IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 October 2025, accepted 17 November 2025, date of publication 25 November 2025,
date of current version 19 December 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3636666

RESEARCH ARTICLE

On the Structural Analysis and Optimal Input Design for Joint State and Parameter Estimation

ARTHUR LEPSIEN[1,3], PHILIPP KÜGLER[2,3], AND ALEXANDER SCHAUM[1,3], (Senior Member, IEEE)
[1]Department of Process Analytics, University of Hohenheim, 70599 Stuttgart, Germany
[2]Institute of Mathematics and Statistics, University of Hohenheim, 70599 Stuttgart, Germany
[3]Computational Science Hub, University of Hohenheim, 70599 Stuttgart, Germany

Corresponding author: Arthur Lepsien (arthur.lepsien@uni-hohenheim.de)

This work was supported by the Funding Program Open Access Publishing of University of Hohenheim.

ABSTRACT This paper addresses the problem of joint state and parameter estimation for nonlinear affine-input systems with positive parameters including the design of a closed-loop optimal input adaptation to increase an identifiability measure for the system. The identifiability itself is considered in the context of structural observability of the system dynamics based on structural analysis of the system including the unknown parameters as additional states. In particular, the network graph-based interpretation of structural observability is employed at this point. This analysis motivates to include time derivatives of the measurements as additional system outputs to enhance the structural observability properties. For this purpose robust exact differentiation is considered, relying on the super-twisting algorithm to obtain finite time convergent estimates of these signals. Using the extended measurement signal, a continuous-discrete Extended Kalman filter is proposed that ensures strictly positive estimates for the parameters. Based on the estimates of states and parameters the input signal is determined using a moving horizon optimal predictive control that evaluates the condition number of the Fisher information matrix, thus maximizing the information content of the measurements with respect to the parameters. The proposed scheme extends and combines different previously discussed approaches from the literature and is evaluated by means of a thermal process example in simulation and experiment, showing high potential for similar system identification problems.

INDEX TERMS State and parameter estimation, nonlinear systems, closed–loop identification, Kalman filtering, process control, optimization, design of experiments.

I. INTRODUCTION

Parameter estimation for mathematical models is a central key for accessing the potential of predictive models [1], [2]. This important topic has consequently received considerable attention over the last about 50 years, starting with classical design of experiments [3], leading to the well-known Fisher information matrix, and advancing to optimal experimental design for dynamical models [4], [5].

For dynamical systems with inputs the topic of feedfoward, feedback and optimal control design for parameter estimation has further paved the way to access the influence of the input on the identifiability based on Fisher information measures. In this regard, and without being exhaustive, e.g., [6] analytically addressed the control design for parameter estimation in a water tank system, optimal control based on Pontryagin's maximum principle has been employed in [7], and model predictive (moving horizon optimal) control, e.g., in [8]. Different application studies can be found in the literature including, e.g., parameter estimation for an unstable delta wing [9], for battery models [10] or a turntable ladder [11], [12]. All these studies highlight the use of the Fisher information matrix and in particular address the problem of maximizing the information content.

The information content itself is related to the fundamental question of identifiability, i.e., whether or not there exists

The associate editor coordinating the review of this manuscript and approving it for publication was Ángel F. García-Fernández.

a unique set of parameters corresponding to the observed input-output signals. The next question is then how to design systems for automatic parameter estimation for a given actuator-system-sensor setup. Both questions have been addressed in particular in the context of observability, a property that is strongly connected to the identifiability. Observability ensures that the states of a system can be uniquely determined on the basis of the knowledge on input and output over a (sufficiently long) time interval.[1] Considering the parameters as additional states without proper dynamics, the joint state and parameter estimation problem thus renders into a pure state estimation problem. The existence of a so-called observer to provide asymptotically converging estimates of states is ensured in the case that the system is observable. The connections between identifiability, observability and the Fisher information matrix have been often recognized and highlighted in the literature, e.g., in [14], [15], and [16] or in the recent survey on specific related studies and methods provided in [17].

It is well-known that observability can be analyzed using sequential time derivatives of the outputs (e.g., in the classical Kalman observability matrix analysis [13]), or in an equivalent manner evaluating forward sensitivities of the initial conditions on the outputs by the observability Gramian, or the Fisher information matrix related to the outputs with respect to the initial conditions (cp. [14], [15], [16]). From the first point of view, the information contained in the time derivatives of the outputs includes information of the local past and future in terms of the slope of the measurement signal. The classical Kalman observability condition [13] shows that a necessary condition for observability is that subsequent time derivatives of the outputs provide independent information about the states. Thus, having access to the time derivatives of the outputs naturally enriches the information content of the measurements.

The determination of time derivatives of signals has itself a considerable history (see, e.g., [18]), and it is almost 30 years that the exact differentiator has been proposed in [19]. The topic has gained momentum over the last years with studies focusing either on algebraic methods and approximation theory (see, e.g., [18]) or sliding mode techniques (see, e.g., [20]) and their analysis using Lyapunov theory [21], [22], as well as discrete-time and filtered implementations [23], [24]. One of the main advantages of the latter approach is the ensurance of convergence in finite time, as long as certain criteria are satisfied which have to be analyzed separately. Sliding mode observers have also already been used in the context of parameter estimation, e.g., in [25] and [26], where they were directly employed for the state and parameter estimator design.

Having these enlightening studies as methodological point of departure, the purpose of the present study is to combine and extend the approaches with focus on nonlinear systems with positive parameters. The positivity assumption is motivated by physical constraints as well as stability properties of the system. Particular focus is put on the interrelation of the output sensitivity and its maximization using a measure of the information content in the Fisher information matrix. The potential of improving information content using higher-order sliding mode differentiators and their use for positive state estimation employing an adapted Kalman filter approach is discussed. As a particular case example a thermal conduction process with two coupled heating elements and two temperature measurements is employed.

II. PROBLEM SETUP

In the following we consider multi-input multi-output (MIMO) systems of the form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{p}) + \sum_{i=1}^{n_u} \boldsymbol{g}_i(\boldsymbol{p}) u_i, \tag{1a}$$

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}), \tag{1b}$$

with the state $\boldsymbol{x}(t) \in \mathbb{R}^{n_x}$ at time $t \geq 0$, initial condition $\boldsymbol{x}(0) = \boldsymbol{x}_0$, unknown parameter vector $\boldsymbol{p} \in \mathbb{R}_+^{n_p}$, vector field $\boldsymbol{f} : \mathbb{R}^{n_x \times n_p} \to \mathbb{R}^{n_x}$ which in the following is assumed to consist of entries f_i that are sufficiently often continuously differentiable with respect to both $\boldsymbol{x}$ and $\boldsymbol{p}$. The same holds for the input vector fields $g_i : \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$. It is assumed that the sign of the parameters are known, so that without loss of generality[2] it is assumed that $\boldsymbol{p} \in \mathbb{R}_+^{n_p}$. The input gain vector associated to the input $u_i \in \mathcal{U}$, $i = 1, \ldots, n_u$ is denoted by $\boldsymbol{g}_i$ and assumed to be sufficiently often continuously differentiable with respect to $\boldsymbol{p}$. Here $\mathcal{U}$ denotes the set of bounded functions $u_i : [0, \infty) \to \mathbb{R}$ which are sufficiently often continuously differentiable. The output $\boldsymbol{y}(t) \in \mathbb{R}^{n_y}$ is determined by the output function $\boldsymbol{h} : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$, which has at least n_x-times continuously differentiable entries h_i. The unknown parameter vector $\boldsymbol{p} \in \mathbb{R}_+^{n_p}$ is considered constant, in the understanding that eventual slow variations are happening on a time scale that is much slower than the one of the state dynamics and can thus be neglected. In the sequel we denote by $\boldsymbol{x}(t; \boldsymbol{x}_0, \boldsymbol{u}, \boldsymbol{p})$ the solution of (1a) for a given parameter vector $\boldsymbol{p}$ and input vector

$$\boldsymbol{u}(t) = \left[u_1(t) \ \ldots \ u_{n_u}(t)\right]^{\mathrm{T}} \in \mathbb{R}^{n_u}.$$

Throughout, we assume that the solution exists and is unique. In special examples this of course may imply additional constraints on the parameters, which will not be further addressed in the sequel.

III. IDENTIFIABILITY AND OBSERVABILITY

The identifiability of system (1) is ensured if one can choose the inputs $\boldsymbol{u}$ as functions of time in such a way that for given $\boldsymbol{x}_0$ there exists a unique parameter vector $\boldsymbol{p}$ for which the solutions of (1a) provide the same output according to (1b)

[1] For linear time-invariant finite-dimensional systems this interval can actually be arbitrary short [13].

[2] Note that any model can be written in such a way, as long as the signs of the parameters are known. Usually, physical parameters are positive, e.g. masses.

as a measurement $\boldsymbol{y}(t)$ does, i.e.,

$$\boldsymbol{y}(t) = \boldsymbol{h}(\boldsymbol{x}(t; \boldsymbol{x}_0, \boldsymbol{u}, \boldsymbol{p})), \ \forall\, t \geq 0. \tag{2}$$

In other words, the only vector $\boldsymbol{\epsilon}_p$ for which

$$\boldsymbol{y}(t) = \boldsymbol{h}(\boldsymbol{x}(t; \boldsymbol{x}_0, \boldsymbol{u}, \boldsymbol{p} + \boldsymbol{\epsilon}_p)), \ \forall\, t \geq 0. \tag{3}$$

holds true is given by $\boldsymbol{\epsilon}_p = \boldsymbol{0}$.

On the other side, observability of system (1) means that for a given set of inputs, outputs and a given parameter vector $\boldsymbol{p}$, the initial condition $\boldsymbol{x}_0$ is the only possibility for which (2) is satisfied. In other words, for a given input-output set $(\boldsymbol{u}, \boldsymbol{y})$ and parameter vector $\boldsymbol{p}$ the only vector $\boldsymbol{\epsilon}_x$ for which

$$\boldsymbol{y}(t) = \boldsymbol{h}(\boldsymbol{x}(t; \boldsymbol{x}_0 + \boldsymbol{\epsilon}_x, \boldsymbol{u}, \boldsymbol{p})), \ \forall\, t \geq 0 \tag{4}$$

holds true is given by $\boldsymbol{\epsilon}_x = \boldsymbol{0}$. Condition (4) is called indistinguishability in the literature [27]. From this consideration it directly becomes clear that both the identifiability and observability are strongly related. Indeed, observability can be interpreted as identifiability if one considers the initial condition $\boldsymbol{x}_0$ as a parameter vector.

Given that very powerful approaches exist for state estimation and the analysis of observability, the problem of joint state and parameter estimation is subsequently addressed in the context of state estimation. For this purpose we consider the extended system

$$\dot{\boldsymbol{z}} = \boldsymbol{f}_z(\boldsymbol{z}) + \sum_{i=1}^{n_u} \boldsymbol{g}_{z,i}(\boldsymbol{z}) u_i, \quad \boldsymbol{z} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{p} \end{bmatrix} \in \mathbb{R}^{n_x + n_p} \tag{5a}$$

$$\boldsymbol{y} = \boldsymbol{h}_z(\boldsymbol{z}), \tag{5b}$$

with

$$\boldsymbol{f}_z = \begin{bmatrix} \boldsymbol{f}(\boldsymbol{z}) \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{g}_{z,i} = \begin{bmatrix} \boldsymbol{g}(\boldsymbol{z}) \\ \boldsymbol{0} \end{bmatrix}, \ i = 1, \ldots, n_u, \tag{5c}$$

$$\boldsymbol{h}_z(\boldsymbol{z}) = \boldsymbol{h}(\boldsymbol{x}), \tag{5d}$$

which corresponds to the assumption of constant parameters. The solution of (5) is denoted by $\boldsymbol{z}(t; \boldsymbol{z}_0, \boldsymbol{u})$ in the following. Note that $z_1(t), \ldots, z_{n_x}(t)$ coincide with $x_1(t), \ldots, x_{n_x}(t)$ if $\left[z_1(0), \ldots, z_{n_x}(0)\right]^{\mathrm{T}} = \boldsymbol{x}_0$ and $\left[z_{n_x+1}(0), \ldots, z_{n_x+n_p}(0)\right]^{\mathrm{T}} = \boldsymbol{p}$. If the system is identifiable the only valid solution for $z_{n_x+1}, \ldots, z_{n_x+n_p}$ for which $\boldsymbol{y}(t) = \boldsymbol{h}_z(\boldsymbol{z}(t)) \ \forall t \geq 0$ will indeed be $z_{n_x+i} = p_i, \ i = 1, \ldots, n_p$ provided that $z_i = x_i, \ i = 1, \ldots, n_x$. The problem of joint state and parameter estimation thus resembles to designing an estimator (or observer) that provides an estimate $\hat{\boldsymbol{z}}$ that asymptotically converges to $\boldsymbol{z}$, i.e.,

$$\lim_{t \to \infty} \|\hat{\boldsymbol{z}}(t) - \boldsymbol{z}(t)\| = 0, \tag{6}$$

implying that both the state and the parameters are uniquely determined at least asymptotically.

In terms of the extended state system formulation (5), the property underlying the state and parameter estimation problem is the observability. System (5) is observable if for a given input-output set $(\boldsymbol{u}, \boldsymbol{y})$ the only vector $\boldsymbol{\epsilon}_z$ for which

$$\boldsymbol{y}(t) = \boldsymbol{h}_z(\boldsymbol{z}(t; \boldsymbol{z}_0 + \boldsymbol{\epsilon}_z, \boldsymbol{u})), \ \forall\, t \geq 0 \tag{7}$$

holds true is given by $\boldsymbol{\epsilon}_z = \boldsymbol{0}$. If for any $\boldsymbol{\epsilon}_z$ where (7) holds true, one has that

$$\lim_{t \to \infty} \|\boldsymbol{z}(t; \boldsymbol{z}_0, \boldsymbol{u}) - \boldsymbol{z}(t; \boldsymbol{z}_0 + \boldsymbol{\epsilon}_z, \boldsymbol{u})\| = 0,$$

then the system (5) is called detectable.

With respect to (6), detectability is a nessesary condition. Actually, from an estimator design perspective [13], it is also sufficient for the design of a local state observer, but as the parameters have no dynamics, in the sequel the focus will be on observability, being the stronger condition. A detailed analysis based on a separation into observable and non-observable parts [28] goes beyond the scope of the present study.

Observability can be analyzed locally around the actual state trajectory using the observability map

$$\boldsymbol{\mathcal{O}}(\boldsymbol{z}, \bar{\boldsymbol{u}}) = \begin{bmatrix} h_{z,1} \\ \vdots \\ h_{z,n_y} \\ L_{\boldsymbol{f}_z} h_{z,1} \\ \vdots \\ L_{\boldsymbol{f}_z} h_{z,n_y} \\ \vdots \\ L_{\boldsymbol{f}_z}^k h_{z,1} \\ \vdots \\ L_{\boldsymbol{f}_z}^k h_{z,n_y} \end{bmatrix}, \tag{8}$$

consisting of subsequent coordinate–wise Lie derivatives

$$L_{\boldsymbol{f}} h(\boldsymbol{z}) = \frac{\partial h}{\partial \boldsymbol{z}} \boldsymbol{f}(\boldsymbol{z}), \quad L_{\boldsymbol{f}}^k h = L_{\boldsymbol{f}} \left(L_{\boldsymbol{f}}^{k-1} h \right), \tag{9}$$

of the output map $\boldsymbol{h}_z$ until the order $k = n_x + n_p - 1$ [28], with derivative augmented input vector

$$\bar{\boldsymbol{u}} = \begin{bmatrix} u_1 & \dot{u}_1 & \ldots & u_1^{(k-1)} & \ldots & u_{n_u}^{(k-1)} \end{bmatrix}^{\mathrm{T}}.$$

Note that the computation of the map (8) in the presence of inputs may also involve the first k time derivatives of the input, what underlines the non-trivial dependency of the observability on the input u in case of nonlinear systems [28]. Local observability at a point $\boldsymbol{z}$ is ensured if there exists a neighborhood of $\boldsymbol{z}$ in which this map is invertible. In virtue of the inverse function theorem this is ensured if the associated Jacobian matrix of (8) satisfies

$$\operatorname{rank} \left(\frac{\partial \boldsymbol{\mathcal{O}}}{\partial \boldsymbol{z}}(\boldsymbol{z}, \bar{\boldsymbol{u}}) \right) = n_x + n_p. \tag{10}$$

As this analysis involves nonlinear functions that depend on the actual state, parameters and input, the analysis typically becomes quite cumbersome at some point. A suitable alternative consists in the analysis of the structural identifiability in

the sense of the structural observability [29], [30], [31] of the extended system dynamics (5), which is also a local property.

For the analysis of the structural observability the structure graph $\Gamma = (V, E)$ of (5) is employed [29], [31], [32]. The structure graph consists of the vertex set

$$V = \{v_1, \dots, v_{n_x+n_p}\},$$

where v_k is associated to the state z_k, $k = 1, \dots, n_x + n_p$, and the (directed) edge set

$$E(\boldsymbol{z}, \boldsymbol{u}) = \left\{ (v_j, v_i) \,\middle|\, \frac{\partial f_{z,i}(\boldsymbol{z})}{\partial z_j} + \sum_{k=1}^{n_u} \frac{\partial g_{z,k,i}(\boldsymbol{z})}{\partial z_j} u_k \neq 0 \right\}.$$

Note that the edge set in principle depends on the particular values of $\boldsymbol{z}$ and $\boldsymbol{u}$ and thus the interpretation is carried out structurally, i.e., the edge is included as long as there exists at least one choice of $\boldsymbol{z}, \boldsymbol{u}$ such that the partial derivative is not zero. Note that by continuity arguments the same holds in a neighborhood and in consequence almost everywhere (see, e.g., [29], [31], [32]). The structure graph is extended by introducing the output vertices v_i, $i = 1, \dots, n_y$ in such a way that a directed edge (v_j, v_i) is included whenever $\frac{\partial h_i}{\partial z_j}(\boldsymbol{z}) \neq 0$ (structurally). Subsequently, the graph is separated into strongly connected components, i.e., subsets in which one can get from any vertex to any other one by following the edges without leaving the set. The condition for structural observability then requires that at least one sensor is connected to every such strongly connected component. In addition the structural rank condition

$$\mathrm{s-rank} \begin{bmatrix} \mathfrak{J}_z \\ \mathfrak{H}_z \end{bmatrix} = n_x + n_p, \tag{11}$$

has to be satisfied. Here $\mathfrak{J}_z$, $\mathfrak{H}_z$ denote the structure matrices of the Jacobians of $\boldsymbol{f}_z$ and $\boldsymbol{h}_z$ with respect to $\boldsymbol{z}$. These matrices carry the information which entries are non–zero for at least one choice of $\boldsymbol{z}, \boldsymbol{u}$, and the structural rank s − rank is defined as the maximum possible rank. The structural observability s − rank condition is a sufficient condition for the local paramter estimation problem to be well posed. Note that in terms of structural identifiablity it is indeed not nessesary.

Note that for $n_y < n_p$ the s-rank can be at most $n_x + n_y < n_x + n_p$, showing the underlying fundamental problem. There are basically two potential solutions to this problem. Either one includes additional sensors and thus output equations, or one introduces successive time derivatives of the existing outputs, also providing additional output equations in terms of Lie derivatives. Whether or not the latter approach provides an increase of the s − rank has to be analyzed for the given system. Anyway, taking into account the considered setup (5) it can be easily seen that an increase in the s − rank is likely by computation of the Jacobians of $y_i = h_{z,i}(\boldsymbol{z})$ and $\dot{y}_i = L_{\boldsymbol{f}_z} h_{z,i}(\boldsymbol{z})$, i.e.,

$$\frac{\partial h_i}{\partial \boldsymbol{z}} = \begin{bmatrix} \frac{\partial h_i}{\partial z_1} & \cdots & \frac{\partial h_i}{\partial z_{n_x}} & 0 & \cdots & 0 \end{bmatrix} \tag{12a}$$

$$\frac{\partial L_{\boldsymbol{f}_z} h_i}{\partial \boldsymbol{z}} = \begin{bmatrix} \frac{\partial L_{\boldsymbol{f}_z} h_i}{\partial z_1} & \cdots & \cdots & \frac{\partial L_{\boldsymbol{f}_z} h_i}{\partial z_{n_x+n_p}} \end{bmatrix}, \tag{12b}$$

with the definition of the Lie derivative from (9). It becomes evident that, since in the current setup the output function depends only on the states $\boldsymbol{x}$ and not on the parameters $\boldsymbol{p}$, the Lie derivative $\dot{y} = L_{\boldsymbol{f}_z} h_i$ generally introduces a coupling with the parameters. By explicitly including these new output functions $h_{z,n_x+i} = L_{\boldsymbol{f}_z} h_i$ additional information is incorporated into the structure graph, which can potentially enhance the systems structural observability. If the output y_i is included up to its r_i-th time derivative, the number of measurement equations is generally given by

$$\mathcal{M} = \sum_{i=1}^{n_y} r_i + 1. \tag{13}$$

The structural rank condition (11) requires $\mathcal{M} \geq n_p$. This part will be evidenced and further discussed later in the case study.

IV. OUTPUT DERIVATIVES

To obtain estimates of the time derivatives of the output y_i up to variable derivative order r_i, $i = 1, \dots, n_y$, the sliding mode based robust exact differentiator [20], [33] can be employed in the form

$$\dot{\zeta}_{i,1} = k_1 \lfloor \sigma_{i,1} \rceil^{\frac{r_i}{r_i+1}} + \zeta_{i,2} \tag{14a}$$

$$\vdots$$

$$\dot{\zeta}_{i,j} = k_j \lfloor \sigma_{i,1} \rceil^{\frac{r_i-j+1}{r_i+1}} + \zeta_{i,j+1} \tag{14b}$$

$$\vdots$$

$$\dot{\zeta}_{i,r_i+1} = k_{r_i+1} \operatorname{sign}(\sigma_{i,1}), \tag{14c}$$

where $\lfloor \cdot \rceil^{\gamma} = |\cdot|^{\gamma} \operatorname{sign}(\cdot)$, k_i are positive gains, $\zeta_{i,j}(t)$ represent the estimate of $y_i^{(j)}(t)$ and

$$\sigma_{i,1} = y_i - \zeta_{i,1}. \tag{15}$$

The solution of the discontinuous differential equation is understood in the sense of Filippov [34].

Introducing the estimation errors for the j-th derivative as

$$\tilde{\zeta}_{i,j} = y_i^{(j)} - \zeta_{i,j} \tag{16}$$

the associated estimation error dynamics take the form

$$\dot{\tilde{\zeta}}_{i,1} = k_1 \lfloor \tilde{\zeta}_{i,1} \rceil^{\frac{r_i}{r_i+1}} + \tilde{\zeta}_{i,2} \tag{17a}$$

$$\vdots$$

$$\dot{\tilde{\zeta}}_{i,j} = k_j \lfloor \tilde{\zeta}_{i,1} \rceil^{\frac{r_i-j+1}{r_i+1}} + \tilde{\zeta}_{i,j+1} \tag{17b}$$

$$\vdots$$

$$\dot{\tilde{\zeta}}_{i,r_i+1} = k_{r_i+1} \operatorname{sign}(\tilde{\zeta}_{i,1}) + y_i^{(r_i+1)}, \tag{17c}$$

allowing for finite-time convergence for suitably chosen gains k_j, according to [21], [35], and [36], as long as $y_i^{(r_i+1)}$ is bounded. As pointed out above, the higher-order output derivatives can depend on the time derivatives of the inputs

as well, so that the requirement that $y_i^{(r+1)}$ is bounded in particular requires that the inputs $\boldsymbol{u}$ are sufficiently often differentiable. There exist methods for discrete-time implementation in the presence of noise (see, e.g., [23]).

V. POSITIVE STATE AND PARAMETER ESTIMATION

The state and parameter estimation problem can be addressed, e.g., using an extended Kalman filter (EKF) scheme (see, e.g., [37], [38] and a recent application for parameter estimation, e.g., in [39]). To account for the particular challenge at hand that the parameters are considered positive, the following transformation is employed

$$p_i = e^{\theta_i} > 0 \quad \forall\, \theta_i \in \mathbb{R}, \qquad \theta_i = \ln(p_i), \tag{18}$$

$$\dot{p}_i = e^{\theta_i}\dot{\theta}_i = 0. \tag{19}$$

This transformation is particularly useful in cases where negative parameter values are physically meaningless, and negative estimates would compromise the stability of both the system and the estimation process. In such situations, the estimator may diverge from the measurements and is unlikely to recover from this error. This consideration leads to the redefined extended state vector $\bar{z} = [\boldsymbol{x} \quad \boldsymbol{\theta}]^{\mathrm{T}}$ with dynamics

$$\dot{\bar{z}} = \bar{\boldsymbol{f}}_z(\bar{z}) + \sum_{i=1}^{n_u} \bar{\boldsymbol{g}}_{z,i}(\bar{z})u_i \tag{20}$$

with $\bar{\boldsymbol{f}}_z(\bar{z}) := \boldsymbol{f}_z(\boldsymbol{x}, e^{\boldsymbol{\theta}})$ and $\bar{\boldsymbol{g}}_{z,i} := \boldsymbol{g}_{z,i}(\boldsymbol{x}, e^{\boldsymbol{\theta}})$. Furthermore, we consider Gaussian process noise $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, Q)$ and measurement noise $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, R)$ to represent model and measurement uncertainties, respectively. The positive definite matrices Q and R define the covariances of these noise realizations.

The EKF provides both an estimate of the state vector $\hat{\bar{z}}(t)$ and for the associated estimation error covariance $P(t)$, i.e., $\hat{\bar{z}}(t) - \bar{z}(t) \sim \mathcal{N}(\boldsymbol{0}, P(t))$ as long as $\hat{\bar{z}}_0 - \bar{z}_0 \sim \mathcal{N}(\boldsymbol{0}, P_0)$ based on the assumption of an underlying Gaussian probability density function.

Considering discrete measurements $\boldsymbol{y}(t_k) = \boldsymbol{h}_z(\boldsymbol{z}(t_k))$ a continuous-discrete EKF is employed, based on periodic sampling at times $t_k = k\Delta t$. The continuous state update is based on the nominal system dynamics and the covariance update is based on the linearized Ricatti equation, leading to

$$\dot{\hat{\bar{z}}} = \bar{\boldsymbol{f}}_z(\hat{\bar{z}}) + \sum_{i=1}^{n_u} \bar{\boldsymbol{g}}_{z,i}(\hat{\bar{z}})u_i \tag{21a}$$

$$\dot{P} = \mathfrak{J}_z(\hat{\bar{z}}, u)P + P\mathfrak{J}_z^{\mathrm{T}}(\hat{\bar{z}}, u) + Q \tag{21b}$$

$$\mathfrak{J}_z(\hat{\bar{z}}, u) = \frac{\partial}{\partial \bar{z}}\left(\bar{\boldsymbol{f}}_z(\bar{z}) + \sum_{i=1}^{n_u} \bar{\boldsymbol{g}}_{z,i}(\bar{z})u_i\right)\Bigg|_{\bar{z}=\hat{\bar{z}}} \tag{21c}$$

for $t \in [(k-1)\Delta t, k\Delta t)$, starting with $\hat{\bar{z}}(0) = \hat{\bar{z}}_0$, $P(0) = P_0$, leading to predictions $\hat{\bar{z}}_k^- = \hat{\bar{z}}(t_k)$, $P_k^- = P(t_k)$ with which the new estimates are obtained as

$$\hat{\bar{z}}(t_k) := \hat{\bar{z}}_k^- - \mathfrak{L}_k(\boldsymbol{h}_z(\hat{\bar{z}}_k^-) - \boldsymbol{y}(t_k)) \tag{21d}$$

$$\mathfrak{L}_k := P_k^- \mathfrak{H}_k^{\mathrm{T}} \left(\mathfrak{H}_k^{\mathrm{T}} P_k^- \mathfrak{H}_k + \frac{R}{\Delta t}\right)^{-1} \tag{21e}$$

$$\mathfrak{H}_k = \frac{\partial \boldsymbol{h}_z}{\partial \boldsymbol{z}}(\hat{\bar{z}}_k^-) \tag{21f}$$

$$P(t_k) := (I - L_k \mathfrak{H}_k^{\mathrm{T}})P_k^-. \tag{21g}$$

The noise covariance of the continuous time process is scaled by the sampling interval Δt as the discrete time noise can be approximated via

$$\boldsymbol{v}_k = \frac{1}{\Delta t}\int_{t_{k-1}}^{t_k} \boldsymbol{v}(t)\,\mathrm{d}t \sim \mathcal{N}\left(\boldsymbol{0}, \frac{R}{\Delta t}\right).$$

This ensures area equivalence between the discrete time noise signal and the continuous white noise over Δt [37]. The EKF outcome can therefore be interpreted as a Gaussian approximation $\mathcal{N}(\hat{\bar{z}}(t), P(t))$ of the underlying state estimation probability density function, where the entries $p_{ii}(t)$, $i = n_x + 1, \ldots, n_x + n_p$ of the covariance matrix $P(t)$ corresponding to the parameter estimates are adjusted according to the transformation in (18).

The state and parameter estimates at time $t \geq 0$ in the original coordinates are then given by

$$\hat{\boldsymbol{x}}(t) = \begin{bmatrix} \hat{\bar{z}}_1(t) \\ \vdots \\ \hat{\bar{z}}_{n_x}(t) \end{bmatrix}, \quad \hat{\boldsymbol{p}}(t) = \begin{bmatrix} e^{\hat{\bar{z}}_{n_x+1}(t)} \\ \vdots \\ e^{\hat{\bar{z}}_{n_x+n_p}(t)} \end{bmatrix}, \tag{22}$$

for which by construction it holds that $\hat{p}_i(t) \geq 0$ for all $t \geq 0$ and all $i = 1, \ldots, n_p$.

VI. OPTIMAL INPUT DESIGN FOR PARAMETER ESTIMATION

As pointed out above in the discussion on observability of the extended system (5), the achievement of jointly estimating states and parameters typically depends on the particular input signals. For the purpose of providing the best conditions for joint state and parameter estimation, in the following a strategy for online maximization of the relevant information content in the measurement signals is proposed.

Consider the nominal state dynamics with augmented input and output vectors (including the time derivatives)

$$\bar{\boldsymbol{u}}(t) := \begin{bmatrix} u_1(t) & \cdots & u_1^{(l_1)}(t) & \cdots & u_q^{(l_q)}(t) \end{bmatrix}^{\mathrm{T}}$$

$$\bar{\boldsymbol{y}}(t) := \begin{bmatrix} y_1(t) & \cdots & y_1^{(r_1)}(t) & \cdots & y_{n_y}^{(r_{n_y})}(t) \end{bmatrix}^{\mathrm{T}}, \tag{23}$$

where l_i, r_i indicate the highest derivative of the input u_i or output y_i, respectively. Note that $\bar{\boldsymbol{u}}(t) \in \mathbb{R}^{\bar{n}_u}, \bar{\boldsymbol{y}}(t) \in \mathbb{R}^{\bar{n}_y}$ with $\bar{n}_u = n_u + \sum_{i=1}^{n_u} l_i, \bar{n}_y = n_y + \sum_{i=1}^{n_y} r_i$. Accordingly, in the following we consider

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{p}) + \sum_{i=1}^{n_u} \boldsymbol{g}_i(\boldsymbol{p})u_i, \qquad \boldsymbol{x}(0) = \boldsymbol{x}_0, \tag{24a}$$

$$\bar{\boldsymbol{y}} = \bar{\boldsymbol{h}}(\boldsymbol{x}, \boldsymbol{p}) + \sum_{i=1}^{\bar{n}_u} \boldsymbol{k}_i(\boldsymbol{p})\bar{u}_i, \tag{24b}$$

The vector fields $\boldsymbol{k}_i(\boldsymbol{p})$ arise from extracting the affine parts of these derivatives with respect to the augmented input vector (23).

Now, denote the solution of (24) for a given initial condition $\boldsymbol{x}_0$ and parameter vector $\boldsymbol{p}$ by $\boldsymbol{x}(t;\boldsymbol{x}_0,\boldsymbol{p},\boldsymbol{u})$ and the corresponding augmented measurement vector by $\bar{\boldsymbol{y}}(t;\boldsymbol{x}_0,\boldsymbol{p},\bar{\boldsymbol{u}})$. The goal is to design the inputs u_i in such a way that the information content of the augmented measurements $\bar{\boldsymbol{y}}$ with respect to the parameters $\boldsymbol{p}$ is maximized. For this, consider the sensitivity of the state trajectories with respect to the parameter vector $\boldsymbol{p}$, i.e. the state sensitivity matrix

$$X_{\boldsymbol{p}}(\boldsymbol{x},u,\boldsymbol{p}) := \frac{\partial \boldsymbol{x}(t;\boldsymbol{x}_0,\boldsymbol{p},\boldsymbol{u})}{\partial \boldsymbol{p}} \in \mathbb{R}^{n_x\times n_p}, \tag{25}$$

which propagates over time according to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}X_{\boldsymbol{p}} &= \frac{\partial^2 \boldsymbol{x}(t;\boldsymbol{x}_0,\boldsymbol{p},\boldsymbol{u})}{\partial t\,\partial \boldsymbol{p}} = \frac{\partial}{\partial \boldsymbol{p}}\dot{\boldsymbol{x}}(t;\boldsymbol{x}_0,\boldsymbol{p},\boldsymbol{u})\\ &= \frac{\partial}{\partial \boldsymbol{p}}\Big(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{p}) + \sum_{i=1}^{n_u}\boldsymbol{g}_i(\boldsymbol{p})u_i\Big)\\ &= \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{p}} + \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{p}} + \sum_{i=1}^{n_u}\frac{\partial \boldsymbol{g}_i}{\partial \boldsymbol{p}}(\boldsymbol{p})u_i\\ &= \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}X_{\boldsymbol{p}} + \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{p}} + \sum_{i=1}^{n_u}\frac{\partial \boldsymbol{g}_i}{\partial \boldsymbol{p}}(\boldsymbol{p})u_i\\ &=: F_{X_{\boldsymbol{p}}}(X_{\boldsymbol{p}},\boldsymbol{x},\boldsymbol{p},\boldsymbol{u})\end{aligned} \tag{26}$$

Note that the sensitivity matrix $X_{\boldsymbol{p}}$ is propagated according to linear time-varying dynamics with the dynamics matrix being given by the Jacobian matrix of the vector field $\boldsymbol{f}$ evaluated along the solution trajectory $\boldsymbol{x}(t;\boldsymbol{x}_0,\boldsymbol{p},\boldsymbol{u})$. The initial condition for the parametric sensitivities $X_{\boldsymbol{p}}$ is zero, as the initial state $\boldsymbol{x}_0$ does not depend on $\boldsymbol{p}$.

As the structural identifiability is verified for the (derivative) augmented measurement vector, the sensitivity matrix should also include these derivatives. Therefore, the measurement sensitivities are given by $\bar{Y}_{\boldsymbol{p}} := \frac{\partial \bar{\boldsymbol{y}}}{\partial \boldsymbol{p}} \in \mathbb{R}^{\bar{n}_y\times n_p}$ giving

$$\begin{aligned}\bar{Y}_{\boldsymbol{p}} &= \frac{\partial \bar{\boldsymbol{h}}}{\partial \boldsymbol{x}}X_{\boldsymbol{p}} + \frac{\partial \bar{\boldsymbol{h}}}{\partial \boldsymbol{p}} + \sum_{i=1}^{\bar{n}_u}\frac{\partial \boldsymbol{k}_i}{\partial \boldsymbol{p}}(\boldsymbol{p})\bar{u}_i,\\ &=: H_{\bar{Y}_{\boldsymbol{p}}}(X_{\boldsymbol{p}},\boldsymbol{x},\boldsymbol{p},\bar{\boldsymbol{u}}).\end{aligned} \tag{27}$$

As by consideration (cp. Section V), the measurements are corrupted by noise characterized by the positive definite covariance matrix $R \succ 0$. With this, the Fisher information matrix is defined as [11], [40], and [41]

$$\mathcal{F}(t) := \mathbb{E}\Big[\bar{Y}_{\boldsymbol{p}}^{\mathrm{T}}\bar{Y}_{\boldsymbol{p}}\Big] = \int_0^t H_{\bar{Y}_{\boldsymbol{p}}}^{\mathrm{T}}R^{-1}H_{\bar{Y}_{\boldsymbol{p}}}\,\mathrm{d}\tau, \tag{28}$$

where the integrand is time-varying. Note, that $\mathcal{F}(0)=0$ and $\mathcal{F}$ is positive semi-definite. Equivalently one can determine it solving

$$\dot{\mathcal{F}} = H_{\bar{Y}_{\boldsymbol{p}}}^{\mathrm{T}}R^{-1}H_{\bar{Y}_{\boldsymbol{p}}}. \tag{29}$$

Summarizing the setup, we have a dynamically evolving system that can be written using the state vector $\boldsymbol{\chi} = \left[\boldsymbol{x}^{\mathrm{T}}\ \mathbf{vec}(X_{\boldsymbol{p}})^{\mathrm{T}}\ \mathbf{vec}(\mathcal{F})^{\mathrm{T}}\right]^{\mathrm{T}}$ with the vectorized form of the sensitivity ($X_{\boldsymbol{p}}$) and Fisher information ($\mathcal{F}$) matrices as

$$\dot{\boldsymbol{\chi}} = \begin{bmatrix}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{p}) + \sum_{i=1}^{n_u}\boldsymbol{g}_i(\boldsymbol{p})u_i\\ \mathbf{vec}\left(F_{X_{\boldsymbol{p}}}(X_{\boldsymbol{p}},\boldsymbol{x},\boldsymbol{p},\boldsymbol{u})\right)\\ \mathbf{vec}\left(H_{\bar{Y}_{\boldsymbol{p}}}^{\mathrm{T}}R^{-1}H_{\bar{Y}_{\boldsymbol{p}}}\right)\end{bmatrix} =: \boldsymbol{f}_{\boldsymbol{\chi}}(\boldsymbol{\chi},\boldsymbol{p},\bar{\boldsymbol{u}}). \tag{30}$$

With this, an optimal control problem with fixed end–time T and variable end–state $\boldsymbol{\chi}(T)$ for maximizing information on the finite horizon $[0,T]$ can be formulated. Associated measures for the identifiability of the system are given by scalar functions Φ of the Fisher information matrix, i.e.

$$\mathcal{J}[\boldsymbol{u}] = \Phi\left(\mathcal{F}(T)\right). \tag{31}$$

These functions are typically constructed such that they give inverse properties to the information content.

This leads to the formulation of the optimal control problem

$$\begin{aligned}&\min_{\boldsymbol{u}}\ \ \mathcal{J}[\boldsymbol{u}]\\ &\text{s.t. (30), (31)}\\ &\qquad \boldsymbol{u}_{\min} \le \boldsymbol{u}(t) \le \boldsymbol{u}_{\max}\end{aligned} \tag{32}$$

Its solution is denoted in the following by

$$\boldsymbol{u}^\star = \arg\min_{\boldsymbol{u}} \mathcal{J}[\boldsymbol{u}], \tag{33}$$

and for simplicity in the following we denote $\mathcal{F} = \mathcal{F}(T)$ by applying (28). From the literature, different optimization criteria are known [10]. The cost functions to be minimized in the optimal control problem for different designs are

$$\mathcal{J}_{\mathrm{D}} := \Phi_{\mathrm{D}} = \log\det\mathcal{F}^{-1}, \tag{D-optimal}$$
$$\mathcal{J}_{\mathrm{A}} := \Phi_{\mathrm{A}} = \operatorname{tr}\mathcal{F}^{-1}, \tag{A-optimal}$$
$$\mathcal{J}_{\mathrm{E}} := \Phi_{\mathrm{E}} = -\lambda_{\min}(\mathcal{F}). \tag{E-optimal}$$

The E-Optimality condition (E-optimal) can be modified to (cp., e.g. [42])

$$\mathcal{J}_{\mathrm{E}'} := \Phi_{\mathrm{E}'} = \log\left(\frac{\lambda_{\max}(\mathcal{F})}{\lambda_{\min}(\mathcal{F})}\right), \tag{modified E-optimal}$$

which in optimum is equal to zero. Assuming unitary eigenvectors $\boldsymbol{v}_m$ for the eigenvalues $\lambda_m(\mathcal{F})$, i.e. $\boldsymbol{v}_m^{\mathrm{T}}\boldsymbol{v}_m = 1$ and algebraic multiplicity equal to the geometric multiplicity.[3] Then, the associated matrix derivatives are given by

$$\frac{\partial\Phi_{\mathrm{D}}}{\partial\mathcal{F}} = -\mathcal{F}^{-1}, \tag{34}$$
$$\frac{\partial\Phi_{\mathrm{A}}}{\partial\mathcal{F}} = -\mathcal{F}^{-2}, \tag{35}$$
$$\frac{\partial\Phi_{\mathrm{E}'}}{\partial\mathcal{F}} = \frac{1}{\lambda_{\max}(\mathcal{F})}\boldsymbol{v}_{\max}\boldsymbol{v}_{\max}^{\mathrm{T}} - \frac{1}{\lambda_{\min}(\mathcal{F})}\boldsymbol{v}_{\min}\boldsymbol{v}_{\min}^{\mathrm{T}}. \tag{36}$$

Remark 1: In the numerical implementation, the input trajectory $\boldsymbol{u}(t)$ is discretized as piecewise constant over

[3]The more general case is out of the scope of the current paper.

intervals of length Δt_{mpc}. For each interval, the input is held constant and the ODEs (30) are solved forward in time. This approach does not affect the structure of the gradient estimation or the optimization algorithm, but simplifies the parametrization of the input and the integration of the system dynamics. The required smoothness of the actual input that is required for the output differentiator ($\boldsymbol{u} \in \mathcal{C}^r$, see Section IV) can easily be ensured, e.g., including a filter of the form

$$u_i^{(r)} = -\sum_{j=2}^{n-1} \kappa_j u_i^{(j)} - \kappa_1 (u_i - u_i^*) \tag{37}$$

where $\kappa_j > 0$ are design degree of freedoms and $u_i^\star$ from (33). If the focus is on the critically damped case, the parameters κ_j can be found by evaluation of the polynomial

$$\begin{aligned}(s+\omega)^r &= s^r + \binom{r}{1}\omega s^{r-1} + \binom{r}{2}\omega^2 s^{r-2} \\ &\quad + \cdots + \binom{r}{r-1}\omega^{r-1} s + \omega^r, \\ &= s^r + \kappa_r s^{r-1} + \kappa_{r-1} s^{r-2} + \cdots + \kappa_2 s + \kappa_1,\end{aligned}$$

where ω is the desired pole location and the coefficients $\kappa_j, j = 1, \ldots, r$ can be determined accordingly. Note that using this approach one arrives at $u \in \mathcal{C}^r$, satisfying one condition for the requirement that the $r+1$-th derivative of the outputs remain bounded.

Remark 2: If the filter bandwidth ω is set such, that the filtering of the input drastically impacts the dynamics of the system, the dynamics of the form

$$\dot{\boldsymbol{\sigma}}_i = A_{u,i}\boldsymbol{\sigma}_i + \boldsymbol{b}_{u,i} u_i^\star,$$

with $\boldsymbol{\sigma}_i = \begin{bmatrix} u_i & \dot{u}_i & \cdots & u_i^{(l_i-1)} \end{bmatrix}^{\mathrm{T}}$ and

$$A_{u,i} = \begin{bmatrix} 0 & 1 & \ldots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \ldots & 1 \\ -\kappa_1 & -\kappa_2 & \ldots & -\kappa_r \end{bmatrix},$$
$$\boldsymbol{b}_{u,i} = \begin{bmatrix} 0 & 0 & \ldots & \kappa_1 \end{bmatrix}^{\mathrm{T}},$$

should be included into (30). These filter dynamics can be exactly discretized for a multiple shooting MPC approach. The discretized matrices for constant set-points $u_i^\star$ between sampling instances with sampling time Δt read

$$A_{u,i}^{\mathrm{d}} = \exp(A\Delta t), \quad b_{u,i}^{\mathrm{d}} = -A^{-1}\left(I - A_{u,i}^{\mathrm{d}}\right).$$

Remark 3: Another possibility to achieve smooth input signals would be to dynamically extend the system dynamics by an integrator chain and optimize the highest order derivative over the horizon. This has the downside that the computational complexity of the optimization problem is drastically increased but it allows for free input dynamics not restricted to e.g. a filtered approach.

A. GRADIENT COMPUTATION–INDIRECT GRADIENT METHOD

To be able to solve (32) in real time, a computation of the gradient is needed. Due to the explicit gradient computation, no numerical approximation by disturbing the input and solving the dynamics for every disturbance is needed. This increases the speed of computation and therefore ensures real-time capability. Additionally, sub–optimal solutions with good gradient estimation still lead to higher information content, in contrast to just numerical approximation. For this we first formulate the augmented Lagrangian $\mathcal{L}$ [43], [44]. For convenience, we use the following notation. For matrices U, V of matching dimensions,

$$\operatorname{tr}(U^{\mathrm{T}}V) = \sum_{i,j} U_{ij}V_{ij} =: \langle U, V \rangle$$

defines the Frobenius inner product $\langle \cdot, \cdot \rangle$. With adjoint variables $\lambda \in \mathbb{R}^{n_x}$, $\Lambda_{X_p} \in \mathbb{R}^{n_x \times n_p}$ and $\Lambda_{\mathcal{F}}^{n_p \times n_p}$ we form the Lagrangian $\mathcal{L}$ with the neglection of non–zero input derivatives, i.e. $u_i^{(k)} = 0$, $\forall k \geq 1$ and optimizing for $\boldsymbol{u}^\star$, as[4]

$$\begin{aligned}\mathcal{L}[\boldsymbol{u}, \lambda, \Lambda_{X_p}, \Lambda_{\mathcal{F}}] &:= \Phi\left(\mathcal{F}(T)\right) \\ &+ \int_0^T \lambda^{\mathrm{T}} \left(\boldsymbol{f} + \sum_{i=1}^{n_u} \boldsymbol{g}_i u_i - \dot{\boldsymbol{x}}\right) \mathrm{d}\tau \\ &+ \int_0^T \operatorname{tr}\left(\Lambda_{X_p}^{\mathrm{T}} \left[F_{X_p} - \dot{X}_{\boldsymbol{p}}\right]\right) \mathrm{d}\tau \\ &+ \int_0^T \operatorname{tr}\left(\Lambda_{\mathcal{F}}^{\mathrm{T}} \left[H_{\tilde{Y}_p}^{\mathrm{T}} R^{-1} H_{\tilde{Y}_p} - \dot{\mathcal{F}}\right]\right) \mathrm{d}\tau.\end{aligned}$$

As we have no cost–to–go in pure information maximization, the corresponding Hamiltonian reads

$$\begin{aligned}\mathcal{H} &:= \lambda^{\mathrm{T}} \left(\boldsymbol{f} + \sum_{i=1}^{n_u} \boldsymbol{g}_i u_i\right) \\ &\quad + \operatorname{tr}\left(\Lambda_{X_p}^{\mathrm{T}} F_{X_p}\right) + \operatorname{tr}\left(\Lambda_{\mathcal{F}}^{\mathrm{T}} H_{\tilde{Y}_p}^{\mathrm{T}} R^{-1} H_{\tilde{Y}_p}\right).\end{aligned}$$

Taking the first variation of the Lagrangian $\delta_{\boldsymbol{\chi}}\mathcal{L}$ with respect to the state $\boldsymbol{\chi}$ leads to the adjoint equations given by

$$\dot{\lambda} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}}, \qquad \lambda(T) = \left.\frac{\partial \Phi}{\partial \boldsymbol{x}}\right|_{\mathcal{F}=\mathcal{F}(T)} = \boldsymbol{0}, \tag{38}$$

$$\dot{\Lambda}_{X_p} = -\frac{\partial \mathcal{H}}{\partial X_{\boldsymbol{p}}}, \qquad \Lambda_{X_p}(T) = \left.\frac{\partial \Phi}{\partial X_{\boldsymbol{p}}}\right|_{\mathcal{F}=\mathcal{F}(T)} = 0, \tag{39}$$

$$\dot{\Lambda}_{\mathcal{F}} = -\frac{\partial \mathcal{H}}{\partial \mathcal{F}} = 0, \qquad \Lambda_{\mathcal{F}}(T) = \left.\frac{\partial \Phi}{\partial \mathcal{F}}\right|_{\mathcal{F}=\mathcal{F}(T)}, \tag{40}$$

where the matrix derivative $\frac{\partial \Phi}{\partial \mathcal{F}}$ depends on the chosen metric Φ at hand. Furthermore, $\dot{\Lambda}_{\mathcal{F}} = 0$ as $\mathcal{H}$ is independent of $\mathcal{F}$ and thus this adjoint variable is constant over time, i.e. $\Lambda_{\mathcal{F}}(t) \equiv \Lambda_{\mathcal{F}}(T)$, $\forall t \in [0, T]$. With the derivation from the

[4]Dependencies of the functions are neglected for readability.

appendix, the adjoint equations are then given by

$$\dot{\Lambda}_{X_p} = -\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} \Lambda_{X_p} - 2R^{-1}\frac{\partial \bar{\boldsymbol{h}}}{\partial \boldsymbol{x}} X_p \frac{\partial \Phi(\mathcal{F})}{\partial \mathcal{F}}, \tag{41}$$

with $\Lambda_{X_p}(T) = 0$ and

$$\begin{aligned}\dot{\lambda} = &-\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} \lambda \\ &- \mathbf{vec}\left(\Lambda_{x_p}^{\mathrm{T}} \frac{\partial^2 \boldsymbol{f}}{\partial \boldsymbol{x}\partial \boldsymbol{x}} X_p\right) - \mathbf{vec}\left(\Lambda_{x_p}^{\mathrm{T}} \frac{\partial^2 \boldsymbol{f}}{\partial \boldsymbol{x}\partial \boldsymbol{p}}\right) \\ &- 2\,\mathbf{vec}\left(\left[\frac{\partial \Phi(\mathcal{F})}{\partial \mathcal{F}}\right]^{\mathrm{T}} H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} \frac{\partial H_{\bar{Y}_p}}{\partial \boldsymbol{x}}\right).\end{aligned} \tag{42}$$

where the second and third terms are tensor contractions[5] over the second derivatives of $\boldsymbol{f}$ with the adjoint and sensitivity matrices. The last term is a contraction over the Fisher adjoint, output sensitivities, and their state derivatives.

1) FORWARD–BACKWARD–SWEEP

If the optimal control problem is well-posed, the variation with respect to the adjoint variables is zero and $\delta\mathcal{J} = \delta\mathcal{L}$. The variation with respect to the control variables is given by

$$\delta_{\boldsymbol{u}}\mathcal{L} = \int_0^T \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{u}}\right) \delta \boldsymbol{u} dt. \tag{43}$$

In the optimal point $\boldsymbol{u}^\star$ this variation is zero. As the optimal control problem is assumed to start somewhere different from the optimal point $\boldsymbol{u}_0 \neq \boldsymbol{u}^\star$, we can use the first order optimality condition to obtain a gradient estimate. This procedure is known as the *forward-backward-sweep* [43], [44], [45], [46], [47]. If the input $\boldsymbol{u}$ is applied to the system, the cost function $\mathcal{J}$ can be parameterized only by $\boldsymbol{u}$ when solving the dynamics forward in time. Then the adjoint equations can be solved backward in time, leading to the adjoint variables λ, Λ_{X_p}, $\Lambda_{\mathcal{F}}$ over the horizon. Then the reduced gradient can be estimated with first order accuracy by

$$\nabla_{\boldsymbol{u}}\mathcal{J}(t) = \int_0^T \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{u}}\right) dt. \tag{44}$$

The derivation of the gradient of the Hamiltonian with respect to the control variables can be found in the appendix.[6]

Remark 4: The forward-backward-sweep algorithm for the gradient computation can also be accomplished by the reverse mode of automatic differentiation [48]. This can be useful, if the right-hand side of the system dynamics include e.g. jumps or case statements which cannot be symbolically differentiated. We choose to use the symbolic computations, as e.g. `CasADi` does not include taking derivatives of the trace of the inverse matrix nor the eigenvalues, as these cannot be symbolically computed for big matrices. But the derivative of the minimal and maximal eigenvalues with respect to the matrix are known, thus this derivative can be symbolically computed without effort. Nonetheless, both symbolic differentiation and automatic differentiation are far more efficient than computation of the gradients via finite differences and impose significant speedup during runtime.

Remark 5: To enforce box constraints on the inputs, i.e., $u_{i,\min} \leq u_i(t) \leq u_{i,\max}$, the unconstrained update $\boldsymbol{u}(t)$ obtained from a gradient step can be projected onto the feasible set (projected gradient method). The projection operator $\mathcal{P}_{[\boldsymbol{u}_{\min},\boldsymbol{u}_{\max}]}$ is defined component-wise as

$$\mathcal{P}_{[u_{i,\min},u_{i,\max}]}(u_i(t)) := \min\left(\max\left(u_i(t), u_{\min,i}\right), u_{\max,i}\right),$$

where $u_i^*(t)$ is the unconstrained input for the i-th input channel, and $u_{\min,i}$, $u_{\max,i}$ are the lower and upper bounds, respectively. This projection ensures that the input always satisfies the box constraints.

VII. CASE STUDY—A THERMAL PROCESS EXAMPLE

In the following consider the problem of jointly estimating states and parameters for a thermal conduction process as provided by the Arduino-based temperature control lab [49]. The system consists of two coupled heating devices with temperature sensors and can be modeled using classical energy balances considering heat transport by conduction and radiation, leading to the following model

$$\begin{aligned} mc_p\dot{T}_1 = &\, UA_{12}(T_2 - T_1) + \sigma\epsilon A_{12}(T_2^4 - T_1^4) \\ &+ UA_s(T_s - T_1) + \sigma\epsilon A_s(T_s^4 - T_1^4) + \alpha_1 u_1 \end{aligned} \tag{45a}$$

$$\begin{aligned} mc_p\dot{T}_2 = &\, UA_{12}(T_1 - T_2) + \sigma\epsilon A_{12}(T_1^4 - T_2^4) \\ &+ UA_s(T_s - T_2) + \sigma\epsilon A_s(T_s^4 - T_2^4) + \alpha_2 u_2 \end{aligned} \tag{45b}$$

with T_1, T_2 being the temperature of subsystem 1 and 2, respectively, m the associated mass, specific heat capacity c_p, U the heat transfer coefficient, A_{12} the interchange area between subsystem 1 and 2, σ the Stefan Boltzmann (heat radiation) constant, ϵ the emissivity of the material, A_s the ambient interface area, T_s the ambient temperature, α_1, α_2 the input gain coefficients associated with the heating elements with activation value $u_1, u_2 \in \mathcal{U} := [0, 100]$.

Some of the parameters can be identified well by the manufacturer, while others show some variations and should be identified accordingly using experimental data. For this purpose, in the following the parameter vector is split into the two components $\boldsymbol{p}_k$ (known) and $\boldsymbol{p}$ (unknown). In the present study we consider

$$\boldsymbol{p} := \begin{bmatrix} U & A_{12} & A_s & \alpha_1 & \alpha_2 \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}_+^{n_p}, \tag{45c}$$

with $n_p = 5$. In the following we consider that both temperatures are measured, i.e.

$$y_i = T_i + v_i, \qquad i = 1, 2, \quad v_i \sim \mathcal{N}(0, r_i) \tag{45d}$$

with normally distributed measurement noise v_i and $n_y = 2$.

[5] Note, that the derivatives $\frac{\partial^2 \boldsymbol{f}}{\partial \boldsymbol{x}\partial \boldsymbol{x}}$, $\frac{\partial^2 \boldsymbol{f}}{\partial \boldsymbol{x}\partial \boldsymbol{p}}$, $\frac{\partial H_{\bar{Y}_p}}{\partial \boldsymbol{x}}$ are tensors. The present operation represents a *tensor contraction*, i.e., an index-wise summation over the common dimensions of the involved tensors. For example, for a second-order derivative tensor $(\partial^2 f_i/\partial x_j \partial x_k)$ and a sensitivity matrix $X_{p,kl}$, the contraction is given by $\sum_{j,k} \frac{\partial^2 f_i}{\partial x_j \partial x_k} X_{p,kl}$.

[6] For numerical implementation this gradient is discretized with the trapezoidal rule and fixed length piecewise constant inputs.

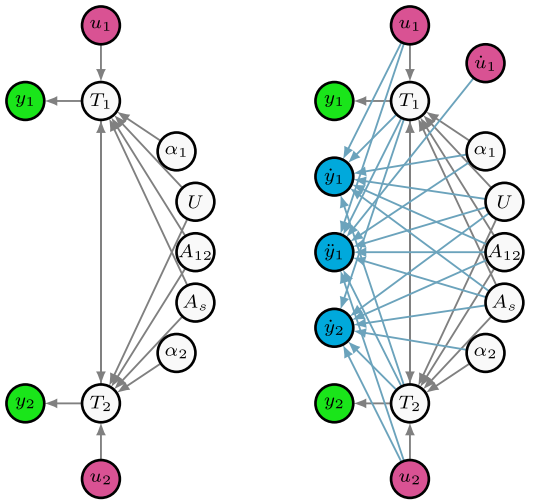

FIGURE 1. Structure graph associated to the extended system dynamics (5) (left) and the one with additional measurement vertices associated to the time derivatives of the outputs (right).

A. STRUCTURAL OBSERVABILITY ANALYSIS

As can be seen by inspecting the structure graph of the extended dynamics (5) for the case at hand, as shown in Fig. 1 (left) there are no information paths from the parameter vertices towards other vertices, implying that they are acting as information sinks. The next step is to consider the time derivatives of the outputs as additional information sources. Choosing the example-specific augmented measurements as

$$\{y_1, \quad y_2, \quad \dot{y}_1, \quad \dot{y}_2, \quad \ddot{y}_1\}, \tag{46}$$

it can be seen, that the structural rank condition can be verified by analyzing the extended structure graph is shown in Fig. 1 (right), from which it becomes clear that all vertices can now be reached by information paths starting from the five output vertices, considering $r_1 = 2$ and $r_2 = 1$, yielding $\mathcal{M} = 5 = n_p$ by (13). The new connections are colored in blue. The evaluation of the structural rank condition (11) based on (46) actually leads to

$$\text{s}-\text{rank}\begin{bmatrix}\mathfrak{J}_z\\ \mathfrak{H}_z\end{bmatrix} = \text{s}-\text{rank}\begin{bmatrix}* & *\\ 0 & 0\\ * & 0\end{bmatrix} = n_x + n_y \tag{47}$$

$$\text{s}-\text{rank}\begin{bmatrix}\mathfrak{J}_z\\ \mathfrak{H}_{z,e}\end{bmatrix} = \text{s}-\text{rank}\begin{bmatrix}* & *\\ 0 & 0\\ * & 0\\ * & *\end{bmatrix} = n_x + \underbrace{2n_y + 1}_{=n_p} \tag{48}$$

showing that in principle, by including the time derivatives of the measurements it is possible to satisfy the condition for local structural observability.

Having the above structural observability analysis as point of departure, and in particular the fact that the time derivatives of the outputs are required in addition to the direct measurements of the two temperatures T_1, T_2, the proposed state and parameter estimation scheme proposed here consists in first estimating the time derivatives $\dot{y}_1, \dot{y}_2, \ddot{y}_1$ of the outputs using exact differentiation (see Section IV) and then using the extended measurement vector

$$\boldsymbol{y}_e := \begin{bmatrix}\boldsymbol{y}\\ \dot{\boldsymbol{y}}\\ \ddot{y}_1\end{bmatrix} = \boldsymbol{h}_e(z) = \begin{bmatrix}\boldsymbol{h}(z)\\ L_f\boldsymbol{h}(z)\\ L_f^2 h_1(z)\end{bmatrix} \tag{49}$$

in an extended Kalman filter discussed in Section V.

B. ALGORITHM

In the following, we outline the steps of the proposed closed-loop MPC Algorithm 1 for optimal input design and joint state and parameter estimation. In the implementation of the numerical study and the experimental study, the D-optimality criterion is chosen for optimization. For noisy measurements, a discrete-time moving average filter with forgetting factor α of the form [50]

$$\varpi_k = \alpha\varpi_{k-1} + 1, \tag{50a}$$

$$\hat{\boldsymbol{y}}_k = \left(1 - \frac{1}{\varpi_k}\right)\hat{\boldsymbol{y}}_{k-1} + \frac{1}{\varpi_k}\boldsymbol{y}(t_k) \tag{50b}$$

can be applied on a finer grid with the measurements $\boldsymbol{y}(k\,\Delta t_{\text{meas}})$. The system dynamics are propagated forward in time using a fourth-order Runge–Kutta (RK4) scheme on a refined time grid of $\Delta t_{\text{mpc}}/n_{\text{Sub}}$, while the adjoint equations are integrated backward in time using an explicit Euler scheme. The control inputs are held constant over each MPC sampling interval Δt_{mpc}, and the prediction horizon is given by $T_{\text{p}} = N\,\Delta t_{\text{mpc}}$. For convenience, we define a control interval $T_{\text{c}} = M\,\Delta t_{\text{mpc}}$, $M < N$ over which the optimizer may treat the inputs as a block for warm-starting or updating. This results in a piecewise-constant input trajectory $\boldsymbol{u}^\star(t)$, which may change every Δt_{mpc}, defined for $t \in [k\,\Delta t_{\text{mpc}}, (k+1)\,\Delta t_{\text{mpc}}]$, $k = 0, 1, \ldots, k_{\max}$.

The optimization is done within `Matlab` with `fmincon` and the interior–point optimizer. For warm starting the optimizer, the piecewise-constant control sequence of horizon length N_h is shifted forward by M control intervals. The first M control inputs are discarded, and the last control action $\boldsymbol{u}_{N_h-1}$ is repeated for the remaining M time instances Δt_{mpc}. This shift operation can be written as

$$U = \begin{bmatrix}\boldsymbol{u}_0\\ \boldsymbol{u}_1\\ \vdots\\ \boldsymbol{u}_{N_h-2}\\ \boldsymbol{u}_{N_h-1}\end{bmatrix} \quad\Longrightarrow\quad \mathcal{S}_{\text{rep},M}(U) = \begin{bmatrix}\boldsymbol{u}_M\\ \boldsymbol{u}_{M+1}\\ \vdots\\ \boldsymbol{u}_{N_h-1}\\ \boldsymbol{u}_{N_h-1}\\ \vdots\\ \boldsymbol{u}_{N_h-1}\end{bmatrix}, \tag{51}$$

where the last control $\boldsymbol{u}_{N_h-1}$ is repeated M times.

The convergence of the optimization can be monitored by checking the optimality conditions of the Hamiltonian, specifically ensuring that the gradient of the cost function with respect to the control inputs is smaller than a pre–defined

threshold, i.e.

$$\int_0^{T_p} \left\| \frac{\partial \mathcal{H}}{\partial \boldsymbol{u}}\bigg|_{\boldsymbol{u}=\boldsymbol{u}^*} \right\| \mathrm{d}t < \epsilon.$$

Additionally, the change in the cost function value between iterations can be used as a convergence criterion. If the change is below a predefined threshold, the algorithm can be considered to have converged. For real–time capability, a suboptimal solution can be accepted after a fixed number of iterations $N_{\max}$ or a maximum computation time $t_{\max}$, ensuring that the control input is updated within the required time frame.

Having a suitable estimate $\hat{\boldsymbol{z}}(t) = [\hat{\boldsymbol{x}}^{\mathrm{T}}(t) \quad \hat{\boldsymbol{p}}^{\mathrm{T}}(t)]^{\mathrm{T}}$, one can solve an associated moving horizon optimal predictive control problem considering a prediction horizon T_p and control horizon $T_c \leq T_p$, over which the calculated control law will be implemented before a new prediction and optimization step starts with $\hat{\boldsymbol{p}}_{u,k} := \hat{\boldsymbol{p}}_u(kT_c)$ and $\hat{\boldsymbol{x}}(kT_c)$, $\hat{\boldsymbol{p}}_u(kT_c)$ provided by the estimates at $t = kT_c$ from the estimator (21) with (14).

C. SIMULATION RESULTS

Algorithm 1 is implemented using Matlab–Simulink with control update time step of $\Delta t_{\mathrm{mpc}} = 300$ s for the MPC, prediction horizon $T_p = 3000$ s, control horizon $T_c = 900$ s and time grid refinement factor $n_{\mathrm{sub}} = 8$. For gradient estimation the system state $\boldsymbol{\chi}$ is forwards propagated with piecewise constant inputs with Runge–Kutta 4 discretization. The backwards solution of the adjoint equations for λ, Λ_{X_p}, $\Lambda_{\mathcal{F}}$ is implemented with explicit Euler discretization backwards in time.

The optimization is carried out as interior–point optimization within `fmincon`, where the gradients are computed by the forward-backwards-sweep with the constant input approximation of (44) via trapezoidal rule.

The exact differentiator is implemented using its discrete-time equivalent as proposed in [23]. Here, the uniform robust exact differentiator was chosen in the Matlab toolbox with the parameter set $r = 1$, filter order 2 and $\mu = 10$. The smoothing of the input was carried out using a fourth order filter with $\omega = 0.1$ to ensure smoothness of the derivative estimates. The first two derivatives of y_1 and the first derivative of y_2 are estimated with the uniform robust exact differentiator. The measurement sampling times are set to $\Delta t = \Delta t_{\mathrm{meas}} = 1$ s. The EKF parameters were subject to tuning and resulted in $P_0 = \mathrm{blkdiag}[P_{x,0}, P_{p,0}]$ with $P_{x,0} = I_2$ and

$$P_{p,0} = \mathrm{diag}\left[3.5 \cdot 10^{-5} 12 \cdot 10^{-4} 1 \cdot 10^{-5} 5 \cdot 10^{-5}\right].$$

Furthermore we set $Q = \mathrm{blkdiag}[Q_x, Q_p]$ with $Q_x = 0.01 I_2$ and $Q_p = 0$ and $R = 2I_5 \cdot 10^{-6}$ in the noise–free case with identity matricies $I_k \in \mathbb{R}^{k \times k}$. The measurement covariance should only compensate for discretization errors of the simulation in the noise–free case. The initial conditions

Algorithm 1 Closed-Loop MPC for Optimal Input Design and Joint State/Parameter Estimation

Require: Initial state estimate $\hat{\boldsymbol{x}}_0$, parameter estimate $\hat{\boldsymbol{p}}_0$, initial input trajectory $\boldsymbol{u}^0(t)$, time step Δt, time grid refinement factor n_{Sub}, control update time step Δt_{mpc}, control horizon $T_c = M \Delta t_{\mathrm{mpc}}$, prediction horizon $T_p = N \Delta t_{\mathrm{mpc}}$, experiment duration $N_{\max}$, input constraints $\boldsymbol{u}_{\min}, \boldsymbol{u}_{\max}$, smoothing parameters k_u, k_d, κ_i, EKF parameters P_0, Q, R, forgetting factor α.

Ensure: Piecewise constant optimal input trajectory $\boldsymbol{u}^*$, updated state and parameter estimates

for each control interval $N = 1, 2, \ldots, N_{\max}$ **do**

Measurement: Acquire output measurements $\boldsymbol{y}(t_k)$

Filtering: Apply moving average filter (50) to measurements $\boldsymbol{y}(t_k)$, $\hat{\boldsymbol{y}}_k \leftarrow \mathrm{MA}(\hat{\boldsymbol{y}}_{k-1}, \boldsymbol{y}(t_k), \alpha)$

Sliding Mode Differentiation: Estimate output derivatives $\hat{\boldsymbol{y}}^{(l_k)}(t_k)$ using robust exact differentiator based on moving averaged measurements $\hat{\boldsymbol{y}}_k$

State/Parameter Estimation: Update state and parameter estimates $(\hat{\boldsymbol{x}}(t_k), \hat{\boldsymbol{p}}(t_k))$ using EKF with extended output $\boldsymbol{y}_e = [\boldsymbol{y}(t_k); \hat{\boldsymbol{y}}_k; \ldots \hat{\boldsymbol{y}}^{(l_k)}]$

Prediction: Predict future state and parameter trajectories over horizon T_p using current estimates

Sensitivity Analysis: Compute sensitivity matrices X_p, $\bar{Y}_p$ and Fisher information $\mathcal{F}$ along predicted trajectory

MPC Optimization: Solve discretized

$$\min_{\boldsymbol{u}} \Phi(\mathcal{F}(T_p))$$

subject to system dynamics and input constraints $\boldsymbol{u}_{\min} \leq \boldsymbol{u} \leq \boldsymbol{u}_{\max}$.

Initial guess/Warm start: choose feasible input sequence $U^{(0)}$ (e.g. shift previous solution or nominal), then enforce feasibility

$$U^{(0)} \leftarrow \mathcal{P}_{[\boldsymbol{u}_{\min}, \boldsymbol{u}_{\max}]}(U^{(0)}).$$

Start with $i = 1$:

while not converged **do**

Compute Fisher information $\mathcal{F}_i$ (28)

Compute cost $\mathcal{J}_i$ (31)

Compute gradient $\nabla \mathcal{J}_i$ (44)

Update input trajectory based on `fmincon` interior–point optimization

$i \leftarrow i + 1$

end while

Input Smoothing: Apply input filter (37) to ensure smooth input

Apply Control: Implement input $\boldsymbol{u}(t)$ over control horizon T_c

Shift Control (warm start): Get next initial guess $U^{(0)}_{\mathrm{next}}$ by applying (51)

Update MPC initial condition: Get starting point $\hat{\boldsymbol{x}}(N \Delta t_{\mathrm{mpc}})$ for prediction based on EKF state estimate

end for

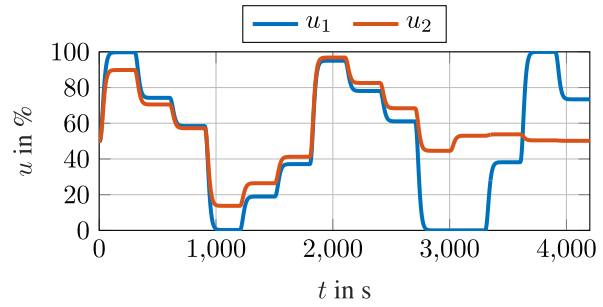

FIGURE 2. Smoothed optimal input signal for the simulation.

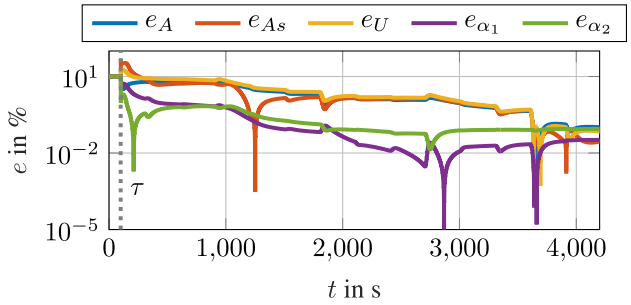

FIGURE 3. Relative parameter errors for the simulation in percent.

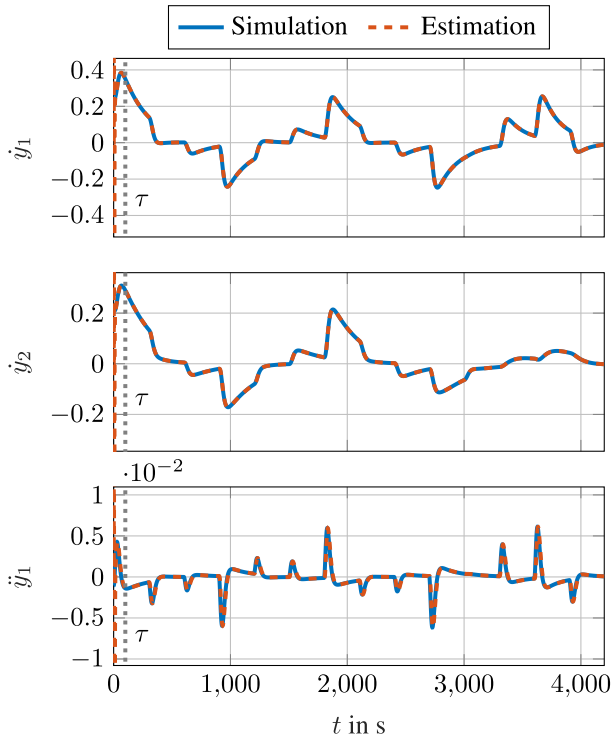

FIGURE 4. Derivative estimations for the simulation.

were set to $T_{1,0} = T_a + 2$ K and $T_{1,0} = T_a - 2$ K. The initial nominal parameter errors are 10 percent. As there is no noise present, no moving average filter is employed in the simulation. The used parameters for the simulation can be found in Table 1.

The correction via the EKF with the derivatives is enabled after $\tau = 100$ s, as there the derivative estimations converged. The resulting optimal input is depicted in Fig. 2, the scaled absolute parameter errors

$$e_i(t) = \left| \frac{\hat{p}_i(t) - p_i}{p_i} \right| \cdot 100\%$$

in Fig. 3 and the derivative estimations in Fig. 4.

D. EXPERIMENTAL VALIDATION

The proposed scheme for online optimal input design for joint state and parameter estimation was implemented using the Arduino-based temperature control lab [49]. Here again, Algorithm 1 is implemented using Matlab–Simulink with sampling interval of $\Delta t_{\text{mpc}} = 300$ s, prediction horizon $T_p = 3000$ s, control horizon $T_c = 900$ s and time grid refinement factor $n_{\text{sub}} = 8$. The exact differentiator is implemented using its discrete-time equivalent as proposed in [23]. Here, the uniform robust exact differentiator was chosen in the Matlab toolbox with the parameter set $r = 1$, filter order 2 and $\mu = 100$. The smoothing of the input was carried out using the same fourth order filter as in the simulation.

The initial parameter values were taken from the literature (see, e.g., [49]). The ambient temperature during the experiment was around $T_a = 23°$ C. The initial conditions for $T_1 = T_2$ were set to T_a, representing the equilibrium point for $u = 0$. The covariance of the EKF was initialized with

$$P_0 = \text{diag}\big[0.01,\ 0.01,\ 1 \cdot 10^{-4},\ 2 \cdot 10^{-4},\ 1 \cdot 10^{-2},\ 2 \cdot 10^{-3},\ 2 \cdot 10^{-3}\big],$$

process noise covariance matrix

$$Q = \text{diag}([1, 1, 0, 0, 0, 0]) \cdot 10^{-3},$$

and measurement covariance matrix

$$R = \text{diag}([1, 1, 0.5, 0.5, 0.25]).$$

A moving average filter with forgetting factor $\alpha = 0.9$ was applied to the measurements to be able to better estimate the time derivatives in the quantized measurement signals of the temperature control lab sensors. The measurement are acquired at $\Delta t_{\text{meas}} = 0.2$ s and the sliding mode differentiator and EKF are updated at a rate of $\Delta t_{\text{est}} = 1$ s. Like in the simulation study, the correction via the EKF is enabled after $\tau = 100$ s, as there the derivative estimations converged.

The results for the closed loop identification problem applied to the real experiment are depicted in Figs. 5 to 7. The optimal piecewise constant inputs obtained by solving the optimization problem within Algorithm 1 are smoothed and the filtered inputs with the resulting state estimates from the EKF (bottom) are shown. In Fig. 6 the noisy derivative estimates are depicted with estimation delay τ. Similarly, the parameter estimates are depicted in Fig. 7. Here the parameter estimates (blue lines) are shown with associated $\pm 3\sqrt{P_{ii}}$ certainty interval (green dotted lines) calculated from the

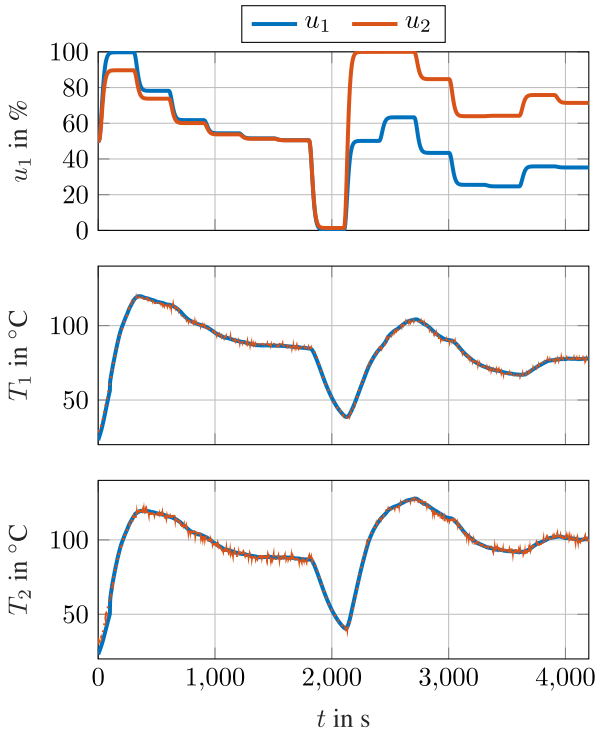

FIGURE 5. Smoothed optimal input signals and state estimates for the experimental validation. The EKF state estimates are shown in blue (continuous lines) and the measurements in orange (dotted lines).

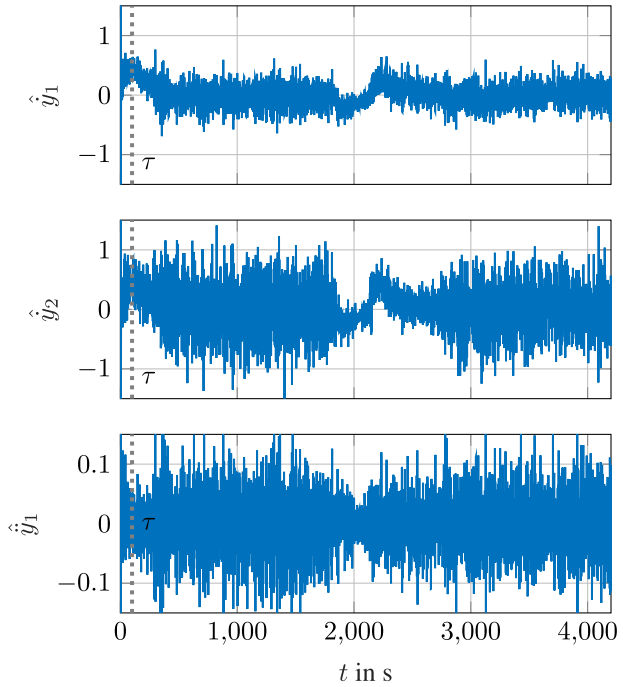

FIGURE 6. Derivative estimations for the experimental validation.

solution of the Ricatti equation in the EKF (21). It should be noted, that the covariance is additive in the logarithmic space, which together with the parameter estimate has to be transformed back to the linear space for the plots.

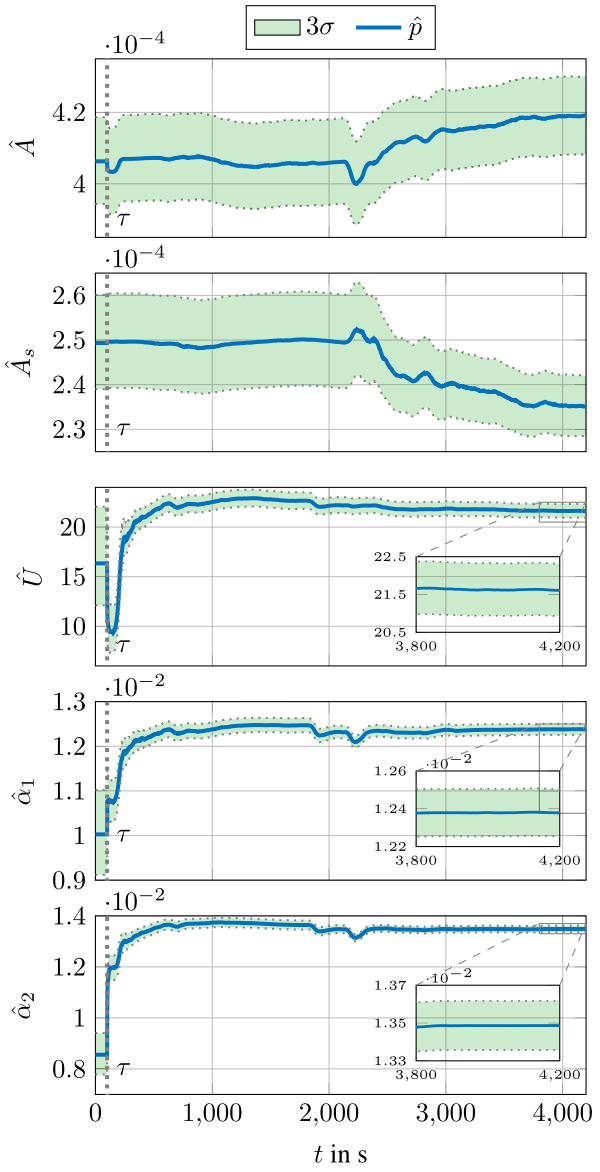

FIGURE 7. Parameter estimates and 3σ confidence intervals for the experimental validation.

For the evaluation of the obtained parameters the mean value $\bar{p}$ of the last 400 s was considered, to filter out measurement noise propagation and fluctuations. The values for $\bar{\boldsymbol{p}}$ and $\boldsymbol{p}_{\mathrm{k}}$ are depicted together with the simulation parameters $\boldsymbol{p}_{\mathrm{sim}}$ in Table 1.

The obtained parameters $\bar{\boldsymbol{p}}$ were then used in a simulation run, using the previous values $\boldsymbol{u}^*$ from the experiment and measured ambient temperature values. The resulting temperatures are compared to the experimental ones. These results are shown in Figure 8 (blue continuous lines) together with the measurements from the preceding experiment (black dotted lines). One observes a relatively good fit during the complete run, with some offset during strong transients. To further evaluate the quality of fit the root mean squared error (RMSE) and the maximum deviations were calculated and summarized in Table 2.

TABLE 1. Parameter values for simulation and experiment.

	Name	Value	Unit
$\boldsymbol{p}_{\mathrm{k}}$	m	0.004	kg
	c_p	500	J/(kg·K)
	T_{a}	297.14	K
	σ	5.67×10^{-8}	J/(s·m^2·K^4)
	ϵ	0.45	–
$\boldsymbol{p}_{0,\mathrm{exp}} = \boldsymbol{p}_{\mathrm{sim}}$	A	0.0004	m^2
	A_s	0.0002	m^2
	U	16.3392	J/(s·m^2·K)
	α_1	0.01	J/s
	α_2	0.0086	J/s
$\bar{\boldsymbol{p}}$	A	4.1878×10^{-4}	m^2
	A_s	2.3546×10^{-4}	m^2
	U	21.6341	J/(s·m^2·K)
	α_1	0.0124	J/s
	α_2	0.0135	J/s

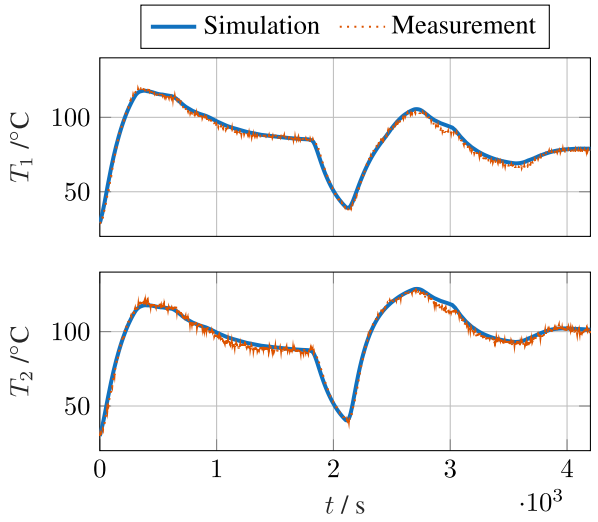

FIGURE 8. Experimental results (orange dotted lines) and simulation without EKF correction (blue continuous lines) using optimized parameter values and the experimental input.

TABLE 2. Fitting results from the experimental evaluation.

Temperature	RMSE in K	$\max \lvert y_i - x_i \rvert$ in K
T_1	1.9	6.7
T_2	2.2	7.8

VIII. CONCLUSION

The problem of optimal input design for joint state and parameter estimation is discussed for nonlinear affine-input systems with positive parameters. It is shown, that with incorporation of output derivative estimates the structural observability improves and therefore a sufficient condition of the identifiability can be fulfilled. For the derivative estimation, higher-order sliding mode differentiators are chosen. Furthermore, to ensure parameter positivity a transformation is employed and a continuous-discrete time EKF is designed for parameter estimation. For optimal input design, a metric based on the Fisher information is chosen and the first-order optimality conditions are derived to subsequently solve the system and its adjoint states in forward-backwards mode and give a symbolically calculable gradient within the forward-backwards-sweep method.

This general scheme is tested in simulation and experiment on a thermal conduction process consisting of two coupled heating elements and two temperature sensors. A model based on energy balances is employed with five unknown parameters, which are estimated online using the positive EKF, enriched by the information of the time derivative estimation of the measured outputs. The structural identifiability is established for the considered setup with the additional measurement information. The optimal input design is based on an moving horizon optimal control scheme with discretized and smoothed inputs. These smoothed in order to ensure the desired convergence behavior of the output differentiator. The numerical study shows the convergence speed of the algorithm in the noise–free case. The functioning of the proposed scheme is then experimentally evaluated using the Arduino-based temperature control lab, leading to convincing results.

The proposed method can easily be adapted for other systems, by considering the structural identifiability condition based on structural observability analysis and output differentiation, taking into account that the input signals are sufficiently smoothed. For large scale models, the usage of automatic differentiation instead of the symbolic adjoint computations for gradient estimation would be needed to be computational efficient.

APPENDIX
DERIVATION OF THE ADJOINT EQUATIONS
A. SECOND ADJOINT EQUATION

Plugging in the definitions of the dynamics of the adjoint sensitivity state variables (39) yields

$$\dot{\Lambda}_{X_p} = -\frac{\partial}{\partial X_p} \operatorname{tr}\left(\Lambda_{X_p}^{\mathrm{T}} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} X_p\right) - \frac{\partial}{\partial X_p} \operatorname{tr}\left(\frac{\partial \Phi(\mathcal{F})}{\partial \mathcal{F}}^{\mathrm{T}} \left(\frac{\partial \bar{\boldsymbol{h}}}{\partial \boldsymbol{x}} X_p\right)^{\mathrm{T}} R^{-1} \left(\frac{\partial \bar{\boldsymbol{h}}}{\partial \boldsymbol{x}} X_p\right)\right).$$

The linear form has the derivative [51]

$$-\frac{\partial}{\partial X_p} \operatorname{tr}\left(\Lambda_{X_p}^{\mathrm{T}} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} X_p\right) = -\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} \Lambda_{X_p},$$

and the quadratic form under the cyclic properties of the trace has the derivative [51]

$$\frac{\partial}{\partial X_p} \operatorname{tr}\left(\frac{\partial \Phi(\mathcal{F})}{\partial \mathcal{F}}^{\mathrm{T}} \left(\frac{\partial \bar{\boldsymbol{h}}}{\partial \boldsymbol{x}} X_p\right)^{\mathrm{T}} R^{-1} \left(\frac{\partial \bar{\boldsymbol{h}}}{\partial \boldsymbol{x}} X_p\right)\right) = 2R^{-1} \frac{\partial \bar{\boldsymbol{h}}}{\partial \boldsymbol{x}} X_p \frac{\partial \Phi(\mathcal{F})}{\partial \mathcal{F}}.$$

B. FIRST ADJOINT EQUATION

The adjoint equation for the system states (42) arises from

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}}\Bigg[\lambda^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{x},\boldsymbol{p}) + \operatorname{tr}\left(\Lambda_{X_p}^{\mathrm{T}}\left[F_{X_p}(X_p,\boldsymbol{x},\boldsymbol{p},\boldsymbol{u})\right]\right) + \operatorname{tr}\left(\Lambda_{\mathcal{F}}^{\mathrm{T}}\left[H_{\bar{Y}_p}^{\mathrm{T}}(X_p,\boldsymbol{x},\boldsymbol{p},\boldsymbol{u}) R^{-1} H_{\bar{Y}_p}(X_p,\boldsymbol{x},\boldsymbol{p},\boldsymbol{u})\right]\right)\Bigg].$$

With the jacobian being the transpose of the gradient compute

$$\frac{\partial}{\partial \boldsymbol{x}} \lambda^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{x},\boldsymbol{p}) = \left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} \lambda.$$

Additionally, the sum formula for the Frobenius inner product gives

$$\operatorname{tr}(\Lambda_{X_p}^{\mathrm{T}} F_{X_p}) = \sum_{i=1}^{n_x}\sum_{j=1}^{n_p} \Lambda_{X_p}^{ij} F_{X_p}^{ij},$$

$$\operatorname{tr}(\Lambda_{\mathcal{F}}^{\mathrm{T}} H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} H_{\bar{Y}_p}) = \sum_{i=1}^{n_x}\sum_{j=1}^{n_p} \Lambda_{\mathcal{F}}^{ij} \left(H^{\mathrm{T}} R^{-1} H\right)^{ij},$$

where $F_{X_p}^{ij}$ is the (i,j)-th entry of the matrix $F_{X_p} \in \mathbb{R}^{n_x \times n_p}$. The derivatives with respect to $\boldsymbol{x}$ are given by

$$\frac{\partial}{\partial \boldsymbol{x}} \operatorname{tr}(\Lambda_{X_p}^{\mathrm{T}} F_{X_p}) = \sum_{i=1}^{n_x}\sum_{j=1}^{n_p} \Lambda_{X_p}^{ij} \frac{\partial F_{X_p}^{ij}}{\partial \boldsymbol{x}},$$

and

$$\frac{\partial}{\partial \boldsymbol{x}} \operatorname{tr}\left(\Lambda_{\mathcal{F}}^{\mathrm{T}}[H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} H_{\bar{Y}_p}]\right) = \sum_{i=1}^{n_p}\sum_{j=1}^{n_p} \Lambda_{\mathcal{F}}^{ij} \frac{\partial}{\partial \boldsymbol{x}} \left(H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} H_{\bar{Y}_p}\right)^{ij},$$

$$= \sum_{i=1}^{n_p}\sum_{j=1}^{n_p} \left(\frac{\partial \Phi(\mathcal{F})}{\partial \mathcal{F}}\right)^{ij} \frac{\partial}{\partial \boldsymbol{x}} \left(H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} H_{\bar{Y}_p}\right)^{ij},$$

Explicitly computing the derivative of $F_{X_p}^{ij}$ leads to

$$\frac{\partial F_{X_p}^{ij}}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}}\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x},\boldsymbol{p}) X_p + \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{p}}(\boldsymbol{x},\boldsymbol{p}) + \sum_{k=1}^{n_u} \frac{\partial \boldsymbol{g}_k}{\partial \boldsymbol{p}}(\boldsymbol{p}) u_k\right)^{ij}$$

$$= \begin{bmatrix} \frac{\partial F_{X_p}^{ij}}{\partial x_1} & \frac{\partial F_{X_p}^{ij}}{\partial x_2} & \cdots & \frac{\partial F_{X_p}^{ij}}{\partial x_{n_x}} \end{bmatrix}^{\mathrm{T}},$$

where each component is given by

$$\frac{\partial F_{X_p}^{ij}}{\partial x_k} = \sum_{s=1}^{n_x} \frac{\partial^2 f_i}{\partial x_k \partial x_s}(\boldsymbol{x},\boldsymbol{p}) X_{p,sj} + \frac{\partial^2 f_i}{\partial x_k \partial p_j}(\boldsymbol{x},\boldsymbol{p})$$

For the third part expand the (i,j) entry of the matrix product $H^{\mathrm{T}} R^{-1} H$ as

$$\left(H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} H_{\bar{Y}_p}\right)^{ij} = \sum_{k=1}^{n_y}\sum_{l=1}^{n_y} H_{\bar{Y}_p}^{ki} (R^{-1})_{kl} H_{\bar{Y}_p}^{lj}.$$

Differentiation with respect to the state vector component x_k gives

$$\frac{\partial}{\partial x_k}\left(H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} H_{\bar{Y}_p}\right)^{ij}$$

$$= \sum_{l=1}^{n_y}\sum_{m=1}^{n_y}\left[\frac{\partial H_{\bar{Y}_p}^{li}}{\partial x_k}(R^{-1})^{lm} H_{\bar{Y}_p}^{mj} + H_{\bar{Y}_p}^{li}(R^{-1})^{lm} \frac{\partial H_{\bar{Y}_p}^{mj}}{\partial x_k}\right].$$

For each entry of H compute

$$\frac{\partial H_{\bar{Y}_p}^{ij}}{\partial x_k} = \sum_{r=1}^{n_x} \frac{\partial^2 \bar{h}_i}{\partial x_k \partial x_r}(\boldsymbol{x},\boldsymbol{p}) X_{p,rj} + \frac{\partial^2 \bar{h}_i}{\partial x_k \partial p_j}(\boldsymbol{x},\boldsymbol{p}) + \sum_{s=1}^{n_u} \frac{\partial^2 k_{s,i}}{\partial x_k \partial p_j}(\boldsymbol{x},\boldsymbol{p})\, u_s.$$

This leads us to

$$\frac{\partial}{\partial x_k}\left(H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} H_{\bar{Y}_p}\right)^{ij}$$

$$= \sum_{m=1}^{n_y}\sum_{l=1}^{n_y}(R^{-1})^{ml}\Bigg\{\Bigg[\sum_{r=1}^{n_x}\frac{\partial^2 \bar{h}_m}{\partial x_k \partial x_r} X_{p,ri} + \frac{\partial^2 \bar{h}_m}{\partial x_k \partial p_i} + \sum_{s=1}^{n_u}\frac{\partial^2 k_{s,m}}{\partial x_k \partial p_i} u_s\Bigg] H_{\bar{Y}_p}^{lj} + H_{\bar{Y}_p}^{mi}\Bigg[\sum_{r=1}^{n_x}\frac{\partial^2 \bar{h}_l}{\partial x_k \partial x_r} X_{p,rj} + \frac{\partial^2 \bar{h}_l}{\partial x_k \partial p_j} + \sum_{s=1}^{n_u}\frac{\partial^2 k_{s,l}}{\partial x_k \partial p_j} u_s\Bigg]\Bigg\}.$$

With this, the derivative of λ_k is given by

$$\begin{aligned}\dot{\lambda}_k = &-\sum_{i=1}^{n_x}\lambda_i \frac{\partial f_i}{\partial x_k} - \sum_{i=1}^{n_x}\sum_{j=1}^{n_p}\sum_{r=1}^{n_x}\Lambda_{x_p}^{ij}\frac{\partial^2 f_i}{\partial x_k \partial x_r} X_{p,rj} \\ &- \sum_{i=1}^{n_x}\sum_{j=1}^{n_p}\Lambda_{x_p}^{ij}\frac{\partial^2 f_i}{\partial x_k \partial p_j} \\ &- \sum_{p=1}^{n_p}\sum_{q=1}^{n_p}\sum_{m=1}^{n_y}\sum_{l=1}^{n_y}(R^{-1})^{ml}\left(\frac{\partial \Phi(\mathcal{F})}{\partial \mathcal{F}}\right)^{pq} \\ &\times\Bigg\{\Bigg[\sum_{r=1}^{n_x}\frac{\partial^2 \bar{h}_m}{\partial x_k \partial x_r} X_{p,rp} + \frac{\partial^2 \bar{h}_m}{\partial x_k \partial p_p} + \sum_{s=1}^{n_u}\frac{\partial^2 k_{s,m}}{\partial x_k \partial p_p} u_s\Bigg] H_{\bar{Y}_p}^{lq} \\ &+ H_{\bar{Y}_p}^{mp}\Bigg[\sum_{r=1}^{n_x}\frac{\partial^2 \bar{h}_l}{\partial x_k \partial x_r} X_{p,rq} + \frac{\partial^2 \bar{h}_l}{\partial x_k \partial p_q} + \sum_{s=1}^{n_u}\frac{\partial^2 k_{s,l}}{\partial x_k \partial p_q} u_s\Bigg]\Bigg\}.\end{aligned}$$

C. HAMILTONIAN CONTROL GRADIENT

The control gradient is given by

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{u}} = \frac{\partial}{\partial \boldsymbol{u}}\Bigg[\lambda^{\mathrm{T}}\left(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{p}) + \sum_{i=1}^{n_u}\boldsymbol{g}_i(\boldsymbol{p}) u_i\right) + \operatorname{tr}\left(\Lambda_{X_p}^{\mathrm{T}}\left[F_{X_p}(X_p,\boldsymbol{x},\boldsymbol{p},\boldsymbol{u})\right]\right) + \operatorname{tr}\left(\Lambda_{\mathcal{F}}^{\mathrm{T}}\left[H_{\bar{Y}_p}^{\mathrm{T}}(X_p,\boldsymbol{x},\boldsymbol{p},\boldsymbol{u}) R^{-1} H_{\bar{Y}_p}(X_p,\boldsymbol{x},\boldsymbol{p},\boldsymbol{u})\right]\right)\Bigg].$$

Componentwise this can be written for the r-th control input u_r from the different input–dependent parts. First

$$\frac{\partial}{\partial u_r}\left(\lambda^{\mathrm{T}}\sum_{s=1}^{n_u} g_s(p) u_s\right) = \lambda^{\mathrm{T}} g_r(p).$$

Second

$$\frac{\partial}{\partial u_r}\operatorname{tr}(\Lambda_{x_p}^{\mathrm{T}} F_{x_p}) = \sum_{i=1}^{n_x}\sum_{j=1}^{n_p} \Lambda_{x_p}^{ij} \frac{\partial}{\partial u_r} F_{x_p}^{ij}, \\ = \sum_{i=1}^{n_x}\sum_{j=1}^{n_p} \Lambda_{x_p}^{ij} \frac{\partial (g_r)_i}{\partial p_j}(\boldsymbol{p}).$$

The third contribution comes from the quadratic form in $H_{\bar{Y}_p}$, using the product rule for the derivative of a quadratic form

$$\frac{\partial}{\partial u_r}\operatorname{tr}(\Lambda_{\mathcal{F}}^{\mathrm{T}} H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} H_{\bar{Y}_p}) = \operatorname{tr}\Big(\Lambda_{\mathcal{F}}^{\mathrm{T}}\big[(\partial_{u_r} H_{\bar{Y}_p})^{\mathrm{T}} R^{-1} H_{\bar{Y}_p} \\ + H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} (\partial_{u_r} H_{\bar{Y}_p})\big]\Big),$$

where the component-wise derivative in the matrix is

$$\left(\frac{\partial}{\partial u_r} H_{\bar{Y}_p}\right)^{ab} = \frac{\partial k_{r,a}}{\partial p_b}(\boldsymbol{x}, \boldsymbol{p}).$$

This gives

$$\frac{\partial}{\partial u_r}\operatorname{tr}(\Lambda_{\mathcal{F}}^{\mathrm{T}} H_{\bar{Y}_p}^{\mathrm{T}} R^{-1} H_{\bar{Y}_p}) = \sum_{p=1}^{n_p}\sum_{q=1}^{n_p}\sum_{a=1}^{n_y}\sum_{b=1}^{n_y} \Lambda_{\mathcal{F}}^{pq} (R^{-1})^{ab} \\ \times \left[\frac{\partial k_{r,a}}{\partial \boldsymbol{p}_p}(x, p)\, H_{\bar{Y}_p}^{bq} + H_{\bar{Y}_p}^{ap} \frac{\partial k_{r,b}}{\partial \boldsymbol{p}_q}(\boldsymbol{x}, \boldsymbol{p})\right].$$

As $\Lambda_{\mathcal{F}}$ is symmetric, the two terms can be combined into a symmetric factor, giving

$$\frac{\partial \mathcal{H}}{\partial u_r} = \lambda^{\mathrm{T}}(g_r(\boldsymbol{p})) + \sum_{i=1}^{n_x}\sum_{j=1}^{n_p} \Lambda_{x_p}^{ij} \frac{\partial (g_r)_i}{\partial p_j}(\boldsymbol{p}) \\ + 2\sum_{p=1}^{n_p}\sum_{a=1}^{n_y}\left[\sum_{q=1}^{n_p}\sum_{b=1}^{n_y} (R^{-1})^{ab} \Lambda_{\mathcal{F}}^{pq} H_{\bar{Y}_p}^{bq}\right] \frac{\partial k_{r,a}}{\partial \boldsymbol{p}_p}(\boldsymbol{x}, \boldsymbol{p}). \tag{52}$$

With this, the gradient of the Hamiltonian with respect to the control inputs can be computed by stacking the components of (52) for each control input u_r. This leads to the gradient

$$\nabla_{\boldsymbol{u}} \mathcal{J}(t) = \int_0^T \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{u}}\right) \mathrm{d}t = \int_0^T \left[\frac{\partial \mathcal{H}}{\partial u_1} \quad \cdots \quad \frac{\partial \mathcal{H}}{\partial u_{n_u}}\right]^{\mathrm{T}} \mathrm{d}t.$$

REFERENCES

[1] E. Walter and L. Pronzato, "Qualitative and quantitative experiment design for phenomenological models—A survey," *Automatica*, vol. 26, no. 2, pp. 195–213, Mar. 1990.

[2] L. Ljung and T. Glad, "On global identifiability for arbitrary model parametrizations," *Automatica*, vol. 30, no. 2, pp. 265–276, Feb. 1994.

[3] R. A. Fisher, *The Design of Experiments*. Cham, Switzerland: Springer, 1971.

[4] A. M. Cappuyns, K. Bernaerts, I. Y. Smets, O. Ona, E. Prinsen, J. Vanderleyden, and J. F. Van Impe, "Optimal experiment design in bioprocess modeling: From theory to practice," *IFAC Proc. Volumes*, vol. 39, no. 2, pp. 535–540, 2006.

[5] S. Bhonsale, P. Nimmegeers, S. Akkermans, D. Telen, I. Stamati, F. Logist, and J. F. M. Van Impe, "Optimal experiment design for dynamic processes," in *Simulation and Optimization in Process Engineering*. Amsterdam, The Netherlands: Elsevier, 2022, pp. 243–271.

[6] H. A. Chianeh, J. D. Stigter, and K. J. Keesman, "Optimal input design for parameter estimation in a single and double tank system through direct control of parametric output sensitivities," *J. Process Control*, vol. 21, no. 1, pp. 111–118, Jan. 2011.

[7] K. J. Keesman, "Optimal input design for parameter estimation in nonlinear state-space models using Pontryagin's minimum principle," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 1319–1324, 2015.

[8] M. Z. Babar and M. Baglietto, "MPC based optimal input design for nonlinear system identification," in *Proc. 20th Int. Conf. Syst. Theory, Control Comput. (ICSTCC)*, Oct. 2016, pp. 619–625.

[9] J. Qian, M. Nadri, and P. Dufour, "Optimal input design for parameter estimation of nonlinear systems: Case study of an unstable delta wing," *Int. J. Control*, vol. 90, no. 4, pp. 873–887, Apr. 2017.

[10] S. Park, D. Kato, Z. Gima, R. Klein, and S. Moura, "Optimal input design for parameter identification in an electrochemical Li-ion battery model," in *Proc. Annu. Amer. Control Conf. (ACC)*, Jun. 2018, pp. 2300–2305.

[11] B. Müller, B. Rolle, and O. Sawodny, "Sensitivity analysis and Fisher-information matrix for a dynamic model of a turntable ladder," *IFAC-PapersOnLine*, vol. 55, no. 27, pp. 172–177, 2022.

[12] B. Müller and O. Sawodny, "Dynamic multibody model of a turntable ladder truck considering unloaded outriggers and sensitivity-based parameter identification," *Math. Comput. Model. Dyn. Syst.*, vol. 30, no. 1, pp. 567–590, Dec. 2024.

[13] T. Kailath, *Linear Systems*. London, U.K.: Pearson Education, 1980.

[14] D. Joubert, J. D. Stigter, and J. Molenaar, "Assessing the role of initial conditions in the local structural identifiability of large dynamic models," *Sci. Rep.*, vol. 11, no. 1, p. 16902, Aug. 2021.

[15] J. D. Stigter, L. G. van Willigenburg, and J. Molenaar, "An efficient method to assess local controllability and observability for non-linear systems," *IFAC-PapersOnLine*, vol. 51, no. 2, pp. 535–540, 2018.

[16] J. D. Stigter, D. Joubert, and J. Molenaar, "Observability of complex systems: Finding the gap," *Sci. Rep.*, vol. 7, no. 1, p. 16566, Nov. 2017.

[17] X. Rey Barreiro and A. F. Villaverde, "Benchmarking tools for a priori identifiability analysis," *Bioinformatics*, vol. 39, no. 2, p. 65, Feb. 2023.

[18] A. Othmane, J. Rudolph, and H. Mounier, "Systematic comparison of numerical differentiators and an application to model-free control," *Eur. J. Control*, vol. 62, pp. 113–119, Nov. 2021.

[19] A. Levant, "Robust exact differentiation via sliding mode technique," *Automatica*, vol. 34, no. 3, pp. 379–384, Mar. 1998.

[20] A. Levant, "Higher-order sliding modes, differentiation and output-feedback control," *Int. J. Control*, vol. 76, nos. 9–10, pp. 924–941, Jan. 2003.

[21] J. A. Moreno and M. Osorio, "Strict Lyapunov functions for the super-twisting algorithm," *IEEE Trans. Autom. Control*, vol. 57, no. 4, pp. 1035–1040, Apr. 2012.

[22] R. Seeber and M. Horn, "Stability proof for a well-established super-twisting parameter setting," *Automatica*, vol. 84, pp. 241–243, Oct. 2017.

[23] M. Reichhartinger, S. K. Spurgeon, M. Forstinger, and M. Wipfler, "A robust exact differentiator toolbox for MATLAB/simulink," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 1711–1716, Jul. 2017.

[24] S. Koch and M. Reichhartinger, "Discrete-time equivalents of the super-twisting algorithm," *Automatica*, vol. 107, pp. 190–199, Sep. 2019.

[25] H. Imine, L. Fridman, H. Shraim, and M. Djemai, *Sliding Mode Based Analysis and Identification of Vehicle Dynamics*, vol. 414. Cham, Switzerland: Springer, 2011.

[26] L. Senkel, A. Rauh, and H. Aschemann, "Optimal input design for online state and parameter estimation using interval sliding mode observers," in *Proc. 52nd IEEE Conf. Decis. Control*, Dec. 2013, pp. 502–507.

[27] H. Nijmeijer and A. Van der Schaft, *Nonlinear Dynamical Control Systems*, vol. 175. Cham, Switzerland: Springer, 1990.

[28] A. Isidori, *Nonlinear Control Systems*. London, U.K.: Springer, 1995.

[29] C.-T. Lin, "Structural controllability," *IEEE Trans. Autom. Control*, vol. AC-19, no. 3, pp. 201–208, Mar. 1974.

[30] J.-M. Dion, C. Commault, and J. van der Woude, "Generic properties and control of linear structured systems: A survey," *Automatica*, vol. 39, no. 7, pp. 1125–1144, Jul. 2003.

[31] Y. Liu, J.-J. Slotine, and A. Barabási, "Observability of complex systems," *Proc. Natl. Acad. Sci. USA*, vol. 110, no. 7, pp. 2460–2465, 2013.

[32] J. Lunze, "Networked control of multi-agent systems," in *Bookmundo Direct*, vol. 1, 2019, ch. 2, pp. 29–56.

[33] A. Levant, M. Livne, and X. Yu, "Sliding-mode-based differentiation and its application," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 1699–1704, Jul. 2017.

[34] A. F. Filippov, *Differential Equations With Discontinuous Righthand Sides: Control Systems*, vol. 18. Cham, Switzerland: Springer, 2013.
[35] R. Seeber and M. Horn, "Necessary and sufficient stability criterion for the super-twisting algorithm," in *Proc. 15th Int. Workshop Variable Struct. Syst. (VSS)*, Jul. 2018, pp. 120–125.
[36] R. Seeber, M. Horn, and L. Fridman, "A novel method to estimate the reaching time of the super-twisting algorithm," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4301–4308, Dec. 2018.
[37] A. Gelb, *Applied Optimal Estimation*. Cambridge, MA, USA: MIT Press, 1978.
[38] K. Reif, S. Guünther, E. Yaz, and R. Unbehauen, "Stochastic stability of the continuous-time extended Kalman filter," *IEE Proc.-Control Theory Appl.*, vol. 147, no. 1, pp. 45–52, Jan. 2000.
[39] T. N. Lundt, A. Lepsien, P. Feketa, T. Meurer, and A. Schaum, "Parameter estimation in adaptively coupled Kuramoto oscillators," *IFAC-PapersOnLine*, vol. 59, no. 26, pp. 305–310, 2025.
[40] N. J. Treloar, N. Braniff, B. Ingalls, and C. P. Barnes, "Deep reinforcement learning for optimal experimental design in biology," *PLOS Comput. Biol.*, vol. 18, no. 11, Nov. 2022, Art. no. e1010695.
[41] S. Pant, "Information sensitivity functions to assess parameter information gain and identifiability of dynamical systems," *J. Roy. Soc. Interface*, vol. 15, no. 142, May 2018, Art. no. 20170871.
[42] I. Y. M. Smets, K. J. E. Versyck, and J. F. M. Van Impe, "Optimal control theory: A generic tool for identification and control of (bio-)chemical reactors," *Annu. Rev. Control*, vol. 26, no. 1, pp. 57–73, Jan. 2002.
[43] M. Papageorgiou, M. Leibold, and M. Buss, *Optimierung*, vol. 4. Cham, Switzerland: Springer, 2015.
[44] D. E. Kirk, *Optimal Control Theory: An Introduction*. Mineola, NY, USA: Dover, 2004.
[45] X. Liu and J. Frank, "Symplectic Runge–Kutta discretization of a regularized forward–backward sweep iteration for optimal control problems," *J. Comput. Appl. Math.*, vol. 383, Feb. 2021, Art. no. 113133.
[46] M. McAsey, L. Mou, and W. Han, "Convergence of the forward-backward sweep method in optimal control," *Comput. Optim. Appl.*, vol. 53, no. 1, pp. 207–226, Sep. 2012.
[47] W. W. Hager and R. Rostamian, "Optimal coatings, bang–bang controls, and gradient techniques," *Optim. Control Appl. Methods*, vol. 8, no. 1, pp. 1–20, Jan. 1987.
[48] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. Philadelphia, PA, USA: SIAM, 2008.
[49] J. Park, R. A. Martin, J. D. Kelly, and J. D. Hedengren, "Benchmark temperature microcontroller for process dynamics and control," *Comput. Chem. Eng.*, vol. 135, Apr. 2020, Art. no. 106736.
[50] J. S. Hunter, "The exponentially weighted moving average," *J. Quality Technol.*, vol. 18, no. 4, pp. 203–210, 1986.
[51] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Dept. Appl. Math. Comput. Sci., Tech. Univ. Denmark, Nov. 2012. [Online]. Available: http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html

ARTHUR LEPSIEN received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from Kiel University (CAU), in 2021 and 2023, respectively, where he is currently pursuing the Ph.D. degree with the Chair of Automatic Control.

Then, he switched to the University of Hohenheim, in September 2023. He is with the University of Hohenheim, working on model-based analysis and synthesis of automation solutions for spray drying. Furthermore, he works on model order reduction for FEM based models, focusing on observer and control design for thermal and viscoelastic processes.

PHILIPP KÜGLER received the Dipl.-Ing. degree in technical physics and industrial mathematics from Johannes Kepler University (JKU) Linz, in 2000, and the Dr.Techn. degree in technical sciences, in 2003.

After visiting positions at the University of Oxford, in 2001, the University of California, Los Angeles, in 2003, and the University of Stuttgart, from 2009 to 2010, he established the research group mathematical methods in molecular and systems biology for the Radon Institute of the Austrian Academy of Sciences, Vienna Biocenter, from 2010 to 2013. Since 2013, he has been a Full Professor (W3) of mathematics, especially for the modeling of complex biological systems, and the Managing Director of the Institute of Applied Mathematics and Statistics, University of Hohenheim. His research interests include computational methods for cardiac drug safety testing and artificial intelligence-based strategies for the evaluation of pharmacology data. He is a Research Member of the Living Heart Project run by Dassault Systémes and serves on various committees of the Safety Pharmacology Society regularly. In 2017, he declined a call for a Full Professorship in mathematical methods for medicine and biosciences at JKU and became a Founding Member of the Computational Science Hub, University of Hohenheim.

ALEXANDER SCHAUM (Senior Member, IEEE) received the Diploma degree in technical cybernetics from the University of Stuttgart, in 2006, the Ph.D. degree in control engineering from Universidad Nacional Autónoma de México (UNAM), Mexico City, in 2009, and the Habilitation degree in control theory and control engineering from Kiel University (CAU), in 2020.

He has led the Chair of Process Analytics, University of Hohenheim, since September 2023. From 2011 to 2014, he was a Guest Professor with the Department for Applied Mathematics and Systems, Universidad Autónoma Metropolitana (UAM), Cuajimalpa, Mexico City. During winter term 2018, he had an Interim Professorship for Networked Electronic Systems and Automation and Control, CAU, in the summer term 2023. His research interests include process modeling for monitoring and control with applications in fermentation and thermal processing, particularly in bio- and food technology, and the development of new methods for observer and control design for nonlinear and distributed parameter systems. He is an IFAC Affiliate and a member of the IEEE Control System Society (CSS), DECHEMA, GDCH, and ARGESIM.