_k)v_k \tag{24}$$

Note: It is worth to see, if (23) does not hold, the (24) will take another form, including $Ww(h_k)$ which then, needs to be known.

Before providing the PDC control design procedure, let us recall the following lemma [4, 6, 7]:

Lemma 1 *The following statements are equivalent*

1. *There exists* $\boldsymbol{X} \succ 0$ *such that*

$$\boldsymbol{V}^T\boldsymbol{X}\boldsymbol{V} - \boldsymbol{W} \prec 0 \tag{25}$$

2. *There exists* $\boldsymbol{X} \succ 0$ *such that*

$$\begin{bmatrix} -\boldsymbol{W} & \boldsymbol{V}^T\boldsymbol{U}^T \\ \boldsymbol{U}\boldsymbol{V} & \boldsymbol{X} - \boldsymbol{U} - \boldsymbol{U}^T \end{bmatrix} \prec 0. \tag{26}$$

Remark 2 Note that the regularity of $\boldsymbol{U}$ is ensured by the last block diagonal element of (26), which implies $\boldsymbol{U} + \boldsymbol{U}^T \succ \boldsymbol{X} \succ \boldsymbol{0}$. This property will be exploited in further deliberations.

The following theorem constitutes the main result of this section.

Theorem 3.2 *For a prescribed disturbance attenuation level* $\mu > 0$ *for the* $\boldsymbol{x}_{f,k}$*, the* $\mathscr{H}_\infty$ *controller design problem for the system (4) is solvable if there exist* $\boldsymbol{U}$*,* $\boldsymbol{N}^i$ *and* $\boldsymbol{P}^i \succ \boldsymbol{0}$ *(*$i = 1, \ldots, M$*) such that the following condition is satisfied:*

$$\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{l=1}^{M} h_i(\boldsymbol{s}_k)h_j(\boldsymbol{s}_k)h_l(\boldsymbol{s}_{k+1})\Upsilon_{i,j}^{l} \prec \boldsymbol{0}, \tag{27}$$

where

$$\Upsilon_{i,j}^{l} = \begin{bmatrix} -\boldsymbol{P}^{i} & \boldsymbol{0} & \boldsymbol{U}^{T}\boldsymbol{A}^{(i)T} - \boldsymbol{N}^{(j)T}\boldsymbol{B}^{(i)T} & \boldsymbol{U}^{T} \\ \boldsymbol{0} & -\mu^{2}\boldsymbol{I}_{r} & \boldsymbol{B}^{(i)T} & \boldsymbol{0} \\ \boldsymbol{A}^{i}\boldsymbol{U} - \boldsymbol{B}^{i}\boldsymbol{N}^{j} & \boldsymbol{B}^{i} & \boldsymbol{P}^{l} - \boldsymbol{U} - \boldsymbol{U}^{T} & \boldsymbol{0} \\ \boldsymbol{U} & \boldsymbol{0} & \boldsymbol{0} & -\boldsymbol{I} \end{bmatrix} \tag{28}$$

with $\boldsymbol{N}^{j} = \boldsymbol{K}^{j}\boldsymbol{U}$.

Proof The problem of $\mathscr{H}_{\infty}$ controller design (cf. [23, 38]) is to determine the gain matrix $\boldsymbol{K}(h_k)$ such that

$$\lim_{k\to\infty} \boldsymbol{x}_{f,k} = \boldsymbol{0} \quad \text{for } v_k = \boldsymbol{0} \tag{29}$$

$$\|\boldsymbol{x}_{f}\|_{l_2} \leq \mu \|v_k\|_{l_2} \quad \text{for } v_k \neq \boldsymbol{0},\ \boldsymbol{e}_0 = \boldsymbol{0}. \tag{30}$$

In order to settle the above problem it is sufficient to find a Lyapunov function V_k such that:

$$\Delta V_k + \boldsymbol{x}_{f,k}^{T}\boldsymbol{x}_{f,k} - \mu^{2} v_k^{T} v_k < 0,\ k = 0, 1, \ldots, \tag{31}$$

where $\Delta V_k = V_{k+1} - V_k$. Indeed, if $v_k = \boldsymbol{0}$ then (31) boils down to

$$\Delta V_k + \boldsymbol{x}_{f,k}^{T}\boldsymbol{x}_{f,k} < 0,\ k = 0, 1, \ldots, \tag{32}$$

and hence $\Delta V_k < 0$, which leads to (29). If $v_k \neq \boldsymbol{0}$ then (31) yields

$$J = \sum_{k=0}^{\infty} \left(\Delta V_k + \boldsymbol{x}_{f,k}^{T}\boldsymbol{x}_{f,k} - \mu^{2} v_k^{T} v_k\right) < 0, \tag{33}$$

which can be written as

$$J = -V_0 + \sum_{k=0}^{\infty} \boldsymbol{x}_{f,k}^{T}\boldsymbol{x}_{f,k} - \sum_{k=0}^{\infty} \mu^{2} v_k^{T} v_k < 0, \tag{34}$$

Knowing that $V_0 = 0$ for $\boldsymbol{x}_{f,0} = \boldsymbol{0}$, (34) leads to (30).

Selecting the Lyapunov function as (cf. Remark 2)

$$V_k = \boldsymbol{x}_{f,k}^{T}\boldsymbol{U}^{-T}\boldsymbol{P}(h_k)\boldsymbol{U}^{-1}\boldsymbol{x}_{f,k}, \tag{35}$$

where

$$\boldsymbol{P}(h_k) = \sum_{i=1}^{M} h_i(\boldsymbol{s}_k)\boldsymbol{P}^{i}, \tag{36}$$

the inequality (31) is

$$\Delta V + \boldsymbol{x}_{f,k}^T \boldsymbol{x}_{f,k} - \mu^2 v_k^T v_k < 0, \tag{37}$$

with

$$\begin{aligned}\Delta V = V_{k+1} - V_k = \boldsymbol{x}_{f,k}^T &\left[\boldsymbol{A}_1(h_k)^T \boldsymbol{U}^{-T} \boldsymbol{P}(h_{k+1}) \boldsymbol{U}^{-1} \boldsymbol{A}_1(h_k) - \boldsymbol{U}^{-T} \boldsymbol{P}(h_k) \boldsymbol{U}^{-1}\right] \boldsymbol{x}_{f,k} \\ &+ \boldsymbol{x}_{f,k}^T \left[\boldsymbol{A}_1(h_k)^T \boldsymbol{U}^{-T} \boldsymbol{P}(h_{k+1}) \boldsymbol{U}^{-1} \boldsymbol{B}(h_k)\right] v_k \\ &+ v_k^T \left[\boldsymbol{B}^T(h_k) \boldsymbol{U}^{-T} \boldsymbol{P}(h_{k+1}) \boldsymbol{U}^{-1} \boldsymbol{A}_1(h_k)\right] \boldsymbol{x}_{f,k} \\ &+ v_k^T \left[\boldsymbol{B}(h_k)^T \boldsymbol{U}^{-T} \boldsymbol{P}(h_{k+1}) \boldsymbol{U}^{-1} \boldsymbol{B}(h_k)\right] v_k. \end{aligned} \tag{38}$$

Note that by Rayleigh quotient and Remark 2:

$$\underline{\alpha} \leq \lambda_i(\boldsymbol{U}^T \boldsymbol{U}) \leq \overline{\alpha}, \quad \underline{\beta} \leq \lambda_i(P) \leq \overline{\beta} \quad i = 1, \ldots, n,$$

where $\lambda(\cdot)$ stands for an eigenvalue of its argument. This implies that

$$\overline{\alpha}\underline{\beta} \boldsymbol{x}_{f,k}^T \boldsymbol{x}_{f,k} \leq V_k \leq \underline{\alpha}\overline{\beta} \boldsymbol{x}_{f,k}^T \boldsymbol{x}_{f,k},$$

which clearly indicated that V_k is a proper Lyapunov candidate matrix. Thus, it can be shown that (37) is equivalent to

$$\begin{bmatrix} \boldsymbol{A}_1(h_k)^T \boldsymbol{U}^{-T} \boldsymbol{P}(h_{k+1}) \boldsymbol{U}^{-1} \boldsymbol{A}_1(h_k) & \boldsymbol{A}_1(h_k)^T \boldsymbol{U}^{-T} \boldsymbol{P}(h_{k+1}) \boldsymbol{U}^{-1} \boldsymbol{B}(h_k) \\ +\boldsymbol{I}_n - \boldsymbol{U}^{-T} \boldsymbol{P}(h_k) \boldsymbol{U}^{-1} & \\ \boldsymbol{B}(h_k)^T \boldsymbol{U}^{-T} \boldsymbol{P}(h_{k+1}) \boldsymbol{U}^{-1} \boldsymbol{A}_1(h_k) & \boldsymbol{B}(h_k)^T \boldsymbol{U}^{-T} \boldsymbol{P}(h_{k+1}) \boldsymbol{U}^{-1} \boldsymbol{B}(h_k) - \mu^2 \boldsymbol{I}_r \end{bmatrix} \prec \boldsymbol{0}. \tag{39}$$

Pre- and post-multiplying it by $\mathrm{diag}(\boldsymbol{U}^T, \boldsymbol{I}_r)$ and $\mathrm{diag}(\boldsymbol{U}, \boldsymbol{I}_r)$ respectively, gives

$$\begin{aligned}&\begin{bmatrix} \boldsymbol{U}^T \boldsymbol{A}_1(h_k)^T \boldsymbol{U}^{-T} \\ \boldsymbol{B}(h_k)^T \boldsymbol{U}^{-T} \end{bmatrix} \boldsymbol{P}(h_{k+1}) \begin{bmatrix} \boldsymbol{U}^{-1} \boldsymbol{A}_1(h_k) \boldsymbol{U} & \boldsymbol{U}^{-1} \boldsymbol{B}(h_k) \end{bmatrix} \\ &+ \begin{bmatrix} \boldsymbol{U}^T \boldsymbol{U} - \boldsymbol{P}(h_k) & \boldsymbol{0} \\ \boldsymbol{0} & -\mu^2 \boldsymbol{I}_r \end{bmatrix} \prec \boldsymbol{0}. \end{aligned} \tag{40}$$

Applying Lemma 1 to (40) yields

$$\begin{bmatrix} \boldsymbol{U}^T\boldsymbol{U} - \boldsymbol{P}(h_k) & \boldsymbol{0} & \boldsymbol{U}^T\boldsymbol{A}_1(h_k)^T \\ \boldsymbol{0} & -\mu^2\boldsymbol{I}_r & \boldsymbol{B}(h_k)^T \\ \boldsymbol{A}_1(h_k)\boldsymbol{U} & \boldsymbol{B}(h_k) & \boldsymbol{P}(h_{k+1}) - \boldsymbol{U} - \boldsymbol{U}^T \end{bmatrix} \prec \boldsymbol{0} \Leftrightarrow$$

$$\begin{bmatrix} \boldsymbol{U}^T \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{I}_n \begin{bmatrix} \boldsymbol{U} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} -\boldsymbol{P}(h_k) & \boldsymbol{0} & \boldsymbol{U}^T\boldsymbol{A}_1(h_k)^T \\ \boldsymbol{0} & -\mu^2\boldsymbol{I}_r & \boldsymbol{B}(h_k)^T \\ \boldsymbol{A}_1(h_k)\boldsymbol{U} & \boldsymbol{B}(h_k) & \boldsymbol{P}(h_{k+1}) - \boldsymbol{U} - \boldsymbol{U}^T \end{bmatrix} \prec \boldsymbol{0}, \tag{41}$$

which by Schur complements leads to

$$\begin{bmatrix} -\boldsymbol{P}(h_k) & \boldsymbol{0} & \boldsymbol{U}^T\boldsymbol{A}_1(h_k)^T & \boldsymbol{U}^T \\ \boldsymbol{0} & -\mu^2\boldsymbol{I}_r & \boldsymbol{B}(h_k)^T & \boldsymbol{0} \\ \boldsymbol{A}_1(h_k)\boldsymbol{U} & \boldsymbol{B}(h_k) & \boldsymbol{P}(h_{k+1}) - \boldsymbol{U} - \boldsymbol{U}^T & \boldsymbol{0} \\ \boldsymbol{U} & \boldsymbol{0} & \boldsymbol{0} & -\boldsymbol{I} \end{bmatrix} \prec \boldsymbol{0}, \tag{42}$$

Finally, substituting $\boldsymbol{A}_1(h_k)\boldsymbol{U} = \sum_{i=1}^{M}\sum_{j=1}^{M} h_i(\boldsymbol{s}_k)h_j(\boldsymbol{s}_k)(\boldsymbol{A}^i\boldsymbol{U} - \boldsymbol{B}^i\boldsymbol{K}^j\boldsymbol{U}) = \sum_{i=1}^{M}\sum_{j=1}^{M} h_i(\boldsymbol{s}_k)h_j(\boldsymbol{s}_k)(\boldsymbol{A}^i\boldsymbol{U} - \boldsymbol{B}^i\boldsymbol{N}^j)$ gives (27), which completes the proof.

Note that (27) requires further relaxation procedure to be efficiently solved within Linear Matrix Inequalities (LMI) framework. A sufficient solution to this problem can be found in [30] and with some improvements in [14]. The conditions provided by [29] lead to a decent compromise between complexity and conservatism. In the case (27) it leads to:

Lemma 2 *Condition* (27) *is fulfilled if:*

$$\Upsilon_{i,i}^{l} \prec \boldsymbol{0}, \quad i = 1, \ldots, M, \tag{43}$$

$$\frac{2}{M-1}\Upsilon_{i,i}^{l} + \Upsilon_{i,j}^{l} + \Upsilon_{j,i}^{l} \prec \boldsymbol{0}, \quad i, j, l = 1, \ldots, M, \quad i \neq j. \tag{44}$$

Finally, the design procedure boils down to solving (43)–(44) with respect to $\boldsymbol{U}$, $\boldsymbol{N}^j$ and $\boldsymbol{P}^i$ $(i = 1, \ldots, M,\ j = 1, \ldots, M)$, and then calculating

$$\boldsymbol{K}^j = \boldsymbol{N}^j\boldsymbol{U}^{-1}, \quad j = 1, \ldots, M \tag{45}$$

The objective of this section was to provide a fault estimation and compensation scheme without taking into the account the control limit. The subsequent section provides a useful description of the invariant set, which takes into account the input constraints, while the Sect. 5 presents an online optimization strategy that can be used for enlarging this set.

4 Derivation of a Robust Invariant Set

As it was mentioned in the previous section, in order to maintain a desired system behaviour, the idea of a robust invariant set of state variables is to be employed [10, 40]. To settle this problem the Quadratic Boundedness (QB) [1] will be recalled along with its further extension called Extended Non-Quadratic Boundedness (EQNB) [41].

Let us assume that $\boldsymbol{P}(h_k) = \boldsymbol{P}$ in, $\boldsymbol{P} \succ \boldsymbol{0}$, which makes it possible to formulate the following definitions (cf. [1]):

Definition 1 The system (24) is strictly quadratically bounded with $\boldsymbol{P} \succ \boldsymbol{0}$ for all allowable $v_k \in \mathscr{E}_v$, $k \geq 0$, if $V_k > 1$ implies $V_{k+1} - V_k < 0$ for any $v_k \in \mathscr{E}_v$.

Definition 2 A set $\mathscr{E}_{\boldsymbol{x}_f}$ is a robust invariant set for the system (24) for all allowable $v_k \in \mathscr{E}_v$ if $\boldsymbol{x}_{f,k} \in \mathscr{E}_{\boldsymbol{x}_f}$ implies $\boldsymbol{x}_{f,k+1} \in \mathscr{E}_{\boldsymbol{x}_f}$, for any $v_k \in \mathscr{E}_v$.

In this section, the ellipsoidal bounding will be used for describing the robust invariant set, i.e.

$$\mathscr{E}_{\boldsymbol{x}_f} = \{\boldsymbol{x}_f \in \mathbb{R}^n \,|\, \boldsymbol{x}_f^T \boldsymbol{P} \boldsymbol{x}_f \leq 1\}. \tag{46}$$

The proposed ellipsoidal bounding strategy can be perceived as an inner bounding of the exact invariant set [12]. An obvious drawback to the proposed approach is that the obtained set is smaller than the exact one. However, the simplicity of the ellipsoidal description will make it possible to use it for online optimization, which will be described in Sect. 5.

Using the above definitions and assumptions, it is possible to recall results provided in [1] that can be directly applied to (24):

Lemma 3 *The following facts are equivalent:*

1. *for all allowable $v_k \in \mathscr{E}_v$, $k \geq 0$, the system (24) is strictly quadratically bounded with $\boldsymbol{P} \succ \boldsymbol{0}$,*
2. *for all allowable $v_k \in \mathscr{E}_v$, $k \geq 0$, the ellipsoid (46) is a robust invariant set for the system (24).*

Nonetheless, there exists a drawback related to the fact that $\boldsymbol{P}(h_k) = \boldsymbol{P}$. To show, how to avoid it, the notion of EQNB(introduced in [41]) should be recalled. In the light of this framework Definitions 1 and 2 can be suitably reformulated as:

Definition 3 The system (24) is strictly non-quadratically bounded for all allowable $v_k \in \mathscr{E}_v$, $k \geq 0$, if $V_k = \boldsymbol{x}_{f,k}^T \boldsymbol{P}(h_k) \boldsymbol{x}_{f,k} > 1$ implies $\Delta V = V_{k+1} - V_k < 0$ for any $v_k \in \mathscr{E}_v$.

Definition 4 A set $\mathscr{E}_{\boldsymbol{x}_{f,k}}$

$$\mathscr{E}_{\boldsymbol{x}_{f,k}} = \{\boldsymbol{x}_{f,k} \in \mathbb{R}^n \,|\, \boldsymbol{x}_{f,k}^T \boldsymbol{P}(h_k) \boldsymbol{x}_{f,k} \leq 1\} \tag{47}$$

is a robust invariant set for the system (24) for all allowable $v_k \in \mathscr{E}_v$ if $\boldsymbol{x}_{f,k} \in \mathscr{E}_{\boldsymbol{x}_{f,k}}$ implies $\boldsymbol{x}_{f,k+1} \in \mathscr{E}_{\boldsymbol{x}_{f,k+1}}$, for any $v_k \in \mathscr{E}_v$.

Following the above reasoning, Lemma 3 can be reformulated in a similar fashion:

Lemma 4 *The following facts are equivalent*

1. *for all allowable $v_k \in \mathscr{E}_v$, $k \geq 0$, the system (24) is strictly non-quadratically bounded,*
2. *for all allowable $v_k \in \mathscr{E}_v$, $k \geq 0$, the ellipsoid (47) is a robust invariant set for the system (24).*

Using Definition 3 and the fact that $v_k^T \boldsymbol{Q} v_k \leq 1$ (cf. (19)), it is possible to write:

$$v_k^T \boldsymbol{Q} v_k < \boldsymbol{x}_{f,k}^T \boldsymbol{P}(h_k) \boldsymbol{x}_{f,k} \Rightarrow \boldsymbol{x}_{f,k+1}^T \boldsymbol{P}(h_{k+1}) \boldsymbol{x}_{f,k+1} - \boldsymbol{x}_{f,k} \boldsymbol{P}(h_k) \boldsymbol{x}_{f,k} < 0. \tag{48}$$

which by Definition 3 gives

$$v_k^T \boldsymbol{Q} v_k < \boldsymbol{x}_{f,k}^T \boldsymbol{P}(h_k) \boldsymbol{x}_{f,k} \Rightarrow \Delta V < 0, \tag{49}$$

which can be written in a matrix form

$$\begin{bmatrix} -\boldsymbol{P}(h_k) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{bmatrix} \prec \boldsymbol{0} \Rightarrow \tag{50}$$

$$\begin{bmatrix} \boldsymbol{A}_1(h_k)^T \boldsymbol{P}(h_{k+1}) \boldsymbol{A}_1(h_k) - \boldsymbol{P}(h_k) & \boldsymbol{A}_1(h_k)^T \boldsymbol{P}(h_{k+1}) \boldsymbol{B}(h_k) \\ \boldsymbol{B}(h_k)^T \boldsymbol{P}(h_{k+1}) \boldsymbol{A}_1(h_k) & \boldsymbol{B}(h_k)^T \boldsymbol{P}(h_{k+1}) \boldsymbol{B}(h_k) \end{bmatrix} \prec \boldsymbol{0}. \tag{51}$$

By applying the S-Lemma [8], the relations (50)–(51) can be written in the form:

$$\begin{bmatrix} \boldsymbol{A}_1(h_k)^T \\ \boldsymbol{B}(h_k)^T \end{bmatrix} \boldsymbol{P}(h_{k+1}) \begin{bmatrix} \boldsymbol{A}_1(h_k) \boldsymbol{B}(h_k) \end{bmatrix} + \begin{bmatrix} -(1-\alpha)\boldsymbol{P}(h_k) & \boldsymbol{0} \\ \boldsymbol{0} & -\alpha \boldsymbol{Q} \end{bmatrix} \prec \boldsymbol{0}, \tag{52}$$

where $\alpha > 0$ is some scalar. Thus, by applying Schur complement it can be written as

$$\begin{bmatrix} -(1-\alpha)\boldsymbol{P}(h_k) & \boldsymbol{0} & \boldsymbol{A}_1(h_k)^T \\ \boldsymbol{0} & -\alpha \boldsymbol{Q} & \boldsymbol{B}(h_k)^T \\ \boldsymbol{A}_1(h_k) & \boldsymbol{B}(h_k) & -\boldsymbol{P}(h_{k+1})^{-1} \end{bmatrix} \prec \boldsymbol{0}, \tag{53}$$

and subsequently, multiplying it from left and right by $\mathrm{diag}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{P}(h_{k+1}))$

$$\begin{bmatrix} -(1-\alpha)\boldsymbol{P}(h_k) & \boldsymbol{0} & \boldsymbol{A}_1(h_k)^T \boldsymbol{P}(h_{k+1}) \\ \boldsymbol{0} & -\alpha \boldsymbol{Q} & \boldsymbol{B}^T \boldsymbol{P}(h_{k+1}) \\ \boldsymbol{P}(h_{k+1}) \boldsymbol{A}_1(h_k) & \boldsymbol{P}(h_{k+1}) \boldsymbol{B}(h_k) & -\boldsymbol{P}(h_{k+1}) \end{bmatrix} \prec \boldsymbol{0}. \tag{54}$$

Applying Lemma 1 to (54) and then substituting $\boldsymbol{A}_1(h_k) = \sum_{i=1}^{M}\sum_{j=1}^{M} h_i(\boldsymbol{s}_k)h_j(\boldsymbol{s}_k)(\boldsymbol{A}^i - \boldsymbol{B}^i\boldsymbol{K}^j) = \sum_{i=1}^{M}\sum_{j=1}^{M} h_i(\boldsymbol{s}_k)h_j(\boldsymbol{s}_k)\boldsymbol{A}_1^{i,j}$ gives

$$\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{l=1}^{M} h_i(\boldsymbol{s}_k)h_j(\boldsymbol{s}_k)h_l(\boldsymbol{s}_{k+1})\Psi_{i,j}^{l} \prec \boldsymbol{0}, \tag{55}$$

where

$$\Psi_{i,j}^{l} = \begin{bmatrix} -(1-\alpha)\boldsymbol{P}^i & \boldsymbol{0} & \boldsymbol{A}_1^{(i,j)T}\boldsymbol{P}^l \\ \boldsymbol{0} & -\alpha\boldsymbol{Q} & \boldsymbol{B}^{(i)T}\boldsymbol{P}^l \\ \boldsymbol{P}^l\boldsymbol{A}_1^{i,j} & \boldsymbol{P}^l\boldsymbol{B}^i & -\boldsymbol{P}^l \end{bmatrix} \tag{56}$$

Similarly as in the previous section, this allows writing the following lemma:

Lemma 5 *Condition* (56) *is fulfilled providing the following conditions hold:*

$$\Psi_{i,i}^{l} \prec \boldsymbol{0}, \quad i \in \{1,\ldots,M\}, \tag{57}$$

$$\frac{2}{M-1}\Psi_{i,i}^{l} + \Psi_{i,j}^{l} + \Psi_{j,i}^{l} \prec \boldsymbol{0}, \quad i,j,l \in \{1,\ldots,M\}, \quad i \neq j, \tag{58}$$

$$0 < \alpha < 1. \tag{59}$$

Note that for a fixed α, the procedure is simply solving LMIs (57)–(58) with respect to $\boldsymbol{P}^i$ $(i = 1,\ldots,M)$.

5 Efficient Predictive FTC

The robust FTC from in Sect. 4 is based on the idea of estimating the fault, and then compensating it with adapted control signal feeding the faulty actuator. Its main drawback is that it does not take into account the fact that actuators are subject to saturation rules. Therefore, the idea behind the approach presented in this section is to prevent a saturation by fixing the control strategy of the remaining actuators in a way increasing the robust invariant set. Thus, making the overall control problem feasible. For this purpose, the efficient predictive control scheme introduced by [22] is utilized. In particular, the proposed scheme is suitably extended to cope with the noises and disturbances, and hence to achieve robustness. Thus, predictions at time k are generated as follows [22]:

$$z_{k+1} = \boldsymbol{Z}(h_k)z_k + \tilde{\boldsymbol{B}}(h_k)v_k. \tag{60}$$

where

$$\boldsymbol{Z}(h_k) = \begin{bmatrix} \boldsymbol{A}(h_k) - \boldsymbol{B}(h_k)\boldsymbol{K}(h_k) & \boldsymbol{B}(h_k)\boldsymbol{T} \\ \boldsymbol{0} & \boldsymbol{M} \end{bmatrix},$$
$$\tilde{\boldsymbol{B}}(h_k) = \begin{bmatrix} \boldsymbol{B}(h_k) \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{M} = \begin{bmatrix} \boldsymbol{0}_{(n_c-1)r\times r} & \boldsymbol{I} \\ \boldsymbol{0}_{r\times r} & \boldsymbol{0}_{r\times(n_c-1)r} \end{bmatrix},$$
$$z_k = \begin{bmatrix} \boldsymbol{x}_{f,k} \\ \boldsymbol{\delta}_k \end{bmatrix}, \quad \boldsymbol{\delta}_k = \begin{bmatrix} \boldsymbol{c}_k \\ \boldsymbol{c}_{k+1} \\ \dots \\ \boldsymbol{c}_{k+n_c-1} \end{bmatrix}, \quad \boldsymbol{T} = \begin{bmatrix} \boldsymbol{I}_{r\times r} & \boldsymbol{0} \vdots \boldsymbol{0} \end{bmatrix}.$$

The stability of (60) is guaranteed by the stability of $\boldsymbol{A}(h_k) - \boldsymbol{B}(h_k)\boldsymbol{K}(h_k)$. After [22], it can be seen that if there exists robust invariant set $\mathscr{E}_{\boldsymbol{x}_f}$ (cf. (47)) for (60), there must exist at least one robust invariant set $\mathscr{E}_z$ for (60). Thus, (56) can be easily adapted for (60)

$$\begin{bmatrix} -(1-\alpha)\boldsymbol{P}^i & \boldsymbol{0} & \boldsymbol{Z}^{(i,j)T}\boldsymbol{P}^l \\ \boldsymbol{0} & -\alpha\boldsymbol{Q} & \tilde{\boldsymbol{B}}^{(i)T}\boldsymbol{P}^l \\ \boldsymbol{P}^l\boldsymbol{Z}^{i,j} & \boldsymbol{P}^l\tilde{\boldsymbol{B}}^i & -\boldsymbol{P}^l \end{bmatrix}, \qquad 0 < \alpha < 1 \tag{61}$$

being a robust invariant set for FTC predictive scheme. Since the robust invariant set for (60) is given, the input constraints (64) can be formulated. It can be done by scaling $\boldsymbol{\delta}_k$ in (60), i.e. $\boldsymbol{\delta}_k$ is replaced by:

$$\bar{\boldsymbol{\delta}}_k = \begin{bmatrix} \boldsymbol{c}_k - \hat{\boldsymbol{f}}_{k-1} \\ \boldsymbol{c}_{k+1} - \hat{\boldsymbol{f}}_{k-1} \\ \vdots \\ \boldsymbol{c}_{k+n_c-1} - \hat{\boldsymbol{f}}_{k-1} \end{bmatrix}. \tag{62}$$

Let us define

$$\boldsymbol{M}(h_k) = [-\boldsymbol{K}(h_k) \quad \boldsymbol{T}], \tag{63}$$

and hence

$$\boldsymbol{u}_{f,k} = \boldsymbol{M}(h_k)z_k. \tag{64}$$

Let $\boldsymbol{e}_i$ denote ith column of the r-order identity matrix, which makes it possible to rewrite the input constraints as follows

$$|\boldsymbol{e}_i^T\boldsymbol{u}_{f,k}| \le \bar{u}_i \quad i = 1, \dots, r \tag{65}$$

Subsequently, it can be observed that for $z_k \in \mathscr{E}_z$ the above inequality implies

$$
\begin{aligned}
|\boldsymbol{e}_i^T \boldsymbol{u}_{f,k}|^2 &= |\boldsymbol{e}_i^T \boldsymbol{M}(h_k)\boldsymbol{z}_k|^2 \\
&= |\boldsymbol{e}_i^T \boldsymbol{M}(h_k)\boldsymbol{P}(h_k)^{-1/2}\boldsymbol{P}(h_k)^{1/2}\boldsymbol{z}_k|^2 \\
&\leq ||\boldsymbol{e}_i^T \boldsymbol{M}(h_k)\boldsymbol{P}(h_k)^{-1/2}||^2 ||\boldsymbol{P}(h_k)^{1/2}\boldsymbol{z}_k||^2 \\
&\leq ||\boldsymbol{e}_i^T \boldsymbol{M}(h_k)\boldsymbol{P}(h_k)^{-1/2}||^2,
\end{aligned}
\tag{66}
$$

and if there exists a symmetric matrix $\boldsymbol{Y}$ such that

$$
\boldsymbol{M}(h_k)\boldsymbol{P}(h_k)^{-1}\boldsymbol{M}(h_k)^T - \boldsymbol{Y} \prec \boldsymbol{0}, \quad \boldsymbol{Y}_{i,i} \leq \bar{u}_i^2, \quad i = 1, \ldots, r \tag{67}
$$

then $|u_{i,f,k}| \leq \bar{u}_i$, $(i = 1, \ldots, r)$. Subsequently, using the Schur complement, inequality (67) can be written as

$$
\begin{bmatrix} -\boldsymbol{Y} & \boldsymbol{M}(h_k) \\ \boldsymbol{M}(h_k)^T & -\boldsymbol{P}(h_k) \end{bmatrix} \prec \boldsymbol{0}, \quad \boldsymbol{Y}_{i,i} \leq \bar{u}_i^2. \tag{68}
$$

Finally to count also the input constraints, (61) should be supplemented with

$$
\begin{bmatrix} -\boldsymbol{Y} & (\boldsymbol{M}^j) \\ (\boldsymbol{M}^j)^T & -\boldsymbol{P}^i \end{bmatrix} \prec \boldsymbol{0}, \quad \boldsymbol{Y}_{a,a} \leq \bar{u}_a^2, \\
a = 1, \ldots, r, \quad i = 1, \ldots, M, \quad j = 1, \ldots, M, \tag{69}
$$

Since invariant sets along with input constraints are described by LMIs, they can be solved along with maximization the invariant set. To accomplish this, different criteria can be selected, e.g.:

- minimization of the determinant of $\boldsymbol{P}(h_k)$, which corresponds to the maximization of volume of the invariant set,
- minimization of the trace of $\boldsymbol{P}(h_k)$, which corresponds to the maximization of the sum of the axes of the ellipsoid describing an invariant set.

Considering the structure of $\boldsymbol{P}(h_k)$, being a weighted sum of matrices, maximization of the size of the $\mathscr{E}_{\boldsymbol{x}_f}$, the following sum of traces is equivalent to:

$$
\min \operatorname{trace}\left(\sum_{i=1}^{M} \left(\boldsymbol{T}\,\boldsymbol{P}^i\,\boldsymbol{T}^T\right)\right) = \min \operatorname{trace}\left(\operatorname{diag}\left(\boldsymbol{T}\,\boldsymbol{P}^1\boldsymbol{T}^T, \ldots, \boldsymbol{T}\,\boldsymbol{P}^M\boldsymbol{T}^T\right)\right) \tag{70}
$$

with

$$
\boldsymbol{x}_{f,k} = \boldsymbol{T}\,\boldsymbol{z}_k, \tag{71}
$$

and constraints (61) and (69). The algorithm for finding $\boldsymbol{c}_k$ in (60) is also inspired by [22] and can be perceived as the following minimization

$$\boldsymbol{\delta}_k^* = \min_{\boldsymbol{\delta}_k} \ \boldsymbol{\delta}_k^T \boldsymbol{\delta}_k, \quad \text{s.t. } z_k^T \boldsymbol{P}(h_k) z_k \leq 1, \tag{72}$$

performed at each sampling time, which can be equivalently written as:

$$\begin{aligned} \boldsymbol{\delta}_k^* = \min_{\boldsymbol{\delta}_k} \ & \boldsymbol{\delta}_k^T \boldsymbol{\delta}_k, \quad \text{s.t. } \boldsymbol{x}_{f,k}^T \boldsymbol{P}_{1,1}(h_k) \boldsymbol{x}_{f,k} \\ & + 2\boldsymbol{x}_{f,k}^T \boldsymbol{P}_{1,2}(h_k) \bar{\boldsymbol{\delta}}_k \\ & + \bar{\boldsymbol{\delta}}_k^T \boldsymbol{P}_{2,2}(h_k) \bar{\boldsymbol{\delta}}_k \leq 1, \end{aligned} \tag{73}$$

where $\boldsymbol{P}_{1,1}(h_k)$, $\boldsymbol{P}_{1,2}(h_k)$ and $\boldsymbol{P}_{2,2}(h_k)$ are block partitions of $\boldsymbol{P}(h_k)$ conformal to the partition of $z_k = [\boldsymbol{x}_{f,k}^T \ \bar{\boldsymbol{\delta}}_k^T]^T$. Thus, if the $\mathscr{H}_\infty$ control is feasible then $\boldsymbol{\delta} = \boldsymbol{0}$, if not, the solution lies on the boundary of $\mathscr{E}_z$ described by (73). Therefore, when $\boldsymbol{\delta} = \boldsymbol{0}$ is contained in $\mathscr{E}_z$, then optimization is not required and the optimal solution is $\boldsymbol{\delta} = \boldsymbol{0}$. Otherwise, the above optimization problem has a unique solution [16, 21].

6 Illustrative Example

The exemplary nonlinear system is based on the Twin-Rotor MIMO System (TRMS), a laboratory set-up [11] developed by Feedback Instruments Limited. Research regarding TRMS itself can be found in [26] and the references therein. The TRMS, as shown in Fig. 2, is driven by two DC motors. It has two perpendicular propellers joined by a beam pivoted on its base. Chassis has 2 degrees of freedom, allows movement the inside imaginary static sphere. The beam can be moved by changing the input voltages of motors that control the rotational speed of the propellers. The system is equipped with a pendulum counterweight fixed to the beam and it

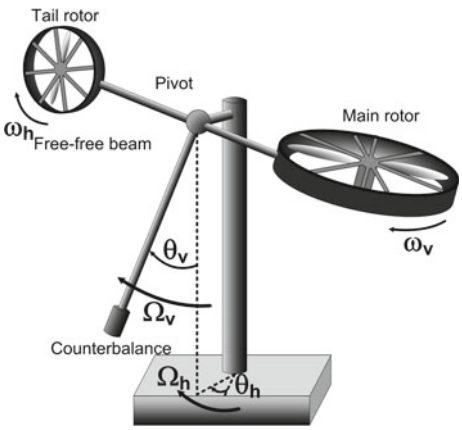


Fig. 2 Components of the twin-rotor MIMO system

determines a stable equilibrium position. The system is balanced in such a way that when the motors are switched off, the main rotor end of the beam is lowered. The behaviour of the TRMS system is similar to a helicopter [26]. For instance, there is a strong cross-coupling between the main rotor (collective) and the tail rotor. At the same time, the system is different from a helicopter in many ways, the main differences being the location of the pivot point (midway between two rotors in TRMS vs. main rotor head in the helicopter), vertical control (speed control of main rotor vs. collective pitch control), yaw control (tail rotor speed vs. pitch angle of tail rotor blades) and lastly, cyclical control (none vs. directional control).

The model of the TRMS can be described by a set of four nonlinear differential equations with two linear differential equations and four nonlinear functions [11]. Some of the parameters can be obtained from manual [11], whereas others should be collected as an experimental results, e.g. inertia, magnitudes of the physical propeller, coefficients of friction and impulse force. The inputs are defined by the input vector $\boldsymbol{u} = [u_h, u_v]^T$, where u_h is the input voltage of the tail motor and u_v is the input voltage of the main motor. The input bounds are $\bar{u}_1 = 1$ and $\bar{u}_2 = 1$. The state vector is defined as $\boldsymbol{x} = [\Theta_h, \phi_h, \theta_h, \Theta_v, \phi_v, \theta_v]^T$, where Θ_h is the angular velocity around the vertical axis, ϕ_h is the azimuth angle of the beam, θ_h is the rotational velocity of the tail rotor, Θ_v is the angular velocity around the horizontal axis, ϕ_v is the pitch angle of the beam, θ_v is the rotational velocity of the main rotor. For the complete physical model of such a system refer to [11, 26]. For the purpose of experiment the nonlinear simulator of the TRMS was used.

A normalized TS model, which approximates the nonlinear TRMS system, is obtained by linearising a system around five operating points [5]. The system can be described in the following way:

$$\boldsymbol{x}_{f,k+1} = \sum_{i=1}^{5} h_i(\boldsymbol{s}_k)[\boldsymbol{A}^i \boldsymbol{x}_{f,k} + \boldsymbol{B}^i \boldsymbol{u}_{f,k} + \boldsymbol{B}^i \boldsymbol{f}_k + \boldsymbol{W}\boldsymbol{w}_k], \tag{74}$$

The matrices $\boldsymbol{A}^i$, and $\boldsymbol{B}^i$, $(i = 1, \ldots, 5)$ are acquired by linearising the initial system around five points chosen in the operating range of the system considered. A detailed description of the model (74) can be found in [5]. Note also that, according to [5] the constant bias arising from the linearization were removed due to their avoidably small values. Moreover, it was verified that the matrices $B^i, i = 1, \ldots, M$ satisfy the conditions of Theorem 3.1, which makes it possible to conduct the remaining design procedure. Five local models guarantee a relatively good approximation of the state of the real system by the TS model within the operating range, also the state vector was used as a premise variables vector. The following numerical values, with the sampling time 50 ms, were used:

$$\boldsymbol{A}^1 = \begin{bmatrix} 0.9812 & -0.0105 & 0.1847 & 0 & 0 & 0 \\ 0 & 0.9657 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8780 & 0 & 0 & 0 \\ 0 & 0.0152 & -0.0254 & 0.9908 & -0.1718 & 0 \\ 0 & 0.0004 & 0.1367 & 0.0498 & 0.9957 & 0 \\ 0.0495 & 0.0276 & 0.0047 & 0 & 0 & 1 \end{bmatrix},$$

$$\boldsymbol{A}^2 = \begin{bmatrix} 0.9814 & -0.0103 & 0.1841 & 0 & 0.0004 & 0 \\ 0 & 0.9657 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8780 & 0 & 0 & 0 \\ 0 & 0.0200 & -0.0254 & 0.9908 & -0.1718 & 0 \\ 0 & 0.0005 & 0.1367 & 0.0498 & 0.9957 & 0 \\ 0.0495 & 0.0274 & 0.0046 & 0 & -0.0010 & 1 \end{bmatrix},$$

$$\boldsymbol{A}^3 = \begin{bmatrix} 0.9818 & -0.0098 & 0.1830 & 0 & 0.0007 & 0 \\ 0 & 0.9657 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8780 & 0 & 0 & 0 \\ 0 & 0.0405 & -0.0254 & 0.9908 & -0.1718 & 0 \\ 0 & 0.0010 & 0.1367 & 0.0498 & 0.9957 & 0 \\ 0.0495 & 0.0268 & 0.0045 & -0.0001 & -0.0020 & 1 \end{bmatrix},$$

$$\boldsymbol{A}^4 = \begin{bmatrix} 0.9826 & -0.0090 & 0.1809 & 0 & 0.0010 & 0 \\ 0 & 0.9657 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8780 & 0 & 0 & 0 \\ 0 & 0.0734 & -0.0254 & 0.9908 & -0.1717 & 0 \\ 0 & 0.0018 & 0.1367 & 0.0498 & 0.9957 & 0 \\ 0.0496 & 0.0256 & 0.0044 & -0.0001 & -0.0030 & 1 \end{bmatrix},$$

$$\boldsymbol{A}^5 = \begin{bmatrix} 0.9837 & -0.0079 & 0.1774 & 0 & 0.0013 & 0 \\ 0 & 0.9657 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8780 & 0 & 0 & 0 \\ 0 & 0.1126 & -0.0254 & 0.9908 & -0.1712 & 0 \\ 0 & 0.0028 & 0.1367 & 0.0498 & 0.9957 & 0 \\ 0.0496 & 0.0239 & 0.0043 & -0.0001 & -0.0039 & 1 \end{bmatrix},$$

$$\boldsymbol{B}^1 = \begin{bmatrix} 0.0047333 & -0.00026413 \\ 0 & 0.049137 \\ 0.046883 & 0 \\ -0.00054491 & 0.00038216 \\ 0.0034973 & 6.3923 \cdot 10^{-6} \\ 7.9866 \cdot 10^{-5} & 0.00069499 \end{bmatrix}, \boldsymbol{B}^2 = \begin{bmatrix} 0.0047179 & -0.00025932 \\ 0 & 0.049137 \\ 0.046883 & 0 \\ -0.0005449 & 0.00050447 \\ 0.0034973 & 8.4382 \cdot 10^{-6} \\ 7.8406 \cdot 10^{-5} & 0.00069104 \end{bmatrix},$$

$$\boldsymbol{B}^3 = \begin{bmatrix} 0.0046897 & -0.00024777 \\ 0 & 0.049137 \\ 0.046883 & 0 \\ -0.00054487 & 0.001021 \\ 0.0034973 & 1.7078 \cdot 10^{-5} \\ 7.6732 \cdot 10^{-5} & 0.0006752 \end{bmatrix}, \boldsymbol{B}^4 = \begin{bmatrix} 0.0046345 & -0.00022722 \\ 0 & 0.049137 \\ 0.046883 & 0 \\ -0.00054472 & 0.0018479 \\ 0.0034973 & 3.0909 \cdot 10^{-5} \\ 7.4631 \cdot 10^{-5} & 0.00064594 \end{bmatrix},$$

$$\boldsymbol{B}^5 = \begin{bmatrix} 0.0045431 & -0.00019841 \\ 0 & 0.049137 \\ 0.046883 & 0 \\ -0.00054422 & 0.0028368 \\ 0.0034973 & 4.745 \cdot 10^{-5} \\ 7.2027 \cdot 10^{-5} & 0.00060318 \end{bmatrix}$$

Experiments appear to show that sampling time of 50 ms is small enough to represent the dynamic and at the same time not overact the computational burden. The robust $\mathscr{H}_\infty$ controller gain matrix $\boldsymbol{K}(h_k)$ has been obtained by solving (28) with a predefined attenuation level $\mu = 0.2$ and

$$\boldsymbol{K}(h_k) = \begin{bmatrix} -0.66072 & 0.02514 & 18.5299 & 1.4495 & -1.2791 & -0.77947 \\ -0.13447 & 19.6351 & 0.035917 & -0.54517 & 0.60774 & -0.34511 \end{bmatrix}$$

for $\boldsymbol{W} = 0.01\boldsymbol{I}$. While the prediction horizon was set to $n_c = 6$, which guarantees a good compromise between the complexity and quality of FTC (as a conclusion of experiments). The actuator faults scenarios, i.e. a decrease in the performance of the main rotor ($\boldsymbol{f}_{1,k}$) and a rotor misalignment in tail electric motor ($\boldsymbol{f}_{2,k}$) are described as follows:

$$\boldsymbol{f}_{1,k} = \begin{cases} -0.2 & 50 \leq k \leq 90, \\ 0 & \text{otherwise} \end{cases}$$

$$\boldsymbol{f}_{2,k} = \begin{cases} 0.5 \sin(2\pi + 1 + 0.1\pi(k - 80)) & 80 \leq k \leq 120, \\ 0 & \text{otherwise} \end{cases}$$

Figure 3 presents the horizontal and vertical angular position of the beam, achieved for the proposed FTC strategy. As a result, Fig. 4 clearly points out that the faults can be estimated with a very high accuracy. The fault estimate exhibits some deviations

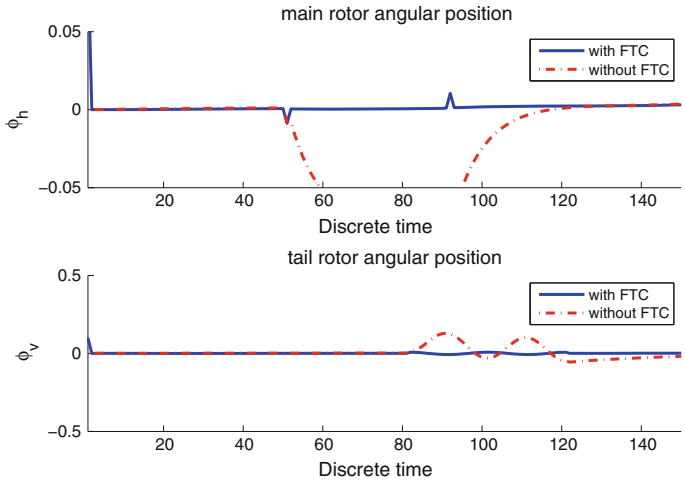


Fig. 3 Selected states of the system with and without FTC

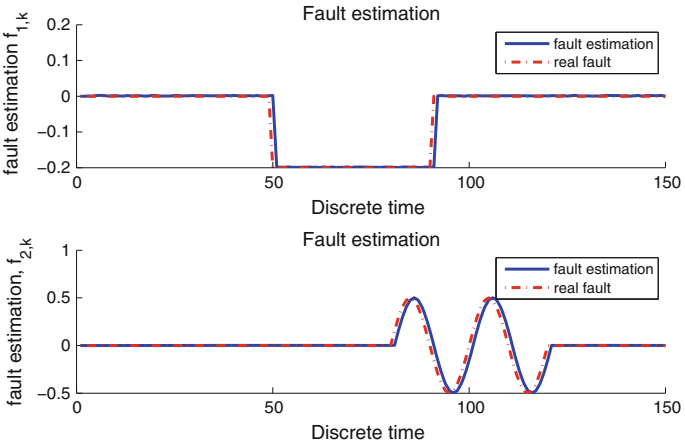


Fig. 4 Faults and their estimates

from the nominal value due to exogenous disturbances and modelling errors (due to the high nonlinearity of the system). Contrary to the non-FTC scheme, the proposed method exhibits a small error only. Figure 5 shows control signals for the chosen fault scenario. Nonlinearities in TRMS have also influence on the faulty behaviour of the system. Thus, Fig. 5 shows appropriate control that should be provided for both inputs in order to stabilize the system behaviour. This case proves the efficiency of the proposed control scheme for respectively large faults. Figure 3 presents stabilization of the beam position with FTC enabled while even small fault evolves into large deviation in the beam position for the non-FTC scheme. Nonetheless, overshoots

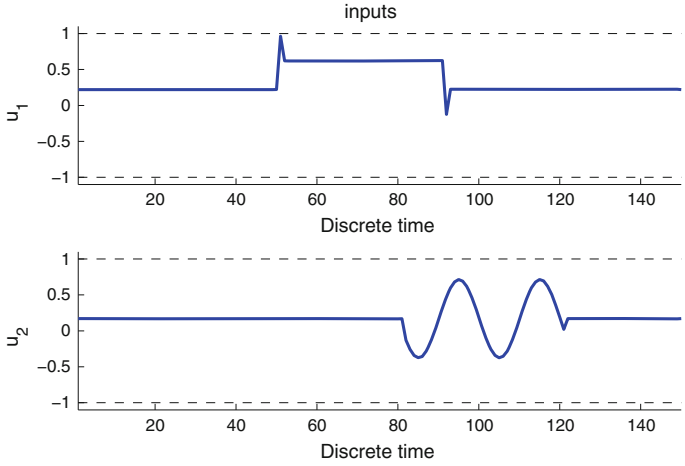


Fig. 5 Control inputs

cannot be totally avoided nor suppressed. Their existence have minor influence on the overall performance of the system, and hence, the proposed scheme can be perceived as a reliable one.

7 Conclusions

The essential contribution of the work is the fault-tolerant scheme for the Takagi-Sugeno discrete-time systems. The presented approach undoubtedly fulfils the goals presumed in the introduction and therefore deals with the actuator faults with given performance limits. The multi-layer scheme consists of the robust controller, fault estimator and predictive controller. The robust controller is designed to prevent influences of the disturbances on the system. Fault estimator allows the fault compensation, while the predictive controller can employ remaining actuators to prevent saturation on faulty ones. As indicated in the state-of-the-art, there have been no efficient solution to this problem so far. All the proposed approaches can be efficiently implemented, i.e. the offline computations boils down to solving a number of linear matrix inequalities while the online computation reduces to the application of the Newton–Raphson method. The proposed approach was applied to the benchmark example of the twin-rotor system. The achieved results show the performance of the high performance of the proposed approach. In spite of the incontestable appeal of the proposed approach, there are still some points which may further improve its effectiveness.

Acknowledgments The work was financed as a research project with the science funds from the National Science Centre in Poland. The authors would like to thank project-related colleagues, who have a significant impact on the shape of this contribution.

Appendix

Table of symbols:

$\boldsymbol{A}$ is a system matrix
$\boldsymbol{B}$ is an input matrix
$\boldsymbol{W}$ is a disturbance matrix
V is Lyapunov function
$\boldsymbol{x}_k$ is a state vector
$\boldsymbol{u}_k$ is an input vector
$\boldsymbol{w}_k$ is a disturbance vector
k is a time instant
$\boldsymbol{K}$ is a controller gain matrix
$\mathscr{E}$ is an invariant set
$\bullet^+$ is a pseudo-inverse operation
Θ_h is the angular velocity around the vertical axis
Θ_v is the angular velocity around the horizontal axis
θ_h is the rotational velocity of the tail rotor
θ_v is the rotational velocity of the main rotor
ϕ_v is the pitch angle of the beam
ϕ_h is an invariant set

References

1. Alessandri A, Baglietto M, Battistelli G (2006) Design of state estimators for uncertain linear systems using quadratic boundedness. Automatica 42(3):497–502
2. Chadli M, Karimi HR (2013) Robust observer design for unknown inputs Takagi-Sugeno models. IEEE Trans Fuzzy Syst 21(1):158–164
3. Chen W, Khan AQ, Abid M, Ding SX (2011) Integrated design of observer-based fault detection for a class of uncertain non-linear systems. Int J Appl Math Comput Sci 21(4):619–636
4. Daafouz J, Riedinger P, Iung C (2002) Stability analysis and control synthesis for switched systems: a switched lyapunov function approach. IEEE Trans Autom Control 47(11):1883–1887
5. de Oca S, Puig V, Witczak M, Dziekan L (2012) Fault-tolerant control strategy for actuator faults using lpv techniques: application to a two degree of freedom helicopter. Int J Appl Math Comput Sci 22(1):161–171
6. de Oliveira MC, Bernussou J, Geromel JC (1999) A new discrete-time robust stability condition. Syst Control Lett 37(4):261–265
7. Delmotte F, Guerra TM, Ksantini M (2007) Continuous Takagi-Sugeno's models: reduction of the number of LMI conditions in various fuzzy control design technics. IEEE Trans Fuzzy Syst 15(3):426–438

8. Derinkuyu K, Pınar M (2006) On the s-procedure and some variants. Math Methods Op Res 64(1):55–77
9. Dziekan L, Witczak M, Korbicz J (2011) Active fault-tolerant control design for Takagi-Sugeno fuzzy systems. Bull Pol Acad Sci : Tech Sci 59(1):93–102
10. Blanchini F (1999) Set invariance in control. In Automatica 35:1747–1767
11. Feedback Instruments Limited (1998) Twin rotor MIMO system advanced teaching manual 1. Crowborough, UK
12. Gilbert EG, Tan KT (1991) Linear systems with state and control constraints: the theory and application of maximal output admissible sets. IEEE Trans Autom Control 36(9):1008–1020
13. Gillijns S, De Moor B (2007) Unbiased minimum-variance input and state estimation for linear discrete-time systems. Automatica 43:111–116
14. Guerra TM, Kruszewski A, Lauber J (2009) Discrete Tagaki-Sugeno models for control: where are we? Annu Rev control 33(1):37–47
15. Ichalal D, Marx B, Ragot J, Maquin D (2012) Advances in observer design for Takagi-Sugeno systems with unmeasurable premise variables. In 20th IEEE mediterranean conference on control & automation (MED), IEEE, pp 848–853
16. Imsland L, Bar N, Foss BA (2005) More efficient predictive control. Automatica 41(8):1395–1403
17. Iserman R (2011) Fault diagnosis applications: model based condition monitoring, actuators, drives, machinery, plants, sensors, and fault-tolerant systems. Springer, Berlin
18. Kolodziejczak B, Szulc T (1999) Convex combinations of matrices. full rank characterization. Linear Algebra Appl 287(1–3):215–222
19. Korbicz J, Kościelny J, Kowalczuk Z, Cholewa W (eds) (2004) Fault diagnosis. models, artificial intelligence, applications. Springer, Berlin
20. Korbicz J, Witczak M, Puig V (2007) LMI-based strategies for designing observers and unknown input observers for non-linear discrete-time systems. Bull Pol Acad Sci : Tech Sci 55(1):31–42
21. Kouvaritakis B, Cannon M, Rossiter JA (2002) Who needs QP for linear MPC anyway? Automatica 38(5):879–884
22. Kouvaritakis B, Rossiter JA, Schuurmans J (2000) Efficient robust predictive control. IEEE Trans Autom Control 45(8):1545–1549
23. Li H, Fu M (1997) A linear matrix inequality approach to robust $\mathscr{H}_\infty$ filtering. IEEE Trans Signal Process 45(9):2338–2350
24. Mahmoud M, Jiang J, Zhang Y (2003) Active fault tolerant control systems: stochastic analysis and synthesis. Springer, Berlin
25. Mrugalski M (2013) An Unscented Kalman Filter in designing dynamic GMDH neural networks for robust fault detection. Int J Appl Math Comput Sci 23(1):157–169
26. Rahideh A, Shaheed MH (2007) Mathematical dynamic modelling of a twin-rotor multiple input-multiple output system. Inst Mech Eng Part I: J Syst Control Eng 227:89–101
27. Takagi T, Sugeno M (1985) Fuzzy identification of systems and its application to modeling and control. IEEE Trans Syst Man Cybern 15(1):116–132
28. Tanaka K, Wang HO (2001) Fuzzy control systems design and analysis: a linear matrix inequality approach. Wiley-Interscience, New York
29. Tuan HD, Apkarian P, Narikiyo T, Yamamoto Y (2001) Parameterized linear matrix inequality techniques in fuzzy control system design. IEEE Trans Fuzzy Syst 9(2):324–332
30. Wang HO, Tanaka K, Griffin MF (1996) An approach to fuzzy control of nonlinear systems: stability and design issues. IEEE Trans Fuzzy Syst 4(1):14–23
31. Witczak M (2007) Modelling and estimation strategies for fault diagnosis of non-linear systems. Springer, Berlin
32. Witczak M (2014) Fault diagnosis and fault-tolerant control strategies for non-linear systems. Lectures Notes in Electrical Engineering, vol 266. Springer International Publisher, Heidelberg, Germany
33. Witczak M, Korbicz J, Jozefowicz R (2013) Design of unknown input observers for non-linear stochastic systems and their application to robust fault diagnosis. Control and Cybernetics 42(1):227–256

34. Witczak M, Pretki P (2007) Design of an extended unknown input observer with stochastic robustness techniques and evolutionary algorithms. Int J Control 80(5):749–762
35. Witczak M, Puig V, de Oca S (2013) A fault-tolerant control strategy for non-linear discrete-time systems: application to the twin-rotor system. Int J Control 86(10):1788–1799
36. Witczak M, Witczak P (2013) Efficient predictive fault-tolerant control for non-linear systems. In: 11th international conference on diagnostics of processes and systems - DPS 2013. Springer, Berlin, Heidelberg, pp 65–76
37. Witczak M, Witczak P, Korbicz J, Aubrun C (2013) Robust and efficient predictive FTC: application to wind turbines. In: Conference on control and fault-tolerant systems - SysTol 2013. Nice, France, pp 371–376
38. Zemouche A, Boutayeb M, Iulia G (2008) Bara. Observer for a class of Lipschitz systems with extension to $\mathscr{H}_\infty$ performance analysis. Syst Control Lett 57(1):18–27
39. Zhang Y, Jiang J (2003) Bibliographical review on reconfigurable fault-tolerant control systems. IFAC symposium fault detection supervision and safety of technical processes. SAFEPROCESSD.C, USA, Washington, pp 265–276
40. Zongli L, Liang L (2007) Set invariance conditions for singular linear systems subject to actuator saturation. IEEE Trans Autom Control 52(12):2351–2355
41. Zou T, Li S (2011) Stabilization via extended nonquadratic boundedness for constrained nonlinear systems in Takagi-Sugeno's form. J Frankl Inst 348(10):2849–2862

Constrained Model Predictive Control of Processes with Uncertain Structure Modeled by Jump Markov Linear Systems

Jens Tonne and Olaf Stursberg

Abstract Linear systems with abrupt changes in its structure, e.g. caused by component failures of a production system, can be modelled by the use of jump Markov linear systems (JMLS). This chapter proposes a finite horizon model predictive control (MPC) approach for discrete-time JMLS considering input constraints as well as constraints for the expectancy of the state trajectory. For the expected value of the state as well as a quadratic cost criterion, recursive prediction schemes are formulated, which consider dependencies on the input trajectory explicitly. Due to the proposed prediction scheme, the MPC problem can be formulated as a quadratic program (QP) exhibiting low computational effort compared to existing approaches. The resulting properties concerning stability as well as computational complexity are investigated and demonstrated by illustrative simulation studies.

1 Introduction

In practice, component breakdowns, link failures and sudden environmental changes abruptly affect technical systems like production lines or communication networks. These effects often lead to significant changes in the system structure. Most of these changes are random and their time of occurrence is unknown in advance, posing significant challenges for control. On the other hand, failure rates or dropout probabilities can typically be obtained from historical data. Hence, one possibility to model the systems is to represent each possible system behaviour and structure by a separate

J. Tonne (✉)
Volkswagen AG, Baunatal, Germany
email: jens.tonne@volkswagen.de; jens.tonne@uni-kassel.de

J. Tonne · O. Stursberg
Department of Electrical Engineering and Computer Science,
Institute of Control and System Theory, University of Kassel, Kassel, Germany
e-mail: stursberg@uni-kassel.de

© Springer International Publishing Switzerland 2016

A. Rauh and L. Senkel (eds.), *Variable-Structure Approaches*,
Mathematical Engineering, DOI 10.1007/978-3-319-31539-3_12

time-discrete LTI-system. The random transitions between these LTI systems can be modelled by means of a Markov chain, leading to the class of JMLS [1]. Especially, when modelling as you did for modelled industrial processes, the resulting JMLS are characterized by high system dimensions, e.g. resulting from discretizations of spatially varying variables.

Furthermore, the facts that inputs are constrained due to actuator limits and that system states have to be hold in safe ranges, renders the control of the systems even more demanding. A common approach for constrained control is MPC, also called receding horizon control (RHC) [2]. In MPC, the future state trajectory over a given horizon is predicted as a function of the future inputs and expected disturbances. The predictions are used to determine an input trajectory by solving a possibly constrained optimization problem.

Due to the importance of both, the concepts of JMLS and MPC, for technical applications, MPC of JMLS has been studied already to some extend in the current literature, but mostly for low-dimensional systems. First approaches to constrained control of JMLS propose to determine state-feedback control laws by minimizing an infinite horizon quadratic cost function offline [3, 4]. While these approaches consider polytopic uncertainties for the initial state and for the transition matrix of the Markov chain, external disturbances are not considered. The resulting controllers can become quite conservative due to the offline determination of the control laws.

At the same time, first contributions to unconstrained RHC of JMLS which determine state-feedback controllers for each time step of the prediction horizon have been proposed [5, 6]. In addition, a robust unconstrained one-step RHC approach with polytopic uncertainties in the system dynamics and transition probabilities is proposed in [7]. In [8, 9], RHC approaches considering constraints for the first and second moments of inputs and states have been presented. These approaches make use of an iterative procedure which involves solving a semi-definite program (SDP) in each iteration. An approach to determine a set of time-dependent feedback laws for a finite horizon considering second moment constraints for states and inputs is presented in [10].

Particle control as proposed in [11, 12] presents an alternative control approach to JMLS, in which the probability distributions of the Markov chain and disturbances are approximated by a finite set of samples (“particles”). This setting has the advantage that chance constraints for the states can be considered easily by the use of mixed integer linear programs (MILP) [11]. However, due to the sampling, the performance and reliability of the proposed approach strongly depends on the number of particles, which typically has to be chosen high in order to approximate the underlying probability distributions properly.

Recently, a focus of attention has been drawn to MPC approaches considering robust stability in case of uncertain dynamics [13, 14], as well as approaches that ensure mean square stability and recursive feasibility [15]. In [16], this procedure is extended to distributed JMLS.

All the aforementioned approaches for constrained control of JMLS are based on SDP or MILP formulations, for which reliable solvers exist. However, the computational time, needed to solve these problems, increases sharply with increasing number

of variables and constraints. Thus, the online implementation of these approaches is not possible for JMLS with fast system dynamics, high-dimensional state-spaces and high numbers of Markov states. For these systems, simpler formulations of the optimization problem are needed.

In [17, 18], MPC approaches resulting in QP formulations for random parameter systems and systems with Markovian jumps of the input matrix are proposed. However, both approaches are not suitable for general JMLS.

One possibility to reduce the computational burden is the truncation of the scenario tree and to use an MPC scheme for the remaining scenarios. In [19], an MPC approach is presented in which the states are predicted on the basis of the most likely Markov states due to the predicted probability distribution of the Markov chain. By means of this prediction, a quadratically constrained QP (QCQP) is formulated. However, the employed prediction scheme may use Markov state trajectories that are not realizable, since the transition probabilities are neglected. A more suitable approach for the robust control of stochastic systems resulting in a QP formulation is presented in [20], where a predefined number of the most probable realizable Markov state trajectories are used for prediction. In [21], the scenario tree truncation approach from [20] is referred to in order to be able to use dynamic programming and parametric piecewise quadratic optimization methods. However, if the scenario tree is not truncated, both approaches lead to relatively high computational effort.

In [22, 23], the current input signal is determined by the combination of an offline computed stabilizing state feedback law and a one-step MPC. The MPC problem in [22] considers robust stability and probabilistic state constraints resulting in a QCQP, while the concept in [23] aims at mean square stability and probabilistic state constraints resulting in a QP. However, due to the restriction to a linear feedback law for all future steps, the current control action can be suboptimal due to the state constraints. In addition, both approaches do not consider input constraints.

In previous work [24], we proposed a constrained, finite horizon MPC approach resulting in a QP (see Appendix B for a brief description). It was demonstrated in [24] that this approach is capable of controlling high-dimensional JMLS. This approach introduced a new formulation of the cost function defining the costs of the expected values of the states. The formulations of the prediction equations are compact at the expense of specifying costs independently of the Markov state. In addition, mean square stability does not follow from convergence of the costs, if the Markov state is unknown. In contrast, a constrained, finite horizon MPC approach is proposed in this contribution, considering the commonly used cost function formulating expected values of a quadratic cost criterion. In opposite to many of the aforementioned contributions, the proposed MPC is able to control large-scale JMLS and thus is suitable for a variety of industrial applications. It is shown how prediction equations for the expectancy of the state and the quadratic cost criterion can be formulated as a function of the future inputs. These results are used to formulate the MPC problem as a QP similar to the condensed MPC formulations for LTI systems. For the proposed approach, no truncation of the scenario tree is necessary.

2 System Dynamics and MPC Formulation

In this chapter, we make different use of bold-faced letters and the bar as in the rest of the book to obtain consistency with existing literature on MPC for JMLS. The bar is used to denote expectancies as in $\bar{x}(k) := \mathrm{E}[x(k)]$. Bold-faced letters are only used for the aggregation of quantities over the whole prediction horizon, e.g.: $\mathbf{u}(k) := \left[u^\mathsf{T}(k), \ldots, u^\mathsf{T}(k+N-1)\right]^\mathsf{T}$ and not for vectors and matrices in general.

System Dynamics Let a discrete-time JMLS be defined by:

$$\mathscr{G} := \begin{cases} x(k+1) = A_{\theta(k)}\, x(k) + B_{\theta(k)}\, u(k) + G_{\theta(k)}\, w(k) \\ \mathscr{M} = \left(\Theta, P, \mu(0)\right) \end{cases}, \tag{1}$$

where $x(k) \in \mathbb{R}^{n_\mathrm{x}}$ denotes the system states, $u(k) \in \mathbb{R}^{n_\mathrm{u}}$ the inputs and $w(k) \in \mathbb{R}^{n_\mathrm{w}}$ the disturbance inputs at time instant $t_0 + kT$ with $t_0 \in \mathbb{R}^{\geq 0}$, $T \in \mathbb{R}^{>0}$ and $k \in \mathbb{N}$. The state of the Markov chain $\mathscr{M}$ and its probability distribution are denoted by $\theta(k) \in \Theta \subset \mathbb{N}$ with $n_\theta := |\Theta|$ and $\mu_i(k) := \mathbf{P}(\theta(k) = i)$, respectively. The system matrices are selected by $\theta(k)$ from the matrix sets $A = \{A_i : i \in \Theta\}$, $B = \{B_i : i \in \Theta\}$ and $G = \{G_i : i \in \Theta\}$. The transition probabilities are defined by $p_{i,m} := \mathbf{P}\big(\theta(k+1) = i \mid \theta(k) = m\big)$ for $i, m \in \Theta$. All transition probabilities form the transition matrix $P := [p_{i,m}]_{i,m \in \Theta}$.

Assumption 1 The continuous state $x(k)$ and the probability distribution $\mu(k)$ are measurable. In addition, it is assumed that the expectancy $\bar{w}(k)$ of the noise process is known and that the process is independent of $x(k)$ and $\theta(k)$.

The measurability assumption is widely-used in the current literature on MPC for JMLS (see for example [3, 8, 10, 14]). If this assumption is violated, the state can be estimated by the use of one of the well investigated filters for JMLS [1]. The Markov state $\theta(k)$ and its probability distribution $\mu(k)$ are often used to model component failures or link breakdowns. Since systems, as modern production lines, communication networks, or energy distribution networks are equipped with highly developed failure detection mechanisms, $\theta(k)$ or $\mu(k)$ can be measured in most cases. Examples for disturbances affecting such systems can be ambient temperature, humidity, vibrations, or load torque. Expected values are typically known for most of these disturbances. Thus, for many applications, particularly in the context of production systems and communication networks, Assumption 1 is justified.

MPC Formulation Assume that recursive application of (1) can be used to predict the expected future system states $x(k+j|k)$ at time k as a function of the future inputs $u(k+l|k)$ with $0 \leq l \leq k$. For the sake of a brief notation, quantities predicted by j time steps in the future at time step k, are abbreviated as follows $x[\![j]\!] := x(k+j|k)$. The same notation is employed for matrices utilized for the prediction, e.g. $M[\![j]\!]$.

Based on the prediction model, MPC seeks to determine an optimal input trajectory $\mathbf{u}(k)$ for the prediction horizon N based on the information available at time step k. To this end, the following optimization problem is solved at time step k:

$$\min_{\mathbf{u}(k)} \quad \mathrm{E}\left[\sum_{j=1}^{N}\left(x^{\mathsf{T}}[\![j]\!]\, Q_{\theta[\![j]\!]}\, x[\![j]\!] + u^{\mathsf{T}}[\![j-1]\!]\, R_{\theta[\![j-1]\!]}\, u[\![j-1]\!]\right)\right] \tag{2a}$$

$$\text{s.t.} \quad C\,u[\![j]\!] \le c \qquad \forall\, j \in \{0,\ldots,N-1\}, \tag{2b}$$

$$D\,\bar{x}[\![j]\!] \le d \qquad \forall\, j \in \{1,\ldots,N\}. \tag{2c}$$

Here, $Q_i > 0$ and $R_i > 0$ for all $i \in \Theta$. The input $u(k) = u[\![0]\!]$ is applied to the system and in time step $k+1$, the procedure is restarted with the measured state $x(k+1)$ and a prediction horizon shifted on step ahead.

Note that (2c) formulates constraints for the expectancy of the state $\bar{x}[\![j]\!]$ instead of $x[\![j]\!]$, what is a common approach when using MPC for JMLS (cf. [8, 9]). If the Markov state $\theta(k)$ is measurable, the one step prediction for $\bar{x}[\![1]\!]$ is deterministic except for the disturbance $w(k)$. Thus, if bounds for the disturbance $w(k)$ are known, which is typically the case for many applications, the constraints for the first prediction step $Dx(k+1) \le d$ become hard constraints. Due to the repeated application of the optimization in each time step, this constraint is fulfilled at every instant k.

To solve (2) efficiently, the problem is reformulated as a standard QP problem. To this end, the cost function and the expectancy of the states, have to be predicted as a function of the input trajectory $\mathbf{u}(k)$. If these would be calculated by directly predicting the states for all possible Markov trajectories, the computational complexity would be $\mathcal{O}(n_\theta^{N+1})$. How the resulting complexity can be reduced by an efficient recursive procedure is shown in the following sections.

3 Prediction of the Expected Value of the State

To be able to specify constraints for the expectancy of the state, a prediction scheme for $\bar{x}[\![j]\!]$ has to be stated. The deduction of the equations was already presented in [24] by the authors. To ease the understanding of the following material, the main results described in Sect. III of [24] are restated in a condensed form in this section.

3.1 Prediction Equation in Matrix Form

Let us define the following conditional expectancies:

$$\bar{x}_i[\![j]\!] := \mathrm{E}\left[x(k+j\,|\,k)\,|\,\theta(k+j-1) = i)\right], \tag{3}$$

describing the predicted expectancy of the state conditioned by the value of the Markov state $\theta(k+j-1)$.

Remark 1 This definition differs from definitions in [1] and other papers: the predicted state for time step $k+j$ is not conditioned on the Markov state in $k+j$ but one time step before. This conditioning is used since $x(k+j)$ is determined by the system matrices selected by $\theta(k+j-1)$.

By means of the indicator function:

$$\mathbb{1}_{\{\theta(k)=i\}} := \begin{cases} 0, & \text{if } \theta(k) \neq i \\ 1, & \text{if } \theta(k) = i \end{cases}, \tag{4}$$

the conditional expectancy $\bar{x}_i[\![j]\!]$ can be expressed as follows [1]:

$$\bar{x}_i[\![j]\!] := \mathrm{E}\left[x[\![j]\!]\ \mathbb{1}_{\{\theta[\![j-1]\!]=i\}}\right]. \tag{5}$$

Despite the different definition of the conditional expectancy, the following relationship between expected value and conditional expectancies from [1] holds:

$$\bar{x}[\![j]\!] = \sum_{i=1}^{n_\theta} \bar{x}_i[\![j]\!]. \tag{6}$$

In order to express $\bar{x}[\![j]\!]$ as a function of the inputs as well as the expected disturbances, the following prediction equation is used:

$$\bar{x}[\![j]\!] = \bar{A}[\![j]\!]\, x(k) + \sum_{l=0}^{j-1} \left(\bar{B}^{(l)}[\![j]\!]\, u[\![l]\!] + \bar{G}^{(l)}[\![j]\!]\, \bar{w}[\![l]\!]\right). \tag{7}$$

To determine the prediction matrices $\bar{A}[\![j]\!]$, $\bar{B}[\![j]\!]$ and $\bar{G}[\![j]\!]$, the conditioned prediction matrices $\tilde{A}_i[\![j]\!]$, $\tilde{B}_i^{(l)}[\![j]\!]$ and $\tilde{G}_i^{(l)}[\![j]\!]$ are introduced. Here, matrices marked with a bar describe the prediction of the expected state. Matrices marked with a tilde are used to describe the prediction of the conditional expectancies of the state. The latter depend on $\theta[\![j-1]\!] = i$ and describe the influence of the current state $x(k)$, the inputs $u[\![l]\!]$ and the disturbances $\bar{w}[\![l]\!]$ with $l \in \{0, \ldots, j-1\}$ on the conditional expectancy $\bar{x}_i[\![j]\!]$, respectively:

$$\bar{x}_i[\![j]\!] = \tilde{A}_i[\![j]\!]\, x(k) + \sum_{l=0}^{j-1} \left(\tilde{B}_i^{(l)}[\![j]\!]\, u[\![l]\!] + \tilde{G}_i^{(l)}[\![j]\!]\, \bar{w}[\![l]\!]\right). \tag{8}$$

How these prediction matrices can be calculated is shown in the following subsection. Due to (6), $\bar{x}[\![j]\!]$ can be determined by summation of the conditional expectancies. Thus, the desired prediction equation (7) can be determined as follows:

$$\bar{x}[\![j]\!] = \sum_{i=1}^{n_\theta} \bar{x}_i[\![j]\!] = \underbrace{\sum_{i=1}^{n_\theta} \tilde{A}_i[\![j]\!]}_{=:\bar{A}[\![j]\!]}\, x(k) + \sum_{l=0}^{j-1} \left(\underbrace{\sum_{i=1}^{n_\theta} \tilde{B}_i^{(l)}[\![j]\!]}_{=:\bar{B}^{(l)}[\![j]\!]}\, u[\![l]\!] + \underbrace{\sum_{i=1}^{n_\theta} \tilde{G}_i^{(l)}[\![j]\!]}_{=:\bar{G}^{(l)}[\![j]\!]}\, \bar{w}[\![l]\!] \right). \tag{9}$$

3.2 Recursive Determination of Prediction Matrices

For the sake of brief notation, the linear operator $\mathscr{V}_i(\phi)$ is introduced for a set ϕ of matrices depending on the Markov state. The operator is defined as the weighted sum over all matrices ϕ_m, weighted by transition probabilities leading to the Markov state $\theta(\cdot) = i$:

$$\mathscr{V}_i(\phi) := \sum_{m=1}^{n_\theta} p_{i,m} \cdot \phi_m \,, \qquad i \in \Theta. \tag{10}$$

Based on this definition, the following Theorem (derived in [24]) states a recursive algorithm to compute the necessary prediction matrices:

Theorem 1 *[24] Let the conditional expectancy $\bar{x}_i[\![j]\!]$ be defined as in (8). Then, the recursion for the calculation of $\tilde{A}_i[\![j]\!]$, $\tilde{B}_i^{(l)}[\![j]\!]$ and $\tilde{G}_i^{(l)}[\![j]\!]$ is initialized by:*

$$\tilde{A}_i[\![1]\!] = \mu_i(k)\, A_i, \qquad \tilde{B}_i^{(j-1)}[\![j]\!] = \mu_i[\![j-1]\!]\, B_i, \qquad \tilde{G}_i^{(j-1)}[\![j]\!] = \mu_i[\![j-1]\!]\, G_i, \tag{11}$$

for all $i \in \Theta,\ j \in \{1, \ldots, N\}$. Here, the probability distribution is predicted by $\mu[\![j]\!] = P^j\, \mu(k)$. The remaining matrices can be calculated by:

$$\tilde{A}_i[\![j+1]\!] = A_i\, \mathscr{V}_i\left(\tilde{A}[\![j]\!]\right), \qquad \tilde{B}_i^{(l)}[\![j+1]\!] = A_i\, \mathscr{V}_i\left(\tilde{B}^{(l)}[\![j]\!]\right), \qquad \tilde{G}_i^{(l)}[\![j+1]\!] = A_i\, \mathscr{V}_i\left(\tilde{G}^{(l)}[\![j]\!]\right) \tag{12}$$

for all $i \in \Theta,\ j \in \{1, \ldots, N-1\}$ and $l \in \{0, \ldots, j-1\}$.

Proof The theorem is proved by induction. Starting point is the prediction of the expectancy for $j = 1$. By inserting the system dynamics (1) for $x(k+1|k) = x[\![1]\!]$ into the definition of the conditional expectancy (5) for $j = 1$, it follows that[1]:

$$\bar{x}_i[\![1]\!] = \mathrm{E}\left[x[\![1]\!]\ \mathbb{1}_{\{\theta(k)=i\}}\right] = \mathrm{E}\left[(A_i\, x(k) + B_i\, u[\![0]\!] + G_i\, w[\![0]\!])\ \mathbb{1}_{\{\theta(k)=i\}}\right]. \tag{13}$$

Since $\theta(k)$, $w[\![0]\!]$ and $x(k)$ are mutually stochastically independent, the expectancy can be calculated by calculating the expectancies of all factors separately:

$$\bar{x}_i[\![1]\!] = A_i\, x(k)\, \mathrm{E}\left[\mathbb{1}_{\{\theta(k)=i\}}\right] + B_i\, u[\![0]\!]\, \mathrm{E}\left[\mathbb{1}_{\{\theta(k)=i\}}\right] + G_i\, \mathrm{E}\left[w[\![0]\!]\right] \mathrm{E}\left[\mathbb{1}_{\{\theta(k)=i\}}\right]. \tag{14}$$

It follows from (4) that $\mathrm{E}\left[\mathbb{1}_{\{\theta(k+j)=i\}}\right] = \mu_i[\![j]\!]$. Hence, Eq. (14) equals:

$$\bar{x}_i[\![1]\!] = \mu_i(k)\, A_i\, x(k) + \mu_i(k)\, B_i\, u[\![0]\!] + \mu_i(k)\, G_i\, \bar{w}[\![0]\!]. \tag{15}$$

[1]The following equations differ from other work, like [1], due to the different definition of the conditional expectancies $\bar{x}_i[\![j]\!]$ (cf. Remark 1).

The expressions $\mu_i(k)\,A_i$, $\mu_i(k)\,B_i$ and $\mu_i(k)\,G_i$ coincide with the initializations in (11). For a general prediction step $j+1$ with $j>0$, one gets the following equations by analog derivations as for $j=1$:

$$\begin{aligned}\bar{x}_i[\![j+1]\!] &= \mathrm{E}\left[x[\![j+1]\!]\cdot \mathbb{1}_{\{\theta[\![j]\!]=i\}}\right] = \mathrm{E}\left[\left(A_i\,x[\![j]\!]+B_i\,u[\![j]\!]+G_i\,w[\![j]\!]\right)\cdot \mathbb{1}_{\{\theta[\![j]\!]=i\}}\right]\\ &= \mathrm{E}\left[A_i\,x[\![j]\!]\;\mathbb{1}_{\{\theta[\![j]\!]=i\}}\right] + \underbrace{\mu_i[\![j]\!]B_i}_{=:\tilde{B}_i^{(j)}[\![j+1]\!]}\,u[\![j]\!] + \underbrace{\mu_i[\![j]\!]G_i}_{=:\tilde{G}_i^{(j)}[\![j+1]\!]}\,\bar{w}[\![j]\!]. \end{aligned} \tag{16}$$

It holds due to the Markov property (cf. [6]):

$$\mathrm{E}\left[\mathbb{1}_{\{\theta(k+j)=i\}}\right] = \mathrm{E}\left[\sum_{m=1}^{n_\theta} p_{i,m}\;\mathbb{1}_{\{\theta(k+j-1)=m\}}\right]. \tag{17}$$

With this relationship, the first term in Eq. (16) can be written as:

$$\mathrm{E}\left[A_i\,x[\![j]\!]\;\mathbb{1}_{\{\theta[\![j]\!]=i\}}\right] = \mathrm{E}\left[A_i\,x[\![j]\!]\sum_{m=1}^{n_\theta} p_{i,m}\;\mathbb{1}_{\{\theta[\![j-1]\!]=m\}}\right] = A_i\sum_{m=1}^{n_\theta} p_{i,m}\;\bar{x}_m[\![j]\!]. \tag{18}$$

Due to definition of $\bar{x}_i[\![j]\!]$ in (5), one can insert (8) in (18) to obtain:

$$\begin{aligned}\mathrm{E}\left[A_i\,x[\![j]\!]\;\mathbb{1}_{\{\theta[\![j]\!]=i\}}\right] &= A_i\sum_{m=1}^{n_\theta} p_{i,m}\left[\tilde{A}_m[\![j]\!]\,x(k)+\sum_{l=0}^{j-1}\left(\tilde{B}_m^{(l)}[\![j]\!]\,u[\![l]\!]+\tilde{G}_m^{(l)}[\![j]\!]\,\bar{w}[\![l]\!]\right)\right]\\ &= \underbrace{A_i\,\mathscr{V}_i\left(\tilde{A}[\![j]\!]\right)}_{=:\tilde{A}_i[\![j+1]\!]}\,x(k)+\sum_{l=0}^{j-1}\left[\underbrace{A_i\,\mathscr{V}_i\left(\tilde{B}^{(l)}[\![j]\!]\right)}_{=:\tilde{B}_i^{(l)}[\![j+1]\!]}\,u[\![l]\!]+\underbrace{A_i\,\mathscr{V}_i\left(\tilde{G}^{(l)}[\![j]\!]\right)}_{=:\tilde{G}_i^{(l)}[\![j+1]\!]}\,\bar{w}[\![l]\!]\right]. \end{aligned} \tag{19}$$

The labelled terms coincide with the recursions in (12). Thus, (19) completes the induction if the prediction matrices are chosen as stated in Theorem 1. □

By the recursive prediction established by Theorem 1, all transitions leading to the same Markov state i are combined and stored in $\bar{x}_i[\![j]\!]$ in each recursion. Hence, only all combinations of $i_1\in\Theta$ and $i_2\in\Theta$ have to be considered for the prediction from $\bar{x}_{i_1}[\![j]\!]$ to $\bar{x}_{i_2}[\![j+1]\!]$. Thus, not the whole set of possible Markov trajectories from k to $k+j$ have to be considered, since these are condensed in the set $\{\bar{x}_{i_1}[\![j]\!]\}$. Therefore, the proposed approach reduces the computational complexity to $\mathscr{O}\left(n_\theta^2\cdot N^2\right)$, compared to the direct calculation of all trajectories with the complexity $\mathscr{O}\left(n_\theta^N\right)$.

4 Reformulation and Computation of the Cost Function

This section shows how the cost function (2a) can be formulated as a function of inputs $u[\![j]\!]$, initial state $x(0)$ and expected disturbances $\bar{w}[\![l]\!]$.

The expected costs can be obtained by calculating the expected costs for each prediction step separately. Since the expectancy of a sum equals the sum of the expectancies of the summands and the inputs $u[\![j]\!]$ are deterministic quantities determined by the optimization, (2a) equals:

$$J(k) := \sum_{j=1}^{N} \Big(\mathrm{E}\left[x^{\mathsf{T}}[\![j]\!]\, Q_{\theta[\![j]\!]}\, x[\![j]\!]\right] + u^{\mathsf{T}}[\![j-1]\!]\, \mathrm{E}\left[R_{\theta[\![j-1]\!]}\right]\, u[\![j-1]\!] \Big). \tag{20}$$

Since the expression is a quadratic form of $x[\![j]\!]$ (which depends on $x(k)$, $u[\![l]\!]$ and $w[\![l]\!]$ with $l < j$), the expected cost can be expressed by quadratic and bi-linear terms of $x(0)$, $u[\![l]\!]$ and $w[\![l]\!]$. The goal is to transform the expectancy in the form:

$$\mathrm{E}\left[x^{\mathsf{T}}[\![j]\!]\; Q_{\theta[\![j]\!]}\; x[\![j]\!]\right] = \mathbf{u}^{\mathsf{T}}(k)\, W'[\![j]\!]\, \mathbf{u}(k) + \left(q_{\mathrm{x}}[\![j]\!] + q_{\mathrm{w}}[\![j]\!]\right)\, \mathbf{u}(k) + \Psi. \tag{21}$$

Here, q_{x} and q_{w} describe the parts that depend on $x(k)$ and $\mathbf{u}(k)$, as well as on $\bar{\mathbf{w}}(k)$ and $\mathbf{u}(k)$, respectively. The variable Ψ contains all costs that cannot be influenced by the inputs (i.e. only depends on $x(k)$ and $\mathbf{w}(k)$). These costs will be neglected in the following, since they do not affect the optimal input trajectory $\mathbf{u}(k)$.

Remark 2 Many $W'[\![j]\!]$, q_{x} and q_{w} are sparse, since the step costs at $k+j$ can be influenced only by the inputs $u[\![l]\!]$ with $l < j$. The resulting structure is as follows:

$$W'[\![1]\!] = \begin{bmatrix} * & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{(N-1)\cdot n_{\mathrm{u}}} \end{bmatrix}, \quad W'[\![2]\!] = \begin{bmatrix} * & * & \mathbf{0} \\ * & * & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{(N-2)\cdot n_{\mathrm{u}}} \end{bmatrix}, \quad \dots \quad W'[\![N]\!] = \begin{bmatrix} * & \cdots & * \\ \vdots & \ddots & \vdots \\ * & \cdots & * \end{bmatrix}$$

$$q_{\mathrm{x}}[\![1]\!] = \begin{bmatrix} * & \mathbf{0}_{1\times(N-1)\cdot n_{\mathrm{u}}} \end{bmatrix}, \quad q_{\mathrm{x}}[\![2]\!] = \begin{bmatrix} * & * & \mathbf{0}_{1\times(N-2)\cdot n_{\mathrm{u}}} \end{bmatrix}, \quad \dots \quad q_{\mathrm{x}}[\![N]\!] = \begin{bmatrix} * & \cdots & * \end{bmatrix},$$

where $\mathbf{0}$ denotes zero matrices of indicated or appropriate dimensions. The $*$ denotes arbitrary matrices, dimensioned according to the size of $x(0)$, $u[\![l]\!]$ and $w[\![l]\!]$. □

The following demonstrates how $W'[\![j]\!]$, $q_{\mathrm{x}}[\![j]\!]$ and $q_{\mathrm{w}}[\![j]\!]$ can be determined recursively. To illustrate the basic idea, the calculation is demonstrated for $j = 1$ and generalized afterwards. Using the fact that costs are scalar, it follows that:

$$\begin{aligned}
&\mathrm{E}\left[x^{\mathsf{T}}[\![1]\!]\, Q_{\theta[\![1]\!]}\, x[\![1]\!]\right] \\
&\quad = \mathrm{E}\Big[\left(A_{\theta(k)}\, x(k) + B_{\theta(k)}\, u[\![0]\!] + G_{\theta(k)}\, w[\![0]\!]\right)^{\mathsf{T}}\, Q_{\theta[\![1]\!]}\, \left(A_{\theta(k)}\, x(k) + B_{\theta(k)}\, u[\![0]\!] + G_{\theta(k)}\, w[\![0]\!]\right) \Big] \\
&\quad = \mathrm{E}\Big[u^{\mathsf{T}}[\![0]\!]\, B_{\theta(k)}^{\mathsf{T}}\, Q_{\theta[\![1]\!]}\, B_{\theta(k)}\, u[\![0]\!] + 2x^{\mathsf{T}}(k)\, A_{\theta(k)}^{\mathsf{T}}\, Q_{\theta[\![1]\!]}\, B_{\theta(k)}\, u[\![0]\!] \\
&\qquad\quad + 2w^{\mathsf{T}}[\![0]\!]\, B_{\theta(k)}^{\mathsf{T}}\, Q_{\theta[\![1]\!]}\, G_{\theta(k)}\, u[\![0]\!] \Big] + \Psi.
\end{aligned} \tag{22}$$

One obtains by calculating the cost for every possible combination of the Markov states $\theta(k)$ and $\theta(k+1)$ and the corresponding transition probabilities:

$$\begin{aligned}\mathrm{E}\left[x^\mathsf{T}[\![1]\!]\, Q_{\theta[\![1]\!]}\, x[\![1]\!]\right] &= \sum_{i=1}^{n_\theta}\sum_{m=1}^{n_\theta} \mu_i(k)\, p_{m,i}\big(u^\mathsf{T}[\![0]\!]\, B_i^\mathsf{T}\, Q_m\, B_i\, u[\![0]\!] \\ &\quad + 2x^\mathsf{T}(k)\, A_i^\mathsf{T}\, Q_m\, B_i\, u[\![0]\!] + 2\bar{w}^\mathsf{T}[\![0]\!]\, B_m^\mathsf{T}\, Q_m\, G_i\, u[\![0]\!]\big) + \Psi \\ &= \sum_{i=1}^{n_\theta} \mu_i(k)\, \big(u^\mathsf{T}[\![0]\!]\, B_i^\mathsf{T}\, \mathscr{T}_i(Q)\, B_i\, u[\![0]\!] + 2x^\mathsf{T}(k)\, A_i^\mathsf{T}\, \mathscr{T}_i(Q)\, B_i\, u[\![0]\!] \\ &\quad + 2\bar{w}^\mathsf{T}[\![0]\!]\, B_m^\mathsf{T}\, \mathscr{T}_i(Q)\, G_i\, u[\![0]\!]\big) + \Psi. \end{aligned} \tag{23}$$

Here, $\mathscr{T}_i$ denotes the following linear operator for the set of cost matrices Q:

$$\mathscr{T}_i(Q) := \sum_{m=1}^{n_\theta} p_{m,i} \cdot Q_m, \qquad i \in \Theta. \tag{24}$$

In contrast to $\mathscr{V}(\cdot)$, this operator uses all probabilities of transitions emanating from the Markov state $\theta(\cdot) = i$ as weights.

The following matrices are introduced to formulate (23) as a function of $\mathbf{u}(k)$:

$$\begin{aligned} \hat{B}_i[1] &:= \begin{bmatrix} B_i & \mathbf{0} \\ \mathbf{0} & I_{(N-1)\cdot n_\mathrm{u}} \end{bmatrix}, \quad \hat{G}_i[1] := \begin{bmatrix} G_i & \mathbf{0} \\ \mathbf{0} & I_{(N-1)\cdot n_\mathrm{w}} \end{bmatrix}, \quad \hat{Q}_{\mathrm{W},i}[1] := \begin{bmatrix} \mathscr{T}_i(Q) & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{(N-1)\cdot n_\mathrm{u}} \end{bmatrix}, \\ \hat{Q}_{\mathrm{q_w},i}[1] &:= \begin{bmatrix} \mathscr{T}_i(Q) & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{(N-1)n_\mathrm{w}\times(N-1)n_\mathrm{u}} \end{bmatrix}, \quad \hat{Q}_{\mathrm{q_x},i}[1] := \begin{bmatrix} \mathscr{T}_i(Q) & \mathbf{0}_{n_\mathrm{x}\times(N-1)n_\mathrm{u}} \end{bmatrix}. \end{aligned} \tag{25}$$

Here, I_n denotes the identity matrix of dimension $n \times n$. The indices W, $\mathrm{q_w}$ and $\mathrm{q_x}$ illustrate the relation of the corresponding matrices and the quadratic and bi-linear terms in (23). Thus, it follows with these definitions:

$$\begin{aligned}\mathrm{E}\left[x^\mathsf{T}[\![1]\!]\; Q_{\theta[\![1]\!]}\; x[\![1]\!]\right] &= \sum_{i=1}^{n_\theta} \mu_i(k)\Big(\mathbf{u}^\mathsf{T}(k)\; \hat{B}_i^\mathsf{T}[1]\; \hat{Q}_{\mathrm{W},i}[1]\; \hat{B}_i[1]\; \mathbf{u}(k) \\ &\quad + 2x^\mathsf{T}(k)\; A_i^\mathsf{T}\; \hat{Q}_{\mathrm{q_x},i}[1]\; \hat{B}_i[1]\; \mathbf{u}(k) + 2\bar{\mathbf{w}}^\mathsf{T}(k)\; \hat{G}_i^\mathsf{T}[1]\; \hat{Q}_{\mathrm{q_w},i}[1]\; \hat{B}_i[1]\; \mathbf{u}(k)\Big) + \Psi. \end{aligned} \tag{26}$$

By comparing coefficients with (21), the equations for the cost matrices follow as:

$$W'[\![1]\!] := \sum_{i=1}^{n_\theta} \mu_i(k)\; \hat{B}_i^\mathsf{T}[1]\; \hat{Q}_{\mathrm{W},i}[1]\; \hat{B}_i[1], \tag{27}$$

$$q_\mathrm{x}[\![1]\!] := 2x^\mathsf{T}(k) \sum_{i=1}^{n_\theta} \mu_i(k)\; A_i^\mathsf{T}\; \hat{Q}_{\mathrm{q_x},i}[1]\; \hat{B}_i[1], \tag{28}$$

$$q_\mathrm{w}[\![1]\!] := 2\bar{\mathbf{w}}^\mathsf{T}(k) \sum_{i=1}^{n_\theta} \mu_i(k)\; \hat{G}_i^\mathsf{T}[1]\; \hat{Q}_{\mathrm{q_w},i}[1]\; \hat{B}_i[1]. \tag{29}$$

Due to the special structure of the matrices defined in (25), the resulting prediction matrices have the structure described in Remark 1.

The equations for $j > 1$ become quite lengthy. Hence, here just the principle of deducing the formulas is described and the general result in terms of an recursive algorithm is stated in Theorem 2. For a detailed deduction of the formulas, the reader is referred to the proof in Appendix A.

The underling principle for the deduction of equations for $j = 1$ can be generalized as follows: First, the expected costs $\mathrm{E}\left[x^{\mathsf{T}}[\![j]\!]\, Q_{\theta[\![j]\!]}\, x[\![j]\!]\right]$ have to be calculated explicitly, like in (23), by summing up the costs for all possible Markov state trajectories $(\theta(k), \ldots, \theta(k+1))$ considering the corresponding occurrence probabilities. This involves computing $j+1$ sums over all n_θ^{j+1} possible Markov state trajectories. To reduce the exponentially growing calculation effort, the Markov property can be employed to reorder the sums. For the resulting nested sums, a recursive calculation scheme for $W'[\![j]\!]$, $q_{\mathrm{x}}[\![j]\!]$, and $q_{\mathrm{w}}[\![j]\!]$ can be deduced, resulting in equations like (27)–(29). An algorithm implementing this principle is given in the following theorem:

Theorem 2 *Let the recursion matrices be defined for all $l \in \{1, \ldots, N\}$ and $i \in \Theta$:*

$$\hat{Q}_{\mathrm{q_x},i}[j] := \big[\underbrace{\mathscr{T}_i(Q) \;\; \ldots \;\; \mathscr{T}_i(Q)}_{j} \;\; \mathbf{0}_{n_{\mathrm{x}} \times (N-j)\cdot n_{\mathrm{u}}}\big], \tag{30}$$

$$\hat{Q}_{\mathrm{W},i}[l] := \begin{bmatrix} \mathbf{1}_l \otimes \mathscr{T}_i(Q) & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{(N-l)\cdot n_{\mathrm{u}}} \end{bmatrix}, \qquad \hat{Q}_{\mathrm{q_w},i}[l] := \begin{bmatrix} \mathbf{1}_l \otimes \mathscr{T}_i(Q) & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{(N-l)\cdot n_{\mathrm{w}} \times (N-l)\cdot n_{\mathrm{u}}} \end{bmatrix},$$

$$\hat{B}_i[l] := \mathrm{diag}\big(\underbrace{A_i, \ldots, A_i}_{l-1}, B_i, I_{(N-l)\cdot n_{\mathrm{u}}}\big), \qquad \hat{G}_i[l] := \mathrm{diag}\big(\underbrace{A_i, \ldots, A_i}_{l-1}, G_i, I_{(N-l)\cdot n_{\mathrm{w}}}\big).$$

Here, $\otimes$ represents the Kronecker product and $\mathbf{1}_l$ denotes an $l \times l$-matrix of ones. Then, $W'[\![j]\!]$, $q_x[\![j]\!]$ and $q_w[\![j]\!]$ can be calculated with the following algorithm:

1. ***Initialization**: Set $m = 1$ and calculate:*

$$\chi_i^{(1)} := A_i^{\mathsf{T}}\; \hat{Q}_{\mathrm{q_x},i}[j]\; \hat{B}_i[j], \qquad \kappa_i^{(1)} := \hat{B}_i^{\mathsf{T}}[j]\; \hat{Q}_{\mathrm{W},i}[j]\; \hat{B}_i[j] \quad \forall i \in \Theta, \tag{31}$$
$$\zeta_i^{(1)} := \hat{G}_i^{\mathsf{T}}[j]\; \hat{Q}_{\mathrm{q_w},i}[j]\; \hat{B}_i[j] \quad \forall i \in \Theta.$$

2. ***Recursion**:*

   (a) If $m < j$:

$$\chi_i^{(m+1)} := A_i^{\mathsf{T}}\; \mathscr{T}_i\left(\chi^{(m)}\right)\; \hat{B}_i[j-m] \quad \forall i \in \Theta, \tag{32a}$$
$$\kappa_i^{(m+1)} := \hat{B}_i^{\mathsf{T}}[j-m]\; \mathscr{T}_i\left(\kappa^{(m)}\right)\; \hat{B}_i[j-m] \quad \forall i \in \Theta, \tag{32b}$$
$$\zeta_i^{(m+1)} := \hat{G}_i^{\mathsf{T}}[j-m]\; \mathscr{T}_i\left(\zeta^{(m)}\right)\; \hat{B}_i[j-m] \quad \forall i \in \Theta, \tag{32c}$$

   else: go to 3.

   (b) Set $m := m+1$ and go to (a).

*3. **Calculation of** $q_{\mathrm{x}}[\![j]\!]$, $W'[\![j]\!]$, **and** $q_{\mathrm{w}}[\![j]\!]$:*

$$q_{\mathrm{x}}[\![j]\!] := 2x^{\intercal}(k) \sum_{i=1}^{n_\theta} \mu_i(k)\ \chi_i^{(j)}, \qquad W'[\![j]\!] := \sum_{i=1}^{n_\theta} \mu_i(k)\ \kappa_i^{(j)}, \tag{33}$$

$$q_{\mathrm{w}}[\![j]\!] := 2\bar{\mathbf{w}}^{\intercal}(k) \sum_{i=1}^{n_\theta} \mu_i(k)\ \zeta_i^{(j)}.$$

Proof See Appendix. □

It can be seen from the recursion equations and the deduction in the proof that the cost prediction matrices can be constructed recursively from inside to outside . This is necessary to guarantee that left and right matrix multiplications are performed for matrices associated with the same Markov state. Due to this, the construction of $W'[\![j]\!]$, $q_{\mathrm{x}}[\![j]\!]$ as well as $q_{\mathrm{w}}[\![j]\!]$ begins with the system matrices for the Markov state $\theta(k+j-1)$ and moves back in time to $\theta(k)$ through the recursion. Finally, the current information at time k, e.g. current state $x(k)$ and probability distribution $\mu(k)$ are used in the last step to calculate the cost matrices. Consequently, the initialization (1) and recursion (2) describe the evolution of the costs and step (3) links this evolution with the current information about continuous state and Markov state.

From the equation for the calculation of $q_{\mathrm{w}}[\![j]\!]$ in (33), it follows that a disturbance with expectancy zero ($\bar{\mathbf{w}}(k) = \mathbf{0}$) does not affect any terms that can be manipulated by the inputs $\mathbf{u}(k)$. Thus, the disturbance has no influence on the input trajectory $\mathbf{u}(k)$ and the matrices $q_w[\![j]\!]$, $\zeta_i^{(m)}$, $\hat{Q}_{q_w,i}[l]$ and $\hat{G}_i[l]$ are not needed.

Computational Complexity of the Recursion Algorithm The computational effort is larger than for the prediction of the expectancy of the state presented in Theorem 1 due to the high dimension of the prediction matrices in (30) and the recursion in (1) and (2). However, the special structure used in Theorem 2 has the advantage that the steps (1) and (2), which dominate the computational effort, are independent of the current information. Provided that the transition probabilities $p_{i,m}$ and the system matrices A_i, B_i and G_i are time invariant, these steps can be performed offline. Just the final calculation (3) of the cost prediction matrices has to be performed online, if the matrices $\chi_i^{(j)}$, $\kappa_i^{(j)}$ and $\zeta_i^{(j)}$ are calculated offline and stored. Hence, for the construction of the cost function, very few computations have to be performed online prior to solving the optimization problem. Thus, the presented prediction algorithm reduces the originally exponentially increasing computational effort drastically, without any truncation of the scenario tree.

One downside of the presented approach is that high powers of the system matrices result for large N. This can lead to numerical problems due to ill-conditioned matrices, similar as for the standard condensed MPC formulation of LTI systems [25].

Remark 3 The calculation time can be reduced considerably if the cost matrices $W'[\![j]\!]$, $q_{\mathrm{x}}[\![j]\!]$ and $q_{\mathrm{w}}[\![j]\!]$ are divided into block matrices according to the inputs $u[\![j]\!]$ for each time step and calculated separately. This reduces the dimension of the multiplied matrices and avoids multiplications with identity and zero matrices.

5 Formulation as Quadratic Optimization Problem

This section demonstrates how the results from Sects. 3 and 4 can be used to formulate the optimization problem (2) as a QP in the following form:

$$\min_{\mathbf{u}(k)} \quad \mathbf{u}^\intercal(k) W(k) \mathbf{u}(k) + q(k) \mathbf{u}(k) \tag{34a}$$

$$\text{s.t.} \quad H\mathbf{u}(k) \leq h. \tag{34b}$$

To this end, the prediction equations, the cost function, and the constraints have to be reformulated as a function of $\mathbf{u}(k)$.

The augmented state vector $\bar{\mathbf{x}}[\![1]\!] := \left[\bar{x}^\intercal[\![1]\!] \ \ldots \ \bar{x}^\intercal[\![N]\!]\right]^\intercal$ containing all expected values of the states for the whole prediction horizon and Eq. (7) are used to formulate the prediction equations:

$$\bar{\mathbf{x}}[\![1]\!] = \mathbf{A}(k)\, x(k) + \mathbf{B}(k)\, \mathbf{u}(k) + \mathbf{G}(k)\, \bar{\mathbf{w}}(k) \tag{35}$$

with:

$$\mathbf{A}(k) := \begin{bmatrix} \bar{A}[\![1]\!] \\ \bar{A}[\![2]\!] \\ \vdots \\ \bar{A}[\![N]\!] \end{bmatrix}, \qquad \mathbf{B}(k) := \begin{bmatrix} \bar{B}^{(0)}[\![1]\!] & 0 & \ldots & 0 \\ \bar{B}^{(0)}[\![2]\!] & \bar{B}^{(1)}[\![2]\!] & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \bar{B}^{(0)}[\![N]\!] & \ldots & \bar{B}^{(N-2)}[\![N]\!] & \bar{B}^{(N-1)}[\![N]\!] \end{bmatrix}, \tag{36}$$

$$\mathbf{G}(k) := \begin{bmatrix} \bar{G}^{(0)}[\![1]\!] & 0 & \ldots & 0 \\ \bar{G}^{(0)}[\![2]\!] & \bar{G}^{(1)}[\![2]\!] & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \bar{G}^{(0)}[\![N]\!] & \ldots & \bar{G}^{(N-2)}[\![N]\!] & \bar{G}^{(N-1)}[\![N]\!] \end{bmatrix}.$$

Once the prediction matrices have been calculated according to Theorem 1, the prediction of the expected values of the state $\bar{x}[\![j]\!]$ is a simple matrix-vector multiplication. In addition, $\bar{\mathbf{x}}[\![1]\!]$ is a linear function of $\mathbf{u}(k)$.

Since $R_{\theta[\![j]\!]}$ does only depend on the current Markov state, the cost function can be rewritten by means of (20) and (21):

$$\begin{aligned} J(k) &= \sum_{j=1}^{N} \left(\mathrm{E}\left[x^\intercal[\![j]\!]\ Q_{\theta[\![j]\!]}\ x[\![j]\!]\right] + u^\intercal[\![j-1]\!] \left(\sum_{i=1}^{n_\theta} \mu_i[\![j-1]\!]\ R_i \right) u[\![j-1]\!] \right) \\ &= \sum_{j=1}^{N} \left(\mathbf{u}^\intercal(k)\, W'[\![j]\!]\, \mathbf{u}(k) + \left(q_\mathrm{x}[\![j]\!] + q_\mathrm{w}[\![j]\!] \right) \mathbf{u}(k) \right) + \mathbf{u}^\intercal(k)\, \mathbf{R}(k)\, \mathbf{u}(k) + \Psi. \end{aligned} \tag{37}$$

The input cost matrix $\mathbf{R}(k)$ is defined as the block diagonal matrix aggregating all expected values of the cost matrices $R_{\theta[\![j]\!]}$:

$$\mathbf{R}(k) := \operatorname{diag}\left(\sum_{i=1}^{n_\theta} \mu_i[\![0]\!] \cdot R_i, \ldots, \sum_{i=1}^{n_\theta} \mu_i[\![N-1]\!] \cdot R_i\right). \tag{38}$$

The following matrices are introduced in order to write the constraints (2b) and (2c) as a function of $\mathbf{u}(k)$:

$$\mathbf{c} := \begin{bmatrix} c^\mathsf{T} & \ldots & c^\mathsf{T} \end{bmatrix}^\mathsf{T}, \quad \mathbf{d} := \begin{bmatrix} d^\mathsf{T} & \ldots & d^\mathsf{T} \end{bmatrix}^\mathsf{T}, \quad \mathbf{C} = \operatorname{diag}(C, \ldots, C), \quad \mathbf{D} = \operatorname{diag}(D, \ldots, D).$$

Using the aforementioned definitions and prediction equations, the QP can be formulated in standard form, as stated in the following theorem:

Theorem 3 *The optimization problem (2) is equivalent to the following QP:*

$$\min_{\mathbf{u}(k)} \quad \mathbf{u}^\mathsf{T}(k) W(k) \mathbf{u}(k) + q(k) \mathbf{u}(k) \tag{39a}$$

$$\text{s.t.} \quad \mathbf{C}\,\mathbf{u}(k) \leq \mathbf{c}, \qquad \mathbf{D}\,(\mathbf{A}(k)\,x(k) + \mathbf{B}(k)\,\mathbf{u}(k) + \mathbf{G}(k)\,\bar{\mathbf{w}}(k)) \leq \mathbf{d}, \tag{39b}$$

when matrices $W(k)$ and $q(k)$ are selected as follows:

$$W(k) = \sum_{j=1}^{N} W'[\![j]\!] + \mathbf{R}(k), \qquad q(k) = \sum_{j=1}^{N} \left(q_\mathrm{x}[\![j]\!] + q_\mathrm{w}[\![j]\!]\right),$$

where $W'[\![j]\!]$, $q_\mathrm{x}[\![j]\!]$ and $q_\mathrm{w}[\![j]\!]$ are determined according to Theorem 2.

Proof The absolute term Ψ in (37) can be neglected, since it does not affect the optimizer. Thus, cost function (39a) directly follows from (37). The equality of the constraints follows by the definitions of $\mathbf{C}$, $\mathbf{c}$, $\mathbf{D}$, and $\mathbf{d}$, as well as the structure of the prediction equation (35). □

Theorem 3 shows that the stated MPC problem (2) for JMLS can be transformed in a QP similar to that arising in standard condensed MPC formulations for LTI systems. Efficient solvers exist for the resulting QP. In addition, the construction of the matrices $\mathbf{A}(k)$, $\mathbf{B}(k)$, $\mathbf{G}(k)$ and $W(k)$, $q(k)$ is based on the recursive computations of the prediction matrices as stated in Theorems 1 and 2 that can be performed offline to some extent. Thus, Theorems 1–3 establish the desired MPC approach with low computational effort, compared to approaches employing SDP or MILP formulations.

Computational Complexity of the MPC The computation time results from the calculation of the prediction matrices on the one hand, and from solving (39) on the other hand. As stated above, the online computational effort grows with complexity[2]

[2]Only the dependencies on n_θ and N are stated, since the statements are just meant to demonstrate the improvement compared to the case where all n_θ^{N+1} Markov state trajectories are calculated.

$\mathcal{O}\left(n_\theta^2 \cdot N^2\right)$ for the computation of the prediction matrices $\mathbf{A}(k)$, $\mathbf{B}(k)$, and $\mathbf{G}(k)$. The computational effort of the online calculations for the cost prediction matrices $W'[\![j]\!]$, $q_\mathrm{x}[\![j]\!]$ and $q_\mathrm{w}[\![j]\!]$ increases with $\mathcal{O}\left(n_\theta \cdot N^3\right)$. The cost matrices can be calculated also by:

$$W(k) = \sum_{j=1}^{N} \sum_{i=1}^{n_\theta} \mu_i(k)\ \kappa_i^{(j)} + \mathbf{R}(k) = \sum_{i=1}^{n_\theta} \mu_i(k) \sum_{j=1}^{N} \kappa_i^{(j)} + \mathbf{R}(k), \tag{40}$$

$$q(k) = \sum_{j=1}^{N} \left(q_x[\![j]\!] + q_w[\![j]\!]\right) = 2x^\mathsf{T}(k) \sum_{i=1}^{n_\theta} \mu_i(k) \sum_{j=1}^{N} \chi_i^{(j)} + 2\bar{\mathbf{w}}^\mathsf{T}(k) \sum_{i=1}^{n_\theta} \mu_i(k) \sum_{j=1}^{N} \zeta_i^{(j)}.$$

In this formulation, the inner sums over $\kappa_i^{(j)}$, $\chi_i^{(j)}$ and $\zeta_i^{(j)}$ can be calculated offline. Hence, the online computational complexity can be reduced further to $\mathcal{O}\left(n_\theta \cdot N^2\right)$. If the Markov state is measurable, the sums over all Markov states can be replaced by just selecting the corresponding matrices. In this case, the online computational effort becomes independent of the number of Markov states.

The prediction matrices $\mathbf{A}(k)$, $\mathbf{B}(k)$ and $\mathbf{G}(k)$ are only needed to formulate the state expectancy constraints in (39b). Thus, the pre-processing effort grows with $\mathcal{O}\left(n_\theta \cdot N^2\right)$ when no state expectancy constraints are present, and with $\mathcal{O}\left(n_\theta^2 \cdot N^2\right)$ otherwise. Thus, despite the more involved reformulation of the cost function, this approach results in a significantly lower calculation time for pre-processing as the approach presented in [24], when no state constraints are considered. The computation time for solving the QP mainly depends on the number of decision variables $(N \cdot n_\mathrm{u})$, the number of constraints and the utilized solver. The resulting computation times for different examples are displayed in Sect. 7.

6 Mean Square Stability

In this section, the presented approach is investigated and compared to the approach presented in [24] concerning stability. As is common for JMLS, the notion of mean square stability (MSS) is employed here. The convergence of the first and second moments of the state $\bar{x}(k)$ and $\mathrm{E}\left[x(k)\ x^\mathsf{T}(k)\right]$ respectively, is required for MSS.

Definition 1 The controlled JMLS is MSS iff for all $x(0)$ and $\theta(0)$ and $w(k) = 0$ the following convergence properties hold [1]:

$$\|\bar{x}(k)\| \to 0 \quad \text{and} \quad \left\|\mathrm{E}\left[x(k)\ x^\mathsf{T}(k)\right]\right\| \to 0 \quad \text{as} \quad k \to \infty\,. \tag{41}$$

Concerning stability, two questions arise and are answered below: Does the convergence of the cost function (2a) to zero lead to MSS (Sect. 6.1)? Is the convergence of the cost function guaranteed by the proposed MPC approach (Sect. 6.2)?

6.1 Cost Convergence and MSS

The following result holds as a criterion for MSS:

Lemma 1 *If the costs (2a) converge to zero, i.e.* $\lim_{k\to\infty} J(k) = 0$*, MSS is implied.*

Proof $J(k) \to 0$ implies $\mathbf{u}(k) \to \mathbf{0}$ and $\mathrm{E}\left[x^\mathsf{T}[\![j]\!]\, Q_{\theta[\![j]\!]}\, x[\![j]\!]\right] \to 0$ for all $j \in \{1, \ldots, N\}$ due to Eq. (37). For the state-dependent step costs, it follows that:

$$\mathrm{E}\left[x^\mathsf{T}[\![j]\!]\, Q_{\theta[\![j]\!]}\, x[\![j]\!]\right] = \mathrm{E}\left[\mathrm{tr}(x^\mathsf{T}[\![j]\!]\, Q_{\theta[\![j]\!]}\, x[\![j]\!])\right] = \mathrm{tr}\left(\mathrm{E}\left[Q_{\theta[\![j]\!]}\, x[\![j]\!]\, x^\mathsf{T}[\![j]\!]\right]\right). \tag{42}$$

The matrices $Q_{\theta[\![j]\!]}$ and $x^\mathsf{T}[\![j]\!]$ are not independent in general, since both depend on $\mu(k)$. Hence, the expectancy of the product can be reformulated as follows:

$$\mathrm{E}\left[x^\mathsf{T}[\![j]\!]\, Q_{\theta[\![j]\!]}\, x[\![j]\!]\right] = \mathrm{tr}\left(\mathrm{E}\left[Q_{\theta[\![j]\!]}\right] \mathrm{E}\left[x[\![j]\!]\, x^\mathsf{T}[\![j]\!]\right] + \mathrm{Cov}(Q_{\theta[\![j]\!]}, x[\![j]\!]\, x^\mathsf{T}[\![j]\!])\right). \tag{43}$$

If the cost matrices are independent of the Markov state, the covariance $\mathrm{Cov}(Q, x[\![j]\!]\, x^\mathsf{T}[\![j]\!])$ is zero.

It holds $\mathrm{E}\left[Q_{\theta[\![j]\!]}\right] > 0$, since $Q_i > 0$ for all $i \in \Theta$. The second moment matrix $\mathrm{E}\left[x[\![j]\!]\, x^\mathsf{T}[\![j]\!]\right]$ and a covariance are always positive semi-definite. Hence, the argument of the trace is positive semi-definite and the whole expression is zero if and only if both summands of the argument are zero. Thus, it holds:

$$\begin{aligned} \mathrm{E}\left[x^\mathsf{T}[\![j]\!]\, Q_{\theta[\![j]\!]}\, x[\![j]\!]\right] = 0 \;&\Rightarrow\; \mathrm{E}\left[Q_{\theta[\![j]\!]}\right] \mathrm{E}\left[x[\![j]\!]\, x^\mathsf{T}[\![j]\!]\right] = \mathbf{0} \\ &\Rightarrow\; \mathrm{E}\left[x[\![j]\!]\, x^\mathsf{T}[\![j]\!]\right] = \mathrm{Cov}\left(x[\![j]\!], x[\![j]\!]\right) + \bar{x}[\![j]\!]\bar{x}^\mathsf{T}[\![j]\!] = \mathbf{0}. \end{aligned} \tag{44}$$

The last equation is only valid if the covariance and $\bar{x}[\![j]\!]$ are zero. Hence, convergence of the costs implies that the convergence of the first and the second moments to zero and MSS follows from the convergence of the cost function (2a) to zero. □

Inserting the definition of the covariance matrix (cf. [26]) into (43) results into the relation between the cost function (2a) and the cost function used in [24]:

$$\begin{aligned} \mathrm{E}\left[x^\mathsf{T}[\![j]\!]\, Q_{\theta[\![j]\!]}\, x[\![j]\!]\right] = \mathrm{tr}\Big(&\mathrm{E}\left[Q_{\theta[\![j]\!]}\right] \mathrm{Cov}\left(x[\![j]\!], x[\![j]\!]\right) + \mathrm{Cov}(Q_{\theta[\![j]\!]}, x[\![j]\!]\, x^\mathsf{T}[\![j]\!])\Big) \\ &+ \bar{x}^\mathsf{T}[\![j]\!]\, \mathrm{E}\left[Q_{\theta[\![j]\!]}\right]\, \bar{x}[\![j]\!]. \end{aligned} \tag{45}$$

In contrast to the approach in [24], the costs (2a) depend on the covariance matrix of the state vector. This means that the MPC approach proposed in this contribution seeks to minimize the expected value of the state as well as the covariance, while the approach in [24] tries to minimize only the expectancy of the state (cf. Appendix B).

Thus, Lemma 1 holds for the approach presented here, but not for the cost function in [24] in general. It is expected that the approach proposed in this contribution will lead to a better performance and is capable of stabilizing a broader class of JMLS.

6.2 Discussion of Cost Convergence

The answer to the question of whether the proposed MPC approach with minimization of (2a) subject to (1), (2b), and (2c) leads to converging costs (implying MSS according to Lemma 1) depends on the system structure, the cost matrices Q_i and R_i, and the horizon length N. If the inputs are penalized too much by high values of R_i compared to Q_i, the computed inputs can be too weak to counteract the uncertainties and disturbances caused by the Markovian switching of the system dynamics. On the other hand, in some rare cases where the inputs are penalized insufficiently by low values of R_i compared to Q_i, the computed inputs may run into the constraints and lead to infeasible optimization problems. With respect to the prediction horizon, an increase of N leads to additional information about the system available to the MPC and a better controller performance. However, due to the increasing uncertainties with increasing prediction horizon, the variance of the predicted states increases. The variance becomes the dominating part of the cost function for larger values of N. Thus, the resulting control actions become more conservative in order to minimize the variance in the system instead of the expected value of the states, and the convergence deteriorates. Hence, the selection of the horizon length N has to establish a compromise between convergence and the minimization of the variance.

However, if the cost matrices and the horizon are suitably selected for the system under consideration, the MPC approach stabilizes the system and exhibits good convergence results for most system structures (as demonstrated by the simulations in Sect. 7). Thus, a simulation phase for the tuning of the cost matrices is recommended to get the desired performance in terms of stability and convergence rates. In addition, MSS stability can be enforced by additional constraints similar as in [20], or by the use of terminal costs comparable to [22].

Uncontrollable Systems An unfavourable case for cost convergence should be mentioned: If the Markov state is not measurable while the JMLS is highly symmetric, the costs cannot be influenced by the inputs. This means that all costs are subsumed in Ψ in (21), and the optimizer is a zero vector. An example is a JMLS with:

$$A_1 = \begin{bmatrix} 1.1 & 0 \\ 0 & 1.1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1.1 \\ 1.1 & 0 \end{bmatrix}, \quad P = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}, \tag{46}$$
$$B_1 = B_2 = I_2, \quad \mu(0) = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}^\mathsf{T}, \quad x(0) = \begin{bmatrix} 1 & -1 \end{bmatrix}^\mathsf{T}.$$

This special configuration leads to a stationary probability distribution $\mu(k)$ with equal probabilities for both Markov states. For both, the system behaviour described by A_1 and A_2 is very different. Thus, basically nothing is known about the system behaviour, but the state $x(k)$. However, this does only occur for very few artificial systems, and is barely relevant for applications.

7 Simulation Results

When considering industrial production processes, the main requirements are minimization of costs, stability of the controlled system under consideration of constraints and disturbances. In addition, the applicability to high-dimensional systems, i.e. low computation times of the MPC, are important. Especially, the last point is crucial for applications, e.g. in hot sheet metal forming processes for car body parts, that exhibit time delays and spatially distributed properties in terms of temperature distribution and geometry. Depending on the sampling time and the resolution of the discretization, large-state space dimensions may result. According to these requirements, the properties of the proposed MPC approach have been investigated by simulation studies. These include a performance comparison with the approach from [24], the demonstration of satisfying constraints, stability in case of disturbances, and an evaluation of times for calculation.

Simulation Setup and Performance Measures In all simulations JMLS are considered where the matrices A_i, B_i and P are determined randomly, such that each of the n_θ LTI-systems is unstable with $1.1 \leq \max(|\mathrm{eig}(A_i)|) \leq 2$. The components of $x(0)$ are taken from $[-6; 9]$. For each case Monte Carlo simulations with $n_{\mathrm{sim}} = 200$ runs and a simulation time of $n_{\mathrm{k}} = 50$ steps were performed. The simulations were implemented with MATLAB and `quadprog` on a standard PC with an Intel i5-4670 processor. The resulting control performance is measured by the average costs:

$$J_{\mathrm{s}} := \frac{1}{n_{\mathrm{sim}}} \sum_{r=1}^{n_{\mathrm{sim}}} \left(\sum_{k=1}^{n_{\mathrm{k}}} x_r^{\mathsf{T}}(k)\, Q_{\theta_r(k)}\, x_r(k) + \sum_{k=0}^{n_{\mathrm{k}-1}} u_r^{\mathsf{T}}(k)\, R_{\theta_r(k)}\, u_r(k) \right), \tag{47}$$

where $x_r(k)$, $u_r(k)$ and $\theta_r(k)$ with $k \in \{0, \ldots, n_{\mathrm{k}}\}$ are the trajectories of the states, inputs and Markov states of the r-th simulation run. To analyze the convergence of states and inputs, the following measures are introduced:

$$\hat{x}(k) := \frac{1}{n_{\mathrm{sim}} \cdot n_{\mathrm{x}}} \sum_{l=1}^{n_{\mathrm{x}}} \sum_{r=1}^{n_{\mathrm{sim}}} \left| x_{r,[l]}(k) \right|, \qquad \hat{u}(k) := \frac{1}{n_{\mathrm{sim}} \cdot n_{\mathrm{u}}} \sum_{l=1}^{n_{\mathrm{u}}} \sum_{r=1}^{n_{\mathrm{sim}}} \left| u_{r,[l]}(k) \right|, \tag{48}$$

Here, $x_{r,[l]}(k)$ and $u_{r,[l]}(k)$ denote the l-th component of $x_r(k)$ and $u_r(k)$ respectively. Since these measures describe the mean of absolute values of all states and inputs, the convergence of the measures shows the convergence of all states and inputs.

Table 1 Comparison of average costs $J_{s,1}$ and $J_{s,2}$ for A1 and A2 as well as relative cost differences for 10 different JMLS and 200 Monte Carlo simulations each

No.	1	2	3	4	5	6	7	8	9	10
$J_{s,1}$	117	178	87	312	181	1.2×10^5	143	2.0×10^9	365	111
$J_{s,2}$	98	170	81	349	166	1.1×10^3	139	946	224	114
$\frac{J_{s,2}-J_{s,1}}{J_{s,1}}$ [%]	−16.7	−4.2	−7.0	11.6	−8.6	(−99.1)	−2.5	(−100.0)	−38.5	3.3

Performance Comparison In this section, the resulting performance of the proposed MPC approach (referred to by A2) and of the approach described in [24] and Appendix B (addressed by A1) are compared. To this end, both approaches have been tested with 10 different JMLS with the dimensions $n_x = 10$, $n_u = 4, n_\theta = 4$. For all tests, the prediction horizon $N = 5$ and the cost matrices $Q_i = 0.5I$ and $R_i = I$ for all $i \in \Theta$ have been used. The inputs are constrained component-wise with $|u_{[i]}| \leq 7$, and $\theta(k)$ is assumed to be measurable. The resulting average costs for all systems (denoted by $J_{s,1}$ for A1 and by $J_{s,2}$ for A2) are shown in Table 1.

The results of the simulations show that A2 outperforms A1 for 8 of the 10 systems. For two cases (6 and 8), A1 leads to an unstable closed loop behaviour, while A2 keeps the states bounded (6) or asymptotically stabilizes the system (8). For all cases in which both approaches lead to an asymptotically stable closed loop system, A2 results in 7.8 % lower costs in average. Thus, due to the use of cost function (2a), the approach proposed in this chapter leads in most cases to a better performance and exhibits better stabilizing properties, as expected by the results from Sect. 6. In addition, A2 has an average computation time of 1.7 ms while A1 needs 4.0 ms. This also demonstrates the advantages in terms of computational effort.

Some systems, like the example used in [4] and [24], exhibit high variances of the predicted states. Since A2 tries also to minimize the covariance in the system with an input trajectory independent of the Markov state, this approach leads to a more conservative and slower control action than for A1, and thus in some cases to higher costs. However, there are configurations for these systems in which A1 does not stabilize the system at all, but A2 does. Thus, the developed approach is more robust at the expense of a slower state convergence.

High-Dimensional Example, Calculation Times, and Scalability As stated before, the capability of controlling JMLS with large state spaces as well as satisfying constraints even in case of disturbances are requirements for an application to production processes. Therefore, the proposed MPC approach is tested in this subsection for high-dimensional systems with $n_\theta = 10, n_x = 150, n_w = 150, n_u = 30$, a prediction horizon of $N = 5$, and the cost matrices $Q_i = 100I$, and $R_i = I$ for all $i \in \Theta$. The Markov state $\theta(k)$ is assumed to be measurable, and $G_i = I$ applies for all $i \in \Theta$. To be able to constrain the states, it is assumed that all components i of the disturbance $w_{[i]}(k)$ are given by independent truncated normal distributions with $w_{[i]}(k) \sim \mathcal{N}(0, 0.05)$ and $-0.4 \leq w_{[i]}(k) \leq 0.4$. In addition, component-wise input constraints $|u_{[i]}| \leq 2.5$ and state expectancy constraints $|\bar{x}_{[i]}| \leq 14.6$ are considered to keep the states within $|x_{[i]}| \leq 15$.

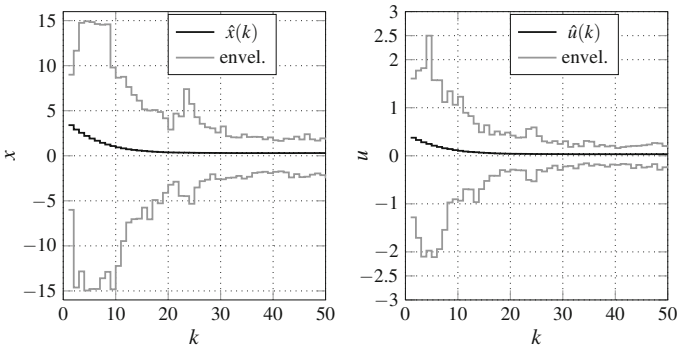


Fig. 1 Results of Monte Carlo simulation showing the measures $\hat{x}$ (*left*) and $\hat{u}$ (*right*) as well as the envelopes containing all trajectories of $x(k)$ and $u(k)$ over all simulation runs. The measures $\hat{x}$ and $\hat{u}$ are not converging to 0 due to the use of the absolute value in (48) and the disturbance $w(k)$

The results for the measures defined in (48) and the envelopes enclosing all components of the states as well as the inputs for all simulation runs are shown in Fig. 1. The results show, that the constraints for the state as well as the inputs are satisfied, and the system is stabilized for all simulation runs in the presence of the defined bounded disturbance. The average computation time is 62 ms.

The simulation of high-dimensional systems showed that the large number of high-dimensional matrices used during the offline calculation as defined in Theorem 2 lead to high memory demands. However, due to the sparse structure of the matrices defined in (30), the necessary memory size can be reduced considerably, if algorithms for sparse matrices are used.

To investigate the scalability of the proposed approach, the computation times for different system dimensions have been measured. The resulting average computation times in milliseconds for pre-processing and solving of the QP are shown in Table 2 for the indicated dimensions, where $n_\mathrm{w} = n_\mathrm{u}$. The first computation time is for A2 with state constraints and the second one for A2 without state constraints, but both with input constraints.

The simulation results show that the resulting computation times are very low even for high system dimensions and large numbers of Markov states. This holds especially for the case where no state expectancy constraints are present. In addition, the computation time is independent of n_θ since the Markov state is measurable. This demonstrates the efficiency of the proposed cost prediction scheme. Thus, for the cases where state constraints are present, the computation times are dominated by the effort for the calculation of the prediction matrices according to Theorem 1. However, it is also possible to use a similar procedure as proposed in this chapter to also calculate the prediction matrices $\bar{A}$, $\bar{B}$ and $\bar{G}$ mainly offline. In addition, the resulting computation times outperform SDP and MILP formulations by far. Just the calculation of stabilizing feedback controllers for a JMLS with $n_\mathrm{w} = n_\mathrm{u} = 15$, $n_\mathrm{x} = 50$ and $n_\theta = 5$ by the commonly used SDP formulations takes about 3 min. Thus, the simulation results show the applicability to large JMLS.

Table 2 Comparison of average computation times for A2 in milliseconds for JMLS of indicated dimensions

N	n_θ	(n_x, n_u)							
		(50,15)		(100,30)		(150,45)		(200,60)	
		SC	No SC	SC	No SC	SC	No SC	SC	No SC
5	5	14	3	30	4	55	6	92	8
	10	25	3	53	4	123	6	182	8
	15	33	3	83	4	175	6	284	8
10	5	40	4	99	8	196	15	345	27
	10	71	4	189	8	394	15	706	26
	15	109	4	291	8	660	15	1167	27

The first computation time in each column is for the case with state constraints (SC) and the second one for the case without state constraints (no SC)

8 Conclusions

In this chapter, an approach to constrained, finite horizon MPC for high-dimensional JMLS has been proposed. These JMLS can be used to model linear systems with abrupt and random changes in its structure. A new recursive prediction scheme for the commonly used cost function, describing the expected value of a quadratic cost criterion, was developed. Due to the special structure of the algorithm, the computational complexity is reduced drastically. Furthermore, most of the computations can be performed offline, and for the proposed approach no truncation of the scenario tree is necessary. By combining this algorithm and the prediction equation developed in [24], the MPC scheme can be formulated as a standard QP.

The advantage of the proposed approach is that due to its recursiveness and the QP formulation it is applicable to high-dimensional systems. These high-dimensional systems often arise when modeling production processes, communication or energy distribution networks. In addition, the proposed algorithms can be extended easily to time varying system matrices and transition probabilities. For most of the existing approaches, this is not possible or leads to quite conservative formulations.

It was shown that converging costs lead to MSS, but the convergence has to be ensured by a careful tuning of the cost matrices. This may be considered a disadvantage of the proposed approach, and the solution is subject to current research. One possibility to guarantee MSS is to add additional constraints enforcing a cost decrease of a Lyapunov function. To this end, concepts proposed in [20] could be used to extended the presented approach to ensure MSS. Future work will also focus on extending the presented approach to be robust against bounded disturbances. Furthermore, formulations guaranteeing recursive feasibility are investigated.

Appendix A—Proof of Theorem 2

This appendix states the proof of Theorem 2. For sake of a brief notation, a certain trajectory of the Markov chain $\mathscr{M}$ defined by $(\theta[\![0]\!] = \theta_0, \ldots, \theta[\![j]\!] = \theta_j)$ is denoted by $(\theta_0, \ldots, \theta_j)$ in this proof. The corresponding realization probability is given by:

$$p_{(\theta_0,\ldots,\theta_j)} := \mathbf{P}\left(\theta[\![0]\!] = \theta_0, \ldots, \theta[\![j]\!] = \theta_j\right) = \mu_{\theta_0}(k) \cdot \prod_{l=0}^{j-1} p_{\theta_{l+1},\theta_l}. \tag{49}$$

Let Λ_j denote the set of all possible Markov state trajectories with j transitions.

By applying the system dynamic (1) j times recursively and consecutive expansion of the products, it follows for the expected cost at time step $k + j$:

$$\begin{aligned}
&\mathrm{E}\left[x^\mathsf{T}[\![j]\!]\, Q_{\theta_j}\, x[\![j]\!]\right] \\
&\quad = \mathrm{E}\left[\left(A_{\theta_{j-1}} x[\![j-1]\!] + B_{\theta_{j-1}} u[\![j-1]\!] + G_{\theta_{j-1}} w[\![j-1]\!]\right)^\mathsf{T} Q_{\theta_j} \cdot \ldots \right. \\
&\qquad \left. \ldots \cdot \left(A_{\theta_{j-1}} x[\![j-1]\!] + B_{\theta_{j-1}} u[\![j-1]\!] + G_{\theta_{j-1}} w[\![j-1]\!]\right)\right] \qquad (50) \\
&\qquad\qquad \vdots \\
&\quad = \sum_{\Lambda_j} p_{(\theta_0,\ldots,\theta_j)} \Bigg(2 \sum_{l=0}^{j-1} x^\mathsf{T}(k) \prod_{c=0}^{j-1} A_{\theta_c}^\mathsf{T} Q_{\theta_j} \prod_{c=1}^{j-l-1} A_{\theta_{j-c}} B_{\theta_l}\, u[\![l]\!] \\
&\qquad + 2 \sum_{l_1=0}^{j-1} \sum_{l_2=0}^{j-1} \bar{w}^\mathsf{T}[\![l_1]\!]\, G_{\theta_{l_1}}^\mathsf{T} \prod_{c=l_1+1}^{j-1} A_{\theta_c}^\mathsf{T} Q_{\theta_j} \prod_{c=1}^{j-l_2-1} A_{\theta_{j-c}} B_{\theta_{l_2}}\, u[\![l_2]\!] \\
&\qquad + \sum_{l_1=0}^{j-1} \sum_{l_2=0}^{j-1} u^\mathsf{T}[\![l_1]\!]\, B_{\theta_{l_1}}^\mathsf{T} \prod_{c=l_1+1}^{j-1} A_{\theta_c}^\mathsf{T} Q_{\theta_j} \prod_{c=1}^{j-l_2-1} A_{\theta_{j-c}} B_{\theta_{l_2}}\, u[\![l_2]\!] \Bigg) + \Psi \\
&\quad = \sum_{\Lambda_{j-1}} p_{(\theta_0,\ldots,\theta_{j-1})} \Bigg(2 \sum_{l=0}^{j-1} x^\mathsf{T}(k) \prod_{c=0}^{j-1} A_{\theta_c}^\mathsf{T} \left(\sum_{\theta_j=1}^{n_\theta} p_{\theta_j,\theta_{j-1}} Q_{\theta_j} \right) \prod_{c=1}^{j-l-1} A_{\theta_{j-c}} B_{\theta_l}\, u[\![l]\!] \\
&\qquad + 2 \sum_{l_1=0}^{j-1} \sum_{l_2=0}^{j-1} \bar{w}^\mathsf{T}[\![l_1]\!]\, G_{\theta_{l_1}}^\mathsf{T} \prod_{c=l_1+1}^{j-1} A_{\theta_c}^\mathsf{T} \left(\sum_{\theta_j=1}^{n_\theta} p_{\theta_j,\theta_{j-1}} Q_{\theta_j} \right) \prod_{c=1}^{j-l_2-1} A_{\theta_{j-c}} B_{\theta_{l_2}}\, u[\![l_2]\!] \\
&\qquad + \sum_{l_1=0}^{j-1} \sum_{l_2=0}^{j-1} u^\mathsf{T}[\![l_1]\!]\, B_{\theta_{l_1}}^\mathsf{T} \prod_{c=l_1+1}^{j-1} A_{\theta_c}^\mathsf{T} \left(\sum_{\theta_j=1}^{n_\theta} p_{\theta_j,\theta_{j-1}} Q_{\theta_j} \right) \prod_{c=1}^{j-l_2-1} A_{\theta_{j-c}} B_{\theta_{l_2}}\, u[\![l_2]\!] \Bigg) + \Psi.
\end{aligned}$$

Here, the variable Ψ contains all costs that cannot be influenced by the inputs. The sums over the cost matrices Q_{θ_j} can be replaced by $\mathscr{T}_{\theta_{j-1}}(Q)$, like in (23). To express the costs as a function of $\mathbf{u}(k)$, the sums over l, l_1 and l_2 are reformulated as matrix multiplications:

$$
\begin{aligned}
&\mathrm{E}\left(x^{\mathsf{T}}[\![j]\!]\, Q_{\theta_j}\, x[\![j]\!]\right) - \Psi \\
&= \sum_{\Lambda_{j-1}} p_{(\theta_0,\ldots,\theta_{j-1})} \left(2\, x^{\mathsf{T}}(k)\, A_{\theta_0}^{\mathsf{T}} \cdot \ldots \cdot A_{\theta_{j-1}}^{\mathsf{T}} \left[\mathscr{T}_{\theta_{j-1}}(Q) \;\ldots\; \mathscr{T}_{\theta_{j-1}}(Q) \right] \begin{bmatrix} A_{\theta_{j-1}} \cdot \ldots \cdot A_{\theta_1} \cdot B_{\theta_0} u[\![0]\!] \\ A_{\theta_{j-1}} \cdot \ldots \cdot A_{\theta_2} \cdot B_{\theta_1} u[\![1]\!] \\ \vdots \\ B_{\theta_{j-1}} u[\![j-1]\!] \end{bmatrix} \right. \\
&\quad + 2 \begin{bmatrix} A_{\theta_{j-1}} \cdot \ldots \cdot A_{\theta_1} \cdot G_{\theta_0} \bar{w}[\![0]\!] \\ A_{\theta_{j-1}} \cdot \ldots \cdot A_{\theta_2} \cdot G_{\theta_1} \bar{w}[\![1]\!] \\ \vdots \\ G_{\theta_{j-1}} \bar{w}[\![j-1]\!] \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathscr{T}_{\theta_{j-1}}(Q) & \ldots & \mathscr{T}_{\theta_{j-1}}(Q) \\ \vdots & \ddots & \vdots \\ \mathscr{T}_{\theta_{j-1}}(Q) & \ldots & \mathscr{T}_{\theta_{j-1}}(Q) \end{bmatrix} \begin{bmatrix} A_{\theta_{j-1}} \cdot \ldots \cdot A_{\theta_1} \cdot B_{\theta_0} u[\![0]\!] \\ A_{\theta_{j-1}} \cdot \ldots \cdot A_{\theta_2} \cdot B_{\theta_1} u[\![1]\!] \\ \vdots \\ B_{\theta_{j-1}} u[\![j-1]\!] \end{bmatrix} \\
&\quad \left. + \begin{bmatrix} A_{\theta_{j-1}} \cdot \ldots \cdot A_{\theta_1} \cdot B_{\theta_0} u[\![0]\!] \\ A_{\theta_{j-1}} \cdot \ldots \cdot A_{\theta_2} \cdot B_{\theta_1} u[\![1]\!] \\ \vdots \\ B_{\theta_{j-1}} u[\![j-1]\!] \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathscr{T}_{\theta_{j-1}}(Q) & \ldots & \mathscr{T}_{\theta_{j-1}}(Q) \\ \vdots & \ddots & \vdots \\ \mathscr{T}_{\theta_{j-1}}(Q) & \ldots & \mathscr{T}_{\theta_{j-1}}(Q) \end{bmatrix} \begin{bmatrix} A_{\theta_{j-1}} \cdot \ldots \cdot A_{\theta_1} \cdot B_{\theta_0} u[\![0]\!] \\ A_{\theta_{j-1}} \cdot \ldots \cdot A_{\theta_2} \cdot B_{\theta_1} u[\![1]\!] \\ \vdots \\ B_{\theta_{j-1}} u[\![j-1]\!] \end{bmatrix} \right)
\end{aligned} \tag{51}
$$

$$
\begin{aligned}
&= \sum_{\Lambda_{j-1}} p_{(\theta_0,\ldots,\theta_{j-1})} \left(2\, x^{\mathsf{T}}(k)\, A_{\theta_0}^{\mathsf{T}} \cdot \ldots \cdot A_{\theta_{j-1}}^{\mathsf{T}} \left[\mathscr{T}_{\theta_{j-1}}(Q) \;\ldots\; \mathscr{T}_{\theta_{j-1}}(Q) \right] \begin{bmatrix} A_{\theta_{j-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \ddots & & \\ \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_{\theta_{j-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & B_{\theta_{j-1}} \end{bmatrix} \cdot \ldots \right. \\
&\quad \ldots \cdot \begin{bmatrix} A_{\theta_{j-2}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \ddots & & & \\ \mathbf{0} & & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_{\theta_{j-2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & B_{\theta_{j-2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & I_{n_{\mathrm{u}}} \end{bmatrix} \cdot \ldots \cdot \begin{bmatrix} A_{\theta_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & B_{\theta_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{(j-2)\cdot n_{\mathrm{u}}} \end{bmatrix} \begin{bmatrix} B_{\theta_0} & \mathbf{0} \\ \mathbf{0} & I_{(j-1)\cdot n_{\mathrm{u}}} \end{bmatrix} \begin{bmatrix} u[\![0]\!] \\ \vdots \\ u[\![j-1]\!] \end{bmatrix} \\
&\quad + 2 \left[\bar{w}^{\mathsf{T}}[\![0]\!] \;\ldots\; \bar{w}^{\mathsf{T}}[\![j-1]\!] \right] \begin{bmatrix} G_{\theta_0}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & I_{(j-1)\cdot n_{\mathrm{w}}} \end{bmatrix} \begin{bmatrix} A_{\theta_1}^{\mathsf{T}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G_{\theta_1}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{(j-2)\cdot n_{\mathrm{w}}} \end{bmatrix} \cdot \ldots \cdot \begin{bmatrix} A_{\theta_{j-1}}^{\mathsf{T}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \ddots & & \\ \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_{\theta_{j-1}}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & G_{\theta_{j-1}} \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} \mathscr{T}_{\theta_{j-1}}(Q) & \ldots & \mathscr{T}_{\theta_{j-1}}(Q) \\ \vdots & \ddots & \vdots \\ \mathscr{T}_{\theta_{j-1}}(Q) & \ldots & \mathscr{T}_{\theta_{j-1}}(Q) \end{bmatrix} \begin{bmatrix} A_{\theta_{j-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \ddots & & \\ \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_{\theta_{j-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & B_{\theta_{j-1}} \end{bmatrix} \cdot \ldots \cdot \begin{bmatrix} B_{\theta_0}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & I_{(j-1)\cdot n_{\mathrm{u}}} \end{bmatrix} \begin{bmatrix} u[\![0]\!] \\ \vdots \\ u[\![j-1]\!] \end{bmatrix} \\
&\quad + \left[u^{\mathsf{T}}[\![0]\!] \;\ldots\; u^{\mathsf{T}}[\![j-1]\!] \right] \begin{bmatrix} B_{\theta_0}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & I_{(j-1)\cdot n_{\mathrm{u}}} \end{bmatrix} \begin{bmatrix} A_{\theta_1}^{\mathsf{T}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & B_{\theta_1}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{(j-2)\cdot n_{\mathrm{u}}} \end{bmatrix} \cdot \ldots \cdot \begin{bmatrix} A_{\theta_{j-1}}^{\mathsf{T}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \ddots & & \\ \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_{\theta_{j-1}}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & B_{\theta_{j-1}} \end{bmatrix} \\
&\quad \left. \cdot \begin{bmatrix} \mathscr{T}_{\theta_{j-1}}(Q) & \ldots & \mathscr{T}_{\theta_{j-1}}(Q) \\ \vdots & \ddots & \vdots \\ \mathscr{T}_{\theta_{j-1}}(Q) & \ldots & \mathscr{T}_{\theta_{j-1}}(Q) \end{bmatrix} \begin{bmatrix} A_{\theta_{j-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \ddots & & \\ \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_{\theta_{j-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & B_{\theta_{j-1}} \end{bmatrix} \cdot \ldots \cdot \begin{bmatrix} B_{\theta_0}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & I_{(j-1)\cdot n_{\mathrm{u}}} \end{bmatrix} \begin{bmatrix} u[\![0]\!] \\ \vdots \\ u[\![j-1]\!] \end{bmatrix} \right),
\end{aligned}
$$

where I and $\mathbf{0}$ denote identity and zero matrices of appropriate dimensions. With the matrices defined in (30), Eq. (51) can be written as a function of $\mathbf{u}(k)$:

$$\begin{aligned}
&\mathrm{E}\left(x^\mathsf{T}[\![j]\!]\, Q_{\theta_j}\, x[\![j]\!]\right) - \Psi \\
&= \sum_{\Lambda_{j-1}} p_{(\theta_0,\dots\theta_{j-1})} \Big(2x^\mathsf{T}(k)\, A_{\theta_0}^\mathsf{T} \cdot \ldots \cdot A_{\theta_{j-1}}^\mathsf{T}\ \hat{Q}_{\mathrm{q_x},\theta_{j-1}}[j]\ \hat{B}_{\theta_{j-1}}[j] \cdot \ldots \cdot \hat{B}_{\theta_0}[1]\, \mathbf{u}(k) \\
&\qquad + 2\bar{\mathbf{w}}^\mathsf{T}(k)\, \hat{G}_{\theta_0}^\mathsf{T}[1] \cdot \ldots \cdot \hat{G}_{\theta_{j-1}}^\mathsf{T}[j]\ \hat{Q}_{\mathrm{q_x},\theta_{j-1}}[j]\ \hat{B}_{\theta_{j-1}}[j] \cdot \ldots \cdot \hat{B}_{\theta_0}[1]\, \mathbf{u}(k) \\
&\qquad + \mathbf{u}^\mathsf{T}(k)\, \hat{B}_{\theta_0}^\mathsf{T}[1] \cdot \ldots \cdot \hat{B}_{\theta_{j-1}}^\mathsf{T}[j]\ \hat{Q}_{\mathrm{W},\theta_{j-1}}[j]\ \hat{B}_{\theta_{j-1}}[j] \cdot \ldots \cdot \hat{B}_{\theta_0}[1]\, \mathbf{u}(k) \Big). \qquad (52)
\end{aligned}$$

Thus, one obtains for the cost prediction matrices:

$$q_\mathrm{x}[\![j]\!] := 2x^\mathsf{T}(k) \sum_{\Lambda_{j-1}} p_{(\theta_0,\dots,\theta_{j-1})}\ A_{\theta_0}^\mathsf{T} \cdot \ldots \cdot A_{\theta_{j-1}}^\mathsf{T}\ \hat{Q}_{\mathrm{q_x},\theta_{j-1}}[j]\ \hat{B}_{\theta_{j-1}}[j] \cdot \ldots \cdot \hat{B}_{\theta_0}[1], \qquad (53)$$

$$q_\mathrm{w}[\![j]\!] := 2\bar{\mathbf{w}}^\mathsf{T}(k) \sum_{\Lambda_{j-1}} p_{(\theta_0,\dots\theta_{j-1})}\ \hat{G}_{\theta_0}^\mathsf{T}[1] \cdot \ldots \cdot \hat{G}_{\theta_{j-1}}^\mathsf{T}[j]\ \hat{Q}_{\mathrm{q_w},\theta_{j-1}}[j]\ \hat{B}_{\theta_{j-1}}[j] \cdot \ldots \cdot \hat{B}_{\theta_0}[1], \qquad (54)$$

$$W'[\![j]\!] := \sum_{\Lambda_{j-1}} p_{(\theta_0,\dots\theta_{j-1})}\ \hat{B}_{\theta_0}^\mathsf{T}[1] \cdot \ldots \cdot \hat{B}_{\theta_{j-1}}^\mathsf{T}[j]\ \hat{Q}_{\mathrm{W},\theta_{j-1}}[j]\ \hat{B}_{\theta_{j-1}}[j] \cdot \ldots \cdot \hat{B}_{\theta_0}[1]. \qquad (55)$$

These equation describe a way to calculate $q_\mathrm{x}[\![j]\!]$, $q_\mathrm{w}[\![j]\!]$, and $W'[\![j]\!]$. However, in this form the summation over all possible Markov trajectories is still employed. The computational effort still depends exponentially on n_θ and N. To reduce the computational effort, the sums are reduced to only the parts that depend on the summation variable. Thus, a nested sum is formed which can be calculated recursively:

$$\begin{aligned}
q_\mathrm{x}[\![j]\!] &= 2x^\mathsf{T}(k) \sum_{\theta_0=1}^{n_\theta} \cdots \sum_{\theta_{j-1}=1}^{n_\theta} p_{(\theta_0,\dots,\theta_{j-1})} A_{\theta_0}^\mathsf{T} \cdot \ldots \cdot A_{\theta_{j-1}}^\mathsf{T} \hat{Q}_{q_x}[j] \hat{B}_{\theta_{j-1}}[j] \cdot \ldots \cdot \hat{B}_{\theta_0}[1] \\
&= 2x^\mathsf{T}(k) \sum_{\theta_0=1}^{n_\theta} \cdots \sum_{\theta_{j-2}=1}^{n_\theta} p_{(\theta_0,\dots,\theta_{j-2})} \cdot A_{\theta_0}^\mathsf{T} \cdot \ldots \cdot A_{\theta_{j-2}}^\mathsf{T} \\
&\qquad \cdot \Bigg(\sum_{\theta_{j-1}=1}^{n_\theta} p_{\theta_{j-1},\theta_{j-2}} \underbrace{A_{\theta_{j-1}}^\mathsf{T} \hat{Q}_{q_x}[j] \hat{B}_{\theta_{j-1}}[j]}_{=:\chi_{\theta_{j-1}}^{(1)}} \Bigg) \hat{B}_{\theta_{j-2}}[j-1] \cdot \ldots \cdot \hat{B}_{\theta_0}[1] \\
&= 2x^\mathsf{T}(k) \sum_{\theta_0=1}^{n_\theta} \cdots \sum_{\theta_{j-2}=1}^{n_\theta} p_{(\theta_0,\dots,\theta_{j-2})} \cdot A_{\theta_0}^\mathsf{T} \cdot \ldots \cdot \underbrace{A_{\theta_{j-2}}^\mathsf{T} \mathscr{T}_{\theta_{j-2}}\left(\chi^{(1)}\right) \hat{B}_{\theta_{j-2}}[j-1]}_{=:\chi_{\theta_{j-2}}^{(2)}} \cdot \ldots \cdot \hat{B}_{\theta_0}[1]
\end{aligned}$$

$$= 2x^{\mathsf{T}}(k) \sum_{\theta_0=1}^{n_\theta} \cdots \sum_{\theta_{j-3}=1}^{n_\theta} p_{(\theta_0,\ldots,\theta_{j-3})} \cdot A_{\theta_0}^{\mathsf{T}} \cdot \ldots \cdot \underbrace{A_{\theta_{j-3}}^{\mathsf{T}} \mathscr{T}_{\theta_{j-3}} \left(\chi^{(2)}\right) \hat{B}_{\theta_{j-3}}[j-2]}_{=:\chi_{\theta_{j-3}}^{(3)}} \cdot \ldots \cdot \hat{B}_{\theta_0}[1]$$

$$\vdots$$

$$= 2x^{\mathsf{T}}(k) \sum_{\theta_0=1}^{n_\theta} \mu_{\theta_0}(k) \underbrace{A_{\theta_0}^{\mathsf{T}} \mathscr{T}_{\theta_0} \left(\chi^{(j-1)}\right) \hat{B}_{\theta_0}[1]}_{=:\chi_{\theta_0}^{(j)}} = 2x^{\mathsf{T}}(k) \sum_{\theta_0=1}^{n_\theta} \mu_{\theta_0}(k) \chi_{\theta_0}^{(j)}. \tag{56}$$

These transformations correspond to the steps defined in Theorem 2. An analogous procedure for $W'[\![j]\!]$ leads to:

$$\begin{aligned}
W'[\![j]\!] &= \sum_{\theta_0=1}^{n_\theta} \cdots \sum_{\theta_{j-1}=1}^{n_\theta} p_{(\theta_0,\ldots,\theta_{j-1})} \cdot \hat{B}_{\theta_0}^{\mathsf{T}}[1] \cdot \ldots \cdot \hat{B}_{\theta_{j-1}}^{\mathsf{T}}[j] \hat{Q}_W[j] \hat{B}_{\theta_{j-1}}[j] \cdot \ldots \cdot \hat{B}_{\theta_0}[1] \\
&= \sum_{\theta_0=1}^{n_\theta} \cdots \sum_{\theta_{j-2}=1}^{n_\theta} p_{(\theta_0,\ldots,\theta_{j-2})} \cdot \hat{B}_{\theta_0}^{\mathsf{T}}[1] \cdot \ldots \cdot \hat{B}_{\theta_{j-2}}^{\mathsf{T}}[j-1] \cdot \ldots \\
&\qquad \ldots \cdot \Bigg(\sum_{\theta_{j-1}=1}^{n_\theta} p_{\theta_{j-1},\theta_{j-2}} \underbrace{\hat{B}_{\theta_{j-1}}^{\mathsf{T}}[j] \hat{Q}_W[j] \hat{B}_{\theta_{j-1}}[j]}_{=:\kappa_{\theta_{j-1}}^{(1)}} \Bigg) \hat{B}_{\theta_{j-2}}[j-1] \cdot \ldots \cdot \hat{B}_{\theta_0}[1] \\
&= \sum_{\theta_0=1}^{n_\theta} \cdots \sum_{\theta_{j-2}=1}^{n_\theta} p_{(\theta_0,\ldots,\theta_{j-2})} \cdot \hat{B}_{\theta_0}^{\mathsf{T}}[1] \cdot \ldots \cdot \underbrace{\hat{B}_{\theta_{j-2}}^{\mathsf{T}}[j-1] \mathscr{T}_{\theta_{j-2}} \left(\kappa^{(1)}\right) \hat{B}_{\theta_{j-2}}[j-1]}_{=:\kappa_{\theta_{j-2}}^{(2)}} \cdot \ldots \cdot \hat{B}_{\theta_0}[1] \\
&\quad \vdots \\
&= \sum_{\theta_0=1}^{n_\theta} \mu_{\theta_0}(k) \underbrace{\hat{B}_{\theta_0}^{\mathsf{T}}[1] \mathscr{T}_{\theta_0} \left(\kappa^{(j-1)}\right) \hat{B}_{\theta_0}[1]}_{=:\kappa_{\theta_0}^{(j)}} = \sum_{\theta_0=1}^{n_\theta} \mu_{\theta_0}(k) \kappa_{\theta_0}^{(j)}.
\end{aligned} \tag{57}$$

For $q_{\mathrm{w}}[\![j]\!]$, the same procedure is used to obtain a recursive algorithm. With these derivations, it is proven that the algorithm in Theorem 2 calculates the cost prediction matrices $q_{\mathrm{x}}[\![j]\!]$, $q_{\mathrm{w}}[\![j]\!]$, and $W'[\![j]\!]$. ■

Appendix B—MPC Approach Proposed in [24]

This section gives a very brief description of the MPC approach given in [24]. The following optimization problem is solved to determine the input trajectory:

$$\min_{\mathbf{u}(k)} \quad \sum_{j=1}^{N} \left(\bar{x}^{\mathsf{T}}[\![j]\!]\, Q_j\, \bar{x}[\![j]\!] + u^{\mathsf{T}}[\![j-1]\!]\, R_{j-1}\, u[\![j-1]\!] \right)$$
$$\text{s.t.} \quad x_{\min,j} \leq \bar{x}[\![j]\!] \leq x_{\max,\mathrm{j}}, \qquad u_{\min,j} \leq u[\![j-1]\!] \leq u_{\max,j} \qquad \forall\, j \in \{1,\ldots,N\}. \tag{58}$$

Here, the cost matrices do not depend on the Markov state, but on the prediction step. With (35) and results from Sect. 3 (58) can be formulated as a QP [24].

References

1. Costa OLV, Fragoso MD, Marques RP (2005) Discrete-time Markov jump linear systems. Probability and its applications. Springer, New York
2. Maciejowski JM (2002) Predictive control with constraints. Prentice-Hall, New Jersey
3. Costa OLV, Filho EOA (1996) Discrete-time constrained quadratic control of Markovian jump linear systems. In: Conference on Decision Control 2:1763–1764
4. Costa OLV, Filho EOA, Boukas EK, Marques RP (1999) Constrained quadratic state feedback control of discrete-time Markovian jump linear systems. Automatica 35(4):617–626
5. do Val JBR, Başar T (1999) Receding horizon control of jump linear systems and a macroeconomic policy problem. J Econ Dyn Control 23(8):1099–131, 1999
6. Vargas AN, do Val JBR, Costa EF (2004) Receding horizon control of Markov jump linear systems subject to noise and unobserved state chain. In: Conference on decision and control, vol 4, pp 4381–4386
7. Park B-G, Kwon WH (2002) Robust one-step receding horizon control of discrete-time Markovian jump uncertain systems. Automatica 38(7):1229–1235
8. Vargas AN, Furloni W, do Val JBR (2006) Constrained model predictive control of jump linear systems with noise and non-observed Markov state. In: American control conference
9. Vargas AN, Furloni W, do Val JBR (2007) Control of Markov jump linear systems with state and input constraints: a necessary optimality condition. In: 3rd IFAC symposium on system, structure and control, vol 3. pp 250–255
10. Vargas AN, Furloni W, do Val JBR (2013) Second moment constraints and the control problem of Markov jump linear systems. Numer Linear Algebra Appl 20(2):357–368
11. Blackmore L, Bektassov A, Ono M, Williams BC (2007) Robust, optimal predictive control of jump Markov linear systems using particles. In: Hybrid systems: computation and control, Lecture notes in computer science, vol 4416. Springer, New York, pp 104–117
12. Blackmore L, Ono M, Bektassov A, Williams BC (2010) A probabilistic particle-control approximation of chance-constrained stochastic predictive control. IEEE Trans Robot, 26(3):502–517
13. Yin Y, Shi Y, Liu F (2013) Constrained model predictive control on convex polyhedron stochastic linear parameter varying systems. Int. Journal of Innovative Computing. Inf Control 9(10):4193–4204
14. Yin Y, Liu Y, Karimi HR (2014) A simplified predictive control of constrained Markov jump system with mixed uncertainties. Abstr Appl Anal Special Issue:1–7
15. Lu J, Li D, Xi Y (2012) Constrained MPC of uncertain discrete-time Markovian jump linear systems. In: 31st Chinese control conference, pp 4131–4136
16. Song Y, Liu S, Wei G (2015) Constrained robust distributed model predictive control for uncertain discrete-time Markovian jump linear system. J Frankl Inst 352(1):73–92
17. Dombrovskii VV, Dombrovskii DV, Lyashenko EA (2005) Predictive control of random-parameter systems with multiplicative noise. Application to investment portfolio optimization. Autom Remote Control 66(4):583–595

18. Dombrovskii VV, Yu T (2011) Ob"edko. Predictive control of systems with Markovian jumps under constraints and its application to the investment portfolio optimization. Autom Remote Control 72(5):989–1003
19. Yan Z, Wang J (2013) Stochastic model predictive control of Markov jump linear systems based on a two-layer recurrent neural network. In: IEEE international conference on information and automation, pp 564–569
20. Bernardini D, Bemporad A (2012) Stabilizing model predictive control of stochastic constrained linear systems. IEEE Trans Autom Control 57(6):1468–1480
21. Patrinos P, Sopasakis P, Sarimveis H, Bemporad A (2014) Stochastic model predictive control for constrained discrete-time Markovian switching systems. Automatica 50(10):2504–2514
22. Lu J, Xi Y, Li D, Cen L (2014) Probabilistic constrained stochastic model predictive control for Markovian jump linear systems with additive disturbance. In: 19th IFAC world congress, vol 19, pp 10469–10474
23. Chitraganti S, Aberkane S, Aubrun C, Valencia-Palomo G, Dragan V (2014) On control of discrete-time state-dependent jump linear systems with probabilistic constraints: a receding horizon approach. Syst Control Lett 74:81–89
24. Tonne J, Jilg M, Stursberg O (2015) Constrained model predictive control of high dimensional jump Markov linear systems. In: American control conference pp 2993–2998
25. Jerez JL, Kerrigan EC, Constantinides GA (2011) A condensed and sparse QP formulation for predictive control. In: 50th IEEE conference on decision and control, pp 5217–5222
26. Seber GAF, Lee AJ (2003) Linear regression analysis. Wiley series in probability and statistics. 2nd edn. Wiley, New York