



Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational
Mathematics (IMACM)

Preprint BUW-IMACM 26/20

Tom Zwerschke, Matthias Ehrhardt, Michael Günther, and Claudia Totzeck

Structure-Preserving Data-Driven Identification of Port-Hamiltonian Differential-Algebraic Systems

September 7, 2026

<http://www.imacm.uni-wuppertal.de>

Structure-Preserving Data-Driven Identification of Port-Hamiltonian Differential-Algebraic Systems

Tom Zwerschke^a, Matthias Ehrhardt^{a,*}, Michael Günther^a, Claudia Totzeck^b

^a*University of Wuppertal, Applied and Computational Mathematics,
Gaußstrasse 20, 42119 Wuppertal, Germany*

^b*University of Wuppertal, Optimization Group,
Gaußstrasse 20, 42119 Wuppertal, Germany*

Abstract

We present a data-driven approach to identifying linear index-1 differential-algebraic pH systems (pH-DAEs) based on input-output measurements. In comparison to the identification of port-Hamiltonian (pH) systems, the algebraic constraint and the index condition pose additional challenges. First, we establish a structure-preserving formulation of the considered pH-DAE class and derive an implicit midpoint discretization that preserves the algebraic constraints and discrete dissipation inequality. We formulate the identification problem as a regularized least-squares minimization problem subject to the pH-DAE dynamics. Exploiting the index-1 structure, we reduce the constrained problem to an unconstrained optimization problem over the system parameters while preserving the port-Hamiltonian structure. Next, we derive an adjoint-based formulation to efficiently evaluate the gradient of the resulting reduced cost functional. This enables us to use gradient-based optimization methods for parameter estimation. Under suitable assumptions on the admissible parameter set, the existence of a minimizer is established. Numerical experiments demonstrate that the proposed approach can identify surrogate pH-DAE systems that accurately reproduce the input-output behavior of reference systems. Further investigations show the approach's potential for identifying reduced-order surrogate models. Cross-validation with independent input signals confirms the predictive capability of the identified models.

Keywords: Differential-algebraic equations, port-Hamiltonian systems, system identification, data-driven modelling, structure-preserving methods, adjoint methods

2020 Mathematics Subject Classification: 93B30, 34A09, 65L80, 90C30

1. Introduction

Mathematical models of physical systems are often characterized by conservation laws, dissipation mechanisms, and interconnections between different physical subsystems. Pre-

*Corresponding author

Email addresses: tom.zwerschke@uni-wuppertal.de (Tom Zwerschke),
ehrhhardt@uni-wuppertal.de (Matthias Ehrhardt), guenther@uni-wuppertal.de (Michael Günther),
totzeck@uni-wuppertal.de (Claudia Totzeck)

serving these structural properties in the mathematical representation is important not only for the physical interpretability of the model, but also for the stability and robustness of numerical simulations. *Port-Hamiltonian (pH) systems* provide a systematic framework for structure-preserving modelling based on the balance of energy and power. In particular, the pH formulation separates the power-conserving part of the dynamics from dissipative effects and external interactions. This structure is preserved under power-conserving interconnections and therefore provides a natural framework for the modelling and coupling of complex multiphysics systems; see, e.g., [26, 20].

In many applications, however, a detailed physical model of a subsystem is not available. This may be caused by incomplete knowledge of the underlying physical parameters, inaccessible internal variables, or the use of proprietary simulation software for which only input-output information is available. In such situations, data-driven system identification provides a means of constructing surrogate models from measurements or simulation data. For port-Hamiltonian systems, it is desirable that the identified model retains the pH structure, since otherwise important properties of the original physical system may be lost during the identification process.

Several approaches for data-driven identification of pH systems have been developed in recent years. Frequency-domain approaches based on the Loewner framework can directly construct pH realizations from suitable frequency response data [7]. A related non-intrusive approach for time-domain data first extracts frequency-response information from input-output measurements and subsequently constructs a pH realization using a Loewner-based procedure [9]. These approaches are attractive when frequency-response data are available or can be reliably inferred from the measurements. An alternative strategy is to formulate the identification directly as an optimization problem. In [15], an adjoint-based time-domain calibration method was developed for linear pH ordinary differential equations (ODEs). The system matrices are determined directly from input-output data by minimizing a data-misfit functional subject to the pH dynamics. Importantly, the pH constraints are incorporated directly into the optimization problem, so that the identified systems retain the skew-symmetry and positive-(semi-)definiteness properties required by the pH formulation. Sensitivity-based identification methods have also been considered for linear pH systems [16].

Differential-algebraic equations (DAEs) arise naturally when dynamics have to obey strict structural constraints, for example in mechanical multibody systems such as robotics. Linear pH descriptor systems were systematically developed in [6], where the pH structure was extended to constrained dynamical systems of arbitrary differentiation index. The authors showed how regularization procedures can be formulated while preserving the pH structure. Numerical methods for DAEs with pH applications have been developed, including structure-preserving splitting techniques; see, e.g., [5, 4].

The data-driven identification of pH-DAE systems, however, is considerably less developed than the problem for pH-ODE systems. One recent approach was proposed by Zaspel and Günther [28], where Gaussian processes are employed to identify nonlinear effort functions of pH-DAE systems from input and state data. The approach was demonstrated for index-1 and index-three pH-DAE systems arising in network design and constrained multibody dynamics. While this provides a data-driven framework for pH-DAEs, it requires state

information and follows a fundamentally different, nonparametric modelling strategy.

Here, we consider the complementary problem of identifying *linear index-1 pH-DAE systems directly from input-output measurements*. Our approach extends the adjoint-based time-domain calibration framework for pH-ODE systems in [15] to the differential-algebraic setting. The main difficulty is that the system dynamics are subject to algebraic constraints and that the differential and algebraic variables must remain compatible with the index-1 condition throughout the identification procedure. We therefore develop a structure-preserving numerical treatment of the underlying pH-DAE and exploit its index-1 structure to formulate the identification problem as an optimization problem in the system parameters.

More precisely, we first formulate the considered class of linear pH-DAE systems and establish the structural properties relevant for the identification problem. For index-1 systems, we derive a numerical discretization that respects the algebraic constraints and the dissipative structure. The resulting data-driven identification problem is posed as a least-squares minimization problem subject to the pH-DAE dynamics. By exploiting the index-1 structure, the constrained problem can be reduced to an unconstrained optimization problem over an appropriately parameterized set of pH system matrices. We derive the gradient of the reduced cost functional using an adjoint formulation. This avoids the explicit computation of sensitivities with respect to every individual system parameter and provides an efficient basis for gradient-based optimization. Under suitable assumptions, we additionally establish the existence of a minimizer of the identification problem.

The proposed approach is assessed numerically using several reference systems. The experiments investigate the influence of the number of differential variables used in the identified model and examine whether lower-dimensional pH-DAE surrogate models can reproduce the observed input-output behaviour. In addition, the identified models are evaluated using independent test signals. The numerical results demonstrate that the identified systems can reproduce the input-output behaviour of the reference systems while retaining the port-Hamiltonian structure. In particular, the cross-validation experiments indicate that the identified models are not merely fitted to the training signal but can provide useful surrogate models for unseen inputs.

The paper is organized as follows. Section 2 introduces the port-Hamiltonian framework and recalls the properties needed in the subsequent analysis. Section 3 discusses differential-algebraic systems and introduces the considered class of linear port-Hamiltonian DAEs, with particular emphasis on the index-1 case. The data-driven identification problem and the existence result are formulated in Section 4. Section 5 derives the parameterization and the adjoint-based gradient used for the numerical optimization. Numerical experiments are presented in Section 6, including the model-reduction and cross-validation studies. Finally, Section 7 summarizes the main results and discusses possible directions for future work.

2. Port-Hamiltonian Systems

We consider the Hamiltonian equations for a mechanical system [26, 3]

$$\dot{q} = \frac{\partial H(p, q)}{\partial p}, \quad \dot{p} = -\frac{\partial H(p, q)}{\partial q} + F, \quad (1)$$

with the generalized coordinates $q = (q_1, \dots, q_k)^\top$, the generalized momenta $p = (p_1, \dots, p_k)^\top$ and the vector of generalized forces F . The function $H(p, q)$ is called *separable Hamiltonian* and is given by the *total energy* $H(p, q) = U(p) + V(q)$, i.e., the sum of kinetic energy U and potential energy V . An important property of such systems is the *energy balance*

$$\frac{d}{dt}H = \frac{\partial^\top H}{\partial q}(q, p)\dot{q} + \frac{\partial^\top H}{\partial p}(q, p)\dot{p} = \frac{\partial^\top H}{\partial p}(q, p)F = \dot{q}^\top F. \quad (2)$$

The Hamiltonian equations (1) without external force F can be expressed in matrix form

$$\dot{x} = J\nabla H(x) \quad \text{with} \quad x = (q, p)^\top \quad \text{and} \quad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \quad (3)$$

Next, we replace the matrix J by an arbitrary skew-symmetric matrix that is again called J , and add a positive semi-definite matrix R to include a dissipative term (energy loss)

$$\dot{x} = (J - R)\nabla H(x). \quad (4)$$

Finally, we include an input and an output to couple the system to the environment.

Definition 1 ([21]). Let $H: \mathbb{R}^n \rightarrow \mathbb{R}$ be a twice continuously differentiable function, $J \in \mathbb{R}^{n \times n}$ a skew-symmetric matrix, i.e., $J^\top = -J$, and $R \in \mathbb{R}^{n \times n}$ a symmetric positive semi-definite matrix, i.e., $R \succeq 0$. The system

$$\dot{x}(t) = (J - R)\nabla H(x(t)) + Bu(t), \quad x(0) = x_0, \quad (5a)$$

$$y = B^\top \nabla H(x(t)), \quad (5b)$$

is called a *port-Hamiltonian system* (pH system) where $u: [0, T] \rightarrow \mathbb{R}^m$ and $y: [0, T] \rightarrow \mathbb{R}^m$ denote an input and output signal with $T > 0$ the time horizon. The matrix $B \in \mathbb{R}^{n \times m}$ is called *port matrix*.

In the linear case, the Hamiltonian H is a quadratic function, i.e., $H(x) = \frac{1}{2}x^\top Qx$, where $Q \in \mathbb{R}^{n \times n}$. Since H represents the energy of the system, we require that $H \geq 0$. Therefore, Q is a symmetric positive semi-definite matrix.

A fundamental property of port-Hamiltonian systems is the *dissipation inequality*, i.e. for the Hamiltonian H and the solutions $x(t)$ and $y(t)$ of the system (5) we have

$$\frac{d}{dt}H(x(t)) \leq y(t)^\top u(t), \quad (6)$$

or in integral form

$$H(x(t+h)) - H(x(t)) \leq \int_t^{t+h} y(\tau)^\top u(\tau) d\tau. \quad (7)$$

Finally, the second important property is the *interconnection*, i.e. the proper coupling of pH systems leads again to a monolithic pH system, see [11] for details.

The next step is to design a numerical scheme that preserves the dissipation inequality (6). For simplicity of notation, we consider a uniform grid for discretization: $0 < t_1 < \dots <$

t_{end} with step size $h = t_{n+1} - t_n > 0$ and numerical solutions $x_n \approx x(t_n)$. A class of methods that satisfied the discrete version of the dissipation inequality (6)

$$\frac{H(x_{n+1}) - H(x_n)}{h} \leq y_{n+1}^\top u_{n+1} \quad (8)$$

is the discrete gradient method.

Definition 2 ([14]). The discrete gradient of a continuously differentiable function $H: \mathbb{R}^n \rightarrow \mathbb{R}$ is the mapping $\bar{\nabla}H: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ with the two properties

$$\bar{\nabla}H(y, x)^\top (y - x) = H(y) - H(x) \quad \text{for all } x, y \in \mathbb{R}^n, \quad (9)$$

$$\bar{\nabla}H(x, x) = \nabla H(x) \quad \text{for all } x \in \mathbb{R}^n. \quad (10)$$

Definition 3. Let $\bar{\nabla}H$ be the discrete gradient of the function H defined in Definition 2. Furthermore, let J be a skew-symmetric matrix, R a positive semi-definite matrix. Then

$$x_{n+1} = x_n + h(J - R)\bar{\nabla}H(x_{n+1}, x_n) + hBu_{n+1} \quad (11)$$

defines the *discrete gradient method*. The definition of u_{n+1} depends on the choice of $\bar{\nabla}H$.

For a quadratic Hamiltonian $H(x) = \frac{1}{2}x^\top Qx$ with $Q = Q^\top$, the (*midpoint*) *discrete gradient* [14] reduces to the midpoint evaluation of the gradient,

$$\bar{\nabla}H(x_{n+1}, x_n) = \nabla H\left(\frac{x_n + x_{n+1}}{2}\right) = Q\left(\frac{x_n + x_{n+1}}{2}\right). \quad (12)$$

Consequently, the resulting discrete-gradient scheme coincides with the implicit midpoint rule; see, e.g., [17, Section 3.1].

3. Port-Hamiltonian Differential-Algebraic Systems

In this section we consider the special case of linear port-Hamiltonian DAE systems which represent a generalized form systems of port-Hamiltonian ODEs. First, we introduce port-Hamiltonian DAE systems and subsequently discuss the properties of such systems.

Definition 4 (pH-DAE,[4]). A linear constant coefficient DAE system of the form

$$E\dot{x} = (J - R)Qx + Bu, \quad (13a)$$

$$y = B^\top Qx, \quad (13b)$$

with $E, Q, J, R \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, and $x = x(t) \in \mathbb{R}^n$, $u = u(t)$, $y = y(t) \in \mathbb{R}^m$ with $t \in [0, T]$, is called a *port-Hamiltonian differential-algebraic system* (pH-DAE), if

1. the differential-algebraic operator

$$Q^\top E \frac{d}{dt} - Q^\top JQ: \mathcal{X} \subset \mathcal{C}^1([0, T], \mathbb{R}^n) \rightarrow \mathcal{C}^0([0, T], \mathbb{R}^n)$$

is skew-adjoint, i.e., $Q^\top J^\top Q = -Q^\top JQ$ and $Q^\top E = E^\top Q$ hold,

2. the product $Q^\top E$ is positive semi-definite, i.e., $Q^\top E \succeq 0$,
3. the passivity matrix $W = (Q^\top RQ)$ is symmetric positive semi-definite, $W \succeq 0$.

The underlying quadratic Hamiltonian $\mathcal{H}: \mathbb{R}^n \rightarrow \mathbb{R}$ of the system is $\mathcal{H}(x) = \frac{1}{2}x^\top Q^\top E x$.

In this work we focus on the identification of pH-DAE systems in the form

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \left(\begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} - \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \right) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} u. \quad (14)$$

We note that, provided Q is regular, any linear pH-DAE system can be transformed into a system (14), which is already given in semi-explicit form. Componentwise we obtain

$$\dot{x}_1 = (J_{11} - R_{11})x_1 + (J_{12} - R_{12})x_2 + B_1 u, \quad (15)$$

$$0 = (J_{21} - R_{21})x_1 + (J_{22} - R_{22})x_2 + B_2 u. \quad (16)$$

The first r components of vector x correspond to the differential variable x_1 , and the remaining $n-r$ components correspond to the algebraic variable x_2 . Under the index 1 assumption that the matrix $(J_{22} - R_{22})$ is regular, the algebraic equation can be solved for x_2 :

$$x_2 = -(J_{22} - R_{22})^{-1}((J_{21} - R_{21})x_1 + B_2 u). \quad (17)$$

If we substitute the algebraic part (17) into the differential part (15), we obtain

$$\begin{aligned} \dot{x}_1 = & \left(J_{11} - R_{11} - (J_{12} - R_{12})(J_{22} - R_{22})^{-1}(J_{21} - R_{21}) \right) x_1 \\ & + \left(B_1 - (J_{12} - R_{12})(J_{22} - R_{22})^{-1}B_2 \right) u, \end{aligned} \quad (18)$$

which is a closed-form expression for x_1 . Hence, we obtain a simple ODE system for x_1 (18) and an explicit equation for x_2 (17). The initial condition for $x_1(0) = x_{1,0}$ can be chosen arbitrarily, while the initial condition for the algebraic variable x_2 is given by

$$x_{2,0} := x_2(0) = -(J_{22} - R_{22})^{-1}((J_{21} - R_{21})x_1(0) + B_2 u(0)). \quad (19)$$

Thus, instead of writing $x(0) = x_0 \in \mathbb{R}^n$ as a consistent initial value, we can write $x_1(0) = x_{1,0} \in \mathbb{R}^r$ and determine $x_2(0) = x_{2,0} \in \mathbb{R}^{n-r}$ using (19).

Remark 1. In [6], it has been shown that the index of the pH-DAE (13) is at most two. Furthermore, it was shown that, under certain assumptions an index 2 pH-DAE can be transformed into an equivalent index 1 pH-DAE using a structure preserving regularization.

In the sequel we will consider again the implicit midpoint and derive the method directly for pH-DAE systems by approximating the derivative $\dot{x}(t_{n+1/2})$ with the finite difference $\frac{x_{n+1} - x_n}{h}$ and $x(t_{n+1/2})$ with the midpoint $\frac{x_n + x_{n+1}}{2}$. This yields

$$E \frac{x_{n+1} - x_n}{h} = (J - R) \frac{x_n + x_{n+1}}{2} + B \frac{u_n + u_{n+1}}{2}. \quad (20)$$

In the sequel, we restrict ourselves to the transformed pH-DAE systems of form (14), i.e.

$$E \dot{x} = (J - R)x + Bu, \quad \text{with} \quad E = \text{diag}(I_r, 0), \quad (21a)$$

$$y = B^\top x. \quad (21b)$$

Definition 5. The implicit midpoint rule for solving the system (21) is given by

$$\begin{aligned} \left(E - \frac{h}{2}(J - R)\right)x_{n+1} &= \left(E + \frac{h}{2}(J - R)\right)x_n + \frac{h}{2}B(u_n + u_{n+1}), \\ x_{n+1} &= \left(E - \frac{h}{2}(J - R)\right)^{-1} \left[\left(E + \frac{h}{2}(J - R)\right)x_n + \frac{h}{2}B(u_n + u_{n+1}) \right]. \end{aligned} \quad (22)$$

It was shown in [4] that the rule (22) is well-defined for index-1 systems and satisfies the algebraic constraints and the discrete form of the dissipation inequality, i.e. the midpoint rule is structure-preserving for pH-DAE systems under certain conditions.

4. Identification Problem

In the previous section we introduced pH-DAE systems. Now, we will investigate how we can reconstruct such systems from measured data $(t_i, u_i, y_i)_{i=0, \dots, N}$. First, we will formulate the minimization problem. We assume that we have an input signal $u: [0, T] \rightarrow \mathbb{R}^m$ and the corresponding output signal $y_{\text{data}}: [0, T] \rightarrow \mathbb{R}^n$ from an unknown system. If only discrete values $(t_i, u_i, y_i)_{i=0, \dots, N}$ are given, the functions $u(t)$ and $y(t)$ are obtained by interpolation.



Figure 1: A linear pH-DAE system to be determined.

The main assumption in the identification process is that the data are generated by a linear index-1 pH-DAE system. We consider the calibration problem with *cost functional*

$$\mathcal{J}(y, w) = \frac{1}{2} \int_0^T |y(t) - y_{\text{data}}(t)|^2 dt + \frac{\lambda}{2} |w - w_{\text{ref}}|^2, \quad \lambda \geq 0, \quad (23)$$

subject to the state constraint or *state equation* (cf. (21)):

$$E\dot{x} = (J - R)x + Bu, \quad x_1(0) = x_{1,0} \in \mathbb{R}^r, \quad (24a)$$

$$y = B^\top x, \quad (24b)$$

where $w = (J, R, B, x_{1,0})$ contains the matrices and the initial data to be identified by minimizing $\mathcal{J}$. If applicable, w_{ref} contains reference data. Otherwise, we set $\lambda = 0$.

Since we assume that we are investigating a pH-DAE, the following properties hold:

$$J, R \in \mathbb{R}^{n \times n}, \quad J^\top = -J, \quad R \succeq 0, \quad B \in \mathbb{R}^{n \times m}, \quad (J_{22} - R_{22}) \text{ is regular.} \quad (25)$$

However, the proof of continuous dependence on the data shows that the index-1 condition on $J_{22} - R_{22}$ to be regular, alone is insufficient. Therefore, we additionally assume that the norm of the inverse matrix is bounded, i.e., there exists $C \gg 0$ such that it holds $\|(J_{22} - R_{22})^{-1}\| \leq C$. Note that this is equivalent to the smallest singular value being

greater than or equal to some threshold $\epsilon > 0$, i.e., $\sigma_{\min}(J_{22} - R_{22}) \geq \epsilon > 0$, cf. [13]. In other words, the matrix remains sufficiently far away from a singularity. Therefore, we define the *admissible parameter set* as

$$\mathcal{U}_{\text{ad}} := \left\{ w = (J, R, B, x_{1,0}) \in W : J^\top = -J, \quad R \succeq 0, \quad \sigma_{\min}(J_{22} - R_{22}) \geq \epsilon > 0 \right\}, \quad (26)$$

where $W = \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times m} \times \mathbb{R}^r$ denotes the parameter space. Since the smallest singular value depends continuously on the matrix entries, $\mathcal{U}_{\text{ad}}$ is closed. Moreover, the condition $\sigma_{\min}(J_{22} - R_{22}) \geq \epsilon$ ensures the uniform bound

$$\|(J_{22} - R_{22})^{-1}\|_2 \leq \epsilon^{-1}.$$

We note that the solution to the identification process is also in $\mathcal{U}_{\text{ad}}$, since we restrict ourselves to index-1 pH-DAEs. Moreover, we define the output space $Y = H^1((0, T), \mathbb{R}^m) = \{y \in L^2((0, T), \mathbb{R}^m) : \dot{y} \in L^2((0, T), \mathbb{R}^m)\}$ equipped with the norm

$$\|y\|_{H^1((0, T), \mathbb{R}^m)} = \left(\|y\|_{L^2((0, T), \mathbb{R}^m)}^2 + \|\dot{y}\|_{L^2((0, T), \mathbb{R}^m)}^2 \right)^{1/2}. \quad (27)$$

This choice of the space can be justified by the fact that y and x have the same structure, since $y = B^\top x$, and x must be differentiable, i.e., $x \in H^1((0, T), \mathbb{R}^n)$. In addition, we are introducing the following notation

$$\mathcal{J}_1(y) := \frac{1}{2} \int_0^T |y(t) - y_{\text{data}}(t)|^2 dt, \quad \mathcal{J}_2(w) := \frac{\lambda}{2} |w - w_{\text{ref}}|^2.$$

If no information on reference data are available, we recommend to set $\lambda = 0$. In this case, the cost function reduces to $\mathcal{J}_1(y)$.

Now we have all important ingredients in order to formulate our calibration task.

Problem 1. Find system matrices and the initial data, $w = (J, R, B, x_{1,0})$, which solve

$$\min_{(y, w) \in Y \times \mathcal{U}_{\text{ad}}} \mathcal{J}(y, w) \quad \text{subject to} \quad (24). \quad (\text{IdP})$$

This minimization problem (IdP) is solved using an iterative method and will be discussed in Section 5. The idea is to start from an initial guess w^0 and solve the state equation to obtain state variable x and output y . Based on this, we evaluate $\mathcal{J}$ and calculate the gradient which is then used for a gradient descent to minimize $\mathcal{J}$. The aim now is to prove that the *identification problem* (IdP) is well-posed and solvable under certain conditions. To this end, we first make some observations and then prove that the solution to the state equation depends continuously on the data.

Remark 2. The identification problem (IdP) still has two free parameters. While m is determined by the dimension of y_{data} , the dimension of the internal state n and the size of the algebraic part $n - r$ are unknown. We therefore assume that n has already been specified. For a deliberately small choice of n , this can be interpreted as a model order reduction for the internal state. To determine the number of differential variables, it is necessary to solve the minimization problem for various r . Then, on the one hand, we can compare the final costs values, and on the other hand, examine the behavior in response to a second test input u_{test} (cross-validation).

Lemma 2. For every $w \in \mathcal{U}_{\text{ad}}$ and $u \in C^1([0, T], \mathbb{R}^m)$, the state equation (24) admits a unique solution $x \in C^1([0, T], \mathbb{R}^n)$. Moreover, the solution depends continuously on the data. More precisely, for $w, \bar{w} \in \mathcal{U}_{\text{ad}}$ and the corresponding solutions $x, \bar{x}$, there exists a constant $C > 0$ such that

$$\|x - \bar{x}\|_{H^1((0, T), \mathbb{R}^n)} \leq C \|w - \bar{w}\|_{\mathcal{U}_{\text{ad}}}.$$

Proof. For given $w \in \mathcal{U}_{\text{ad}}$ and $u \in C([0, T], \mathbb{R}^m)$ the existence of a unique solution to (24) can be easily shown. For the second statement we estimate $\|x - \bar{x}\|_{L_2((0, T), \mathbb{R}^n)}$ and $\left\|\frac{d}{dt}x - \frac{d}{dt}\bar{x}\right\|_{L_2((0, T), \mathbb{R}^n)}$ separately.

We consider x again in terms of its differential and algebraic parts and obtain

$$|x(t) - \bar{x}(t)|^2 = |x_1(t) - \bar{x}_1(t)|^2 + |x_2(t) - \bar{x}_2(t)|^2, \quad (28)$$

with

$$\begin{aligned} \dot{x}_1 &= (J_{11} - R_{11} - (J_{12} - R_{12})(J_{22} - R_{22})^{-1}(J_{21} - R_{21}))x_1 \\ &\quad + (B_1 - (J_{12} - R_{12})(J_{22} - R_{22})^{-1}B_2)u \\ &=: Mx_1 + Nu, \\ x_2 &= -(J_{22} - R_{22})^{-1}(J_{21} - R_{21})x_1 - (J_{22} - R_{22})^{-1}B_2u \\ &=: Kx_1 + Lu, \end{aligned}$$

according to equations (18) and (17). We define $\bar{M}, \bar{N}, \bar{K}$ and $\bar{L}$ analogous for the solution corresponding to $\bar{w}$. The integral representation of the ODE then yields

$$|x_1(t) - \bar{x}_1(t)|^2 \leq 2|x_{1,0} - \bar{x}_{1,0}|^2 + 2T \int_0^t |Mx_1(s) - \bar{M}\bar{x}_1(s) + (N - \bar{N})u(s)|^2 ds.$$

To estimate the integral, we add the term $\bar{M}x_1 - \bar{M}\bar{x}_1$ to the integrand and obtain

$$\begin{aligned} &|Mx_1 - \bar{M}\bar{x}_1 - \bar{M}x_1 + \bar{M}\bar{x}_1 + (N - \bar{N})u|^2 \\ &= |(M - \bar{M})x_1 + \bar{M}(x_1 - \bar{x}_1) + (N - \bar{N})u|^2 \\ &\leq 3\|M - \bar{M}\|^2|x_1|^2 + 3\|\bar{M}\|^2|x_1 - \bar{x}_1|^2 + 3\|N - \bar{N}\|^2|u|^2. \end{aligned}$$

Inserting into the integral yields

$$\begin{aligned} |x_1(t) - \bar{x}_1(t)|^2 &\leq 2|x_{1,0} - \bar{x}_{1,0}|^2 + 6T \int_0^t \|M - \bar{M}\|^2|x_1|^2 ds + 6T \int_0^t \|\bar{M}\|^2|x_1 - \bar{x}_1|^2 ds \\ &\quad + 6T \int_0^t \|N - \bar{N}\|^2|u|^2 ds \\ &\leq C_2|w - \bar{w}|^2 + \int_0^t C_3|x_1 - \bar{x}_1|^2 ds. \end{aligned}$$

An application of Gronwall inequality yields

$$|x_1(t) - \bar{x}_1(t)|^2 \leq C_4|w - \bar{w}|^2. \quad (29)$$

For the algebraic part x_2 , we obtain the following estimate

$$\begin{aligned} |x_2(t) - \bar{x}_2(t)|^2 &= |(K - \bar{K})x_1 + \bar{K}(x_1 - \bar{x}_1) + (L - \bar{L})u|^2 \\ &\leq 3\|K - \bar{K}\|^2|x_1|^2 + 3\|\bar{K}\|^2|x_1 - \bar{x}_1|^2 + 3\|L - \bar{L}\|^2|u|^2 \\ &\leq C_5|w - \bar{w}|^2. \end{aligned}$$

In the final step, we used the estimate (29) to estimate the norm of $K - \bar{K}$ and $L - \bar{L}$.

By an integration over $[0, T]$ and using equation (2), we obtain

$$\|x - \bar{x}\|_{L^2((0,T),\mathbb{R}^n)} \leq C_6|w - \bar{w}|. \quad (30)$$

The second part of the proof consists analogously of estimate the L^2 -difference of the derivatives. For the differential part, we obtain

$$|\frac{d}{dt}(x_1(t) - \bar{x}_1(t))|^2 = C_7|w - \bar{w}|^2 + \int_0^t C_8|\frac{d}{ds}(x_1 - \bar{x}_1)|^2 ds.$$

For the algebraic part, we obtain

$$\begin{aligned} |\frac{d}{dt}(x_2(t) - \bar{x}_2(t))|^2 &\leq 3\|K - \bar{K}\|^2|\frac{d}{dt}x_1|^2 + 3\|\bar{K}\|^2|\frac{d}{dt}|x_1 - \bar{x}_1|^2 + 3\|L - \bar{L}\|^2|\frac{d}{dt}u(t)|^2 \\ &\leq C_9|w - \bar{w}|^2. \end{aligned}$$

Note C_9 depends on $\max_{t \in [0,T]} |\frac{d}{dt}u(t)|^2$, which is finite, since $u \in C^1([0, T], \mathbb{R}^m)$. Again, Gronwall lemma and integration of both terms yields

$$\left\| \frac{d}{dt}(x - x') \right\|_{L^2((0,T),\mathbb{R}^n)} \leq C_{10}|w - w'|^2. \quad (31)$$

Adding (30) and (31) and taking square root yields the desired result. $\square$

Thus, the state equation (24) is well-defined, and we can introduce the *solution operator*.

Definition 6. The *solution operator* S is defined as

$$S: \mathcal{U}_{\text{ad}} \rightarrow H^1((0, T), \mathbb{R}^n), \quad S(w) = x.$$

Using this solution operator $S(w)$, we define the *reduced cost functional*

$$\hat{\mathcal{J}}(w) = \frac{1}{2} \int_0^T |B^\top S(w)(t) - y_{\text{data}}(t)|^2 dt + \frac{\lambda}{2} |w - w_{\text{ref}}|^2. \quad (32)$$

Thus, the minimization Problem 1 reduces to

$$\min \hat{\mathcal{J}}(w), \quad w \in \mathcal{U}_{\text{ad}},$$

and the DAE constraint (24) is only handled implicitly. More specifically, to evaluate the cost functional $\hat{\mathcal{J}}$, the state constraint is first solved, followed by the calculation of the gradient with y . The gradient-descent method used for the numerical results is based on the reduced problem formulation involving $\hat{\mathcal{J}}$.

Next, we define the *state operator* and then show that it is weakly continuous.

Definition 7. The mapping

$$e: H^1((0, T), \mathbb{R}^n) \times \mathcal{U}_{\text{ad}} \rightarrow L^2((0, T), \mathbb{R}^n) \times \mathbb{R}^r, \quad (33)$$

$$(x, w) \mapsto \begin{pmatrix} E \frac{d}{dt} x - (J - R)x - Bu \\ x_1(0) - x_{1,0} \end{pmatrix} \quad (34)$$

with $w = (J, R, B, x_{1,0})$ and $E = \text{diag}(I_r, 0)$ defines the state operator.

Note that the state operator allows us to rewrite our optimization problem as

$$\min_{x \in H^1((0, T), \mathbb{R}^n), w \in \mathcal{U}_{\text{ad}}} J(x, w) \quad \text{subject to} \quad e(x, w) = 0.$$

This formulation of our identification problem will be used later for the Lagrange approach.

Lemma 3. *The state operator e , defined in (33), is weakly continuous.*

Proof. We consider $\{x_n\}_n \subset H^1((0, T), \mathbb{R}^n)$ with $x_n \rightharpoonup x$ and $\{w_n\}_n \subset \mathcal{U}_{\text{ad}}$ with $w_n \rightharpoonup w$. Note that $\mathcal{U}_{\text{ad}}$ is finite dimensional, thus weak and strong convergence coincide. Then for $\varphi \in L^2((0, T), \mathbb{R}^n)$ we obtain

$$\begin{aligned} & \int_0^T \left(E \frac{d}{dt} x_n - (J_n - R_n)x_n(t) - B_n u(t) \right) \cdot \varphi(t) dt \\ & - \int_0^T \left(E \frac{d}{dt} x - (J - R)x(t) - Bu(t) \right) \cdot \varphi(t) dt \\ & = \int_0^T \left(\left(E \frac{d}{dt} x_n - E \frac{d}{dt} x \right) + (B - B_n)u(t) \right) \cdot \varphi(t) dt \\ & + \int_0^T \left[Jx(t) - J_n x_n(t) + R_n x_n(t) - Rx(t) + \underbrace{Jx_n(t) - Jx_n(t) + Rx_n(t) - Rx_n(t)}_{=0} \right] \cdot \varphi(t) dt \\ & = \int_0^T \left(\left(E \frac{d}{dt} x_n - E \frac{d}{dt} x \right) + (B - B_n)u(t) \right) \cdot \varphi(t) dt \\ & + \int_0^T \left[((J - J_n) - (R - R_n))x_n(t) \cdot \varphi(t) + (J - R)(x(t) - x_n(t)) \cdot \varphi(t) \right] dt. \end{aligned}$$

By the weak convergence of x_n and w_n we conclude that all terms tend to zero as $n \rightarrow \infty$. Note that weak convergence of $\{x_n\}_n$ implies boundedness of the sequence $\{x_n\}_n$. Moreover, by the embedding $H^1((0, T), \mathbb{R}^n) \hookrightarrow C([0, T], \mathbb{R}^n)$ we conclude that $x_n(0) \rightarrow x(0)$ and further we obtain $(x_{1,0})_n \rightarrow x_{1,0}$ by the convergence of w_n , which completes the proof. $\square$

Theorem 4. *Let $\lambda > 0$ or $\mathcal{U}_{\text{ad}}$ bounded. Then, the problem (1) admits a global minimum.*

Proof. With the continuous dependence on the data shown in Lemma 2 and the weak continuity of the state operator discussed in Lemma 3, the statement follows as in [15]. $\square$

We note that the global minimum does not have to be unique since our problem is generally non-convex. The non-convexity stems from the nonlinearity introduced by the reformulation using the inverse of $(J_{22} - R_{22})$. In the following section, we will discuss how to find a global minimum using a gradient-based algorithm.

5. Numerical Optimization

Here, we show how to determine an optimal parameter vector w_{opt} that minimizes the cost functional locally, by means of an iterative, gradient-based algorithm. The main challenge is to remain within the admissible set $\mathcal{U}_{\text{ad}}$ at each step. Below, we discuss two approaches: a barrier term in the cost functional and a heuristic approach that exploits the structure of J_{22} and R_{22} . In the following we assume that the dimension of the internal state n and the number of differential variables r is known. If this is not the case, it may be necessary to solve several optimization problems with different hyperparameters n and r . We will examine this in the numerical examples. Let us recall that the minimization problem reads

$$\min \hat{\mathcal{J}}(w), \quad w \in \mathcal{U}_{\text{ad}},$$

with

$$\mathcal{U}_{\text{ad}} := \left\{ w = (J, R, B, x_{1,0}) \in W : \quad J^\top = -J, \quad R \succeq 0, \quad \sigma_{\min}(J_{22} - R_{22}) \geq \epsilon > 0 \right\}.$$

Neglecting the regularity condition, the set $\mathcal{U}_{\text{ad}}$ describes a type of matrix manifold. Consequently, optimization techniques on matrix manifolds can be applied, as was done in [1]. Since we have to deal with the regularity constraint, which couples the conditions on J and R , we choose to follow a parametrization approach [10] in the following.

That is, instead of restricting the optimization to skew-symmetric matrices J , we optimize over all matrices $K \in \mathbb{R}^{n \times n}$ and define $J = K - K^\top$. Similarly, the symmetric positive semi-definite matrix R can be parameterized by $LL^\top$ (where $L \in \mathbb{R}^{n \times n}$), since $x^\top(LL^\top)x = \|L^\top x\|^2 \geq 0$. One disadvantage here is that the linear structure is lost and there are more degrees of freedom. To avoid increasing the degree of freedom, we could restrict L to lower triangular matrices and K to strictly lower triangular matrices. However, for simplicity, we will use the parametrization introduced above and consider

$$\min \mathcal{J}(\bar{w}), \quad \bar{w} = (K, L, B, x_{1,0}) \in W := \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times m} \times \mathbb{R}^r,$$

subject to

$$\begin{aligned} E\dot{x} &= (K - K^\top - LL^\top)x + Bu, & x_1(0) &= x_{1,0} \in \mathbb{R}^r, \\ y &= B^\top x. \end{aligned}$$

In the numerical implementation, we concatenate the matrices L, K and B as well as the initial condition $x_{1,0}$ and work with $v \in \mathbb{R}^z$ where $z = 2n + mn + r$ instead.

The main challenge is to deal with the constraint that $\sigma_{\min}(J_{22} - R_{22}) \geq \epsilon$ or $J_{22} - R_{22}$ is regular. Below we will discuss two possible approaches. The first one is based on a barrier term, which is added to the cost functional to prevent the iterates v^k from approaching the singularities, see [22]. In our setting, a possible penalty term is given by

$$-\mu \log \det(A^\top A), \quad \text{with } A = (J_{22} - R_{22}) \quad \text{and} \quad \mu > 0. \quad (35)$$

We can rewrite the term in (35) as

$$-\mu \log \det(A^\top A) = -\mu \log(\det(A)^2) = -2\mu \log(|\det(A)|) = -2\mu \log(|\det(J_{22} - R_{22})|).$$

Specifically, as $J_{22} - R_{22}$ approaches a singularity, we have $|\det(J_{22} - R_{22})| \rightarrow 0^+$. Consequently, the penalty term tends to infinity, which prevents the iterates from reaching singular matrices. Furthermore, this method is consistent with our original condition, since

$$\log \det(A^\top A) = \log \prod_{i=1}^k \sigma_i^2 = \sum_{i=1}^k \log \sigma_i^2.$$

However, this approach has two disadvantages. First, a sequence of $\{\mu_i\}$ converging to zero must be chosen to solve the original problem

$$\min \mathcal{J}_1(y) + \mathcal{J}_2(w) + \left(-\mu_i \log(|\det(J_{22} - R_{22})|) \right) \quad \text{with } \mu_i \rightarrow 0 \quad \text{for } i = 1, 2, \dots$$

On the other hand, the gradient of the term $-\log |\det(A)|$ is given by $A^{-\top}$. Consequently, the full inverse $(J_{22} - R_{22})^{-1}$ must be computed at every evaluation of $\hat{\mathcal{J}}$, which can be very expensive. Specifically, solving the matrix equation $AX = I_{n-r}$ is required to obtain $A^{-1} := X \in \mathbb{R}^{(n-r) \times (n-r)}$.

The second idea for avoiding singularities is a heuristic approach. Since the matrix J_{22} is skew-symmetric and R_{22} is positive semi-definite, we have $x^\top J_{22}x - x^\top R_{22}x = -x^\top R_{22}x \leq 0$ for all $x \in \mathbb{R}^{n-r}$. Consequently, $J_{22} - R_{22}$ can only be singular if there exists a nonzero vector x such that $J_{22}x = 0$ and $R_{22}x = 0$ and therefore $\ker(J_{22}) \cap \ker(R_{22})$ is nontrivial. Since the solution to our problem is computed iteratively using a gradient-based method, it is difficult to keep track of this condition. However, suppose that v^k is an admissible iterate, meaning that $J_{22} - R_{22}$ is regular. We then compute a new iterate v^{k+1} using a search direction s^k and a step size σ_k , which indicates how far we move in that direction, and set $v^{k+1} := v^k + \sigma_k s^k$. Note that the cost functional can only be evaluated if the state equation is solvable. One way to avoid singularities is therefore to return an error message when the state equation cannot be solved and reject the corresponding step size.

Remark 3. We report that $v^k \in \ker(J_{22}) \cap \ker(R_{22})$ rarely occurred in the numerical examples.

5.1. The Gradient Descent Method

Since we reformulated our identification problem as an unconstrained optimization problem, we aim to apply standard techniques from unconstrained nonlinear optimization (see e.g. [19, 22]) to the reduced cost functional $\hat{\mathcal{J}}(w)$. Since its definition involves the state solution operator, we determine $\nabla \hat{\mathcal{J}}$ with the help of an adjoint-based approach. In [16], this approach was used for to solve the calibration task for pH-ODEs.

We follow the lines of [15, 19] and define the Lagrange function $L: H^1((0, T), \mathbb{R}^n) \times \mathcal{U}_{\text{ad}} \times Z^* \rightarrow \mathbb{R}$ given by

$$L(x, w, \mathbf{p}) = \mathcal{J}(x, w) - \langle \mathbf{p}, e(x, w) \rangle_{Z^*, Z},$$

where $\mathbf{p} \in Z^*$ denotes the Lagrange multiplier or adjoint variable, and $e(x, w)$ is the state operator (33). Note that $Z = L^2((0, T), \mathbb{R}^n) \times \mathbb{R}^r$ and the dual space is given by $Z^* = L^2((0, T), \mathbb{R}^n) \times \mathbb{R}^r$. Thus, the dual pairing can be written as follows:

$$\langle \mathbf{p}, e(x, w) \rangle_{Z^*, Z} = \langle p, \frac{d}{dt}x - (J - R)x - Bu \rangle_{L^2((0, T), \mathbb{R}^n)} + \langle p_{1,0}, x_1(0) - x_{1,0} \rangle_{\mathbb{R}^r}.$$

For simplicity, we use the notation $\mathbf{p} = (p, p_{1,0}) \in Z^*$. The first-order optimality system is derived by solving $dL = 0$. For $h_x \in H^1((0, T), \mathbb{R}^n)$, we compute the directional derivative

$$\begin{aligned} d_x L(x, w, \mathbf{p})[h_x] &= d_x \mathcal{J}(x, w)[h_x] - d_x \langle e(x, w), \mathbf{p} \rangle \\ &= \frac{1}{2} \int_0^T d_x |B^\top x - y_{\text{data}}|^2[h_x] dt \\ &\quad - \int_0^T d_x \left(\left(E \frac{d}{dt}x - (J - R)x - Bu \right) \cdot p \right)[h_x] dt - d_x \left((x_1(0) - x_{1,0}) \cdot p_{1,0} \right)[h_x] \\ &= \int_0^T B(y - y_{\text{data}}) \cdot h_x + (J - R)h_x \cdot p dt - \int_0^T \left(E \frac{d}{dt}h_x \right) \cdot p dt - h_{1,x}(0) \cdot p_{1,0} \end{aligned}$$

Next, we use integration by parts for the second integral and obtain

$$\begin{aligned} &\int_0^T B(y - y_{\text{data}}) \cdot h_x + (J - R)h_x \cdot p dt - \left[E h_x \cdot p \right]_0^\top + \int_0^T E h_x \cdot \frac{d}{dt}p dt - h_{1,x}(0) \cdot p_{1,0} \\ &= \int_0^T B(y - y_{\text{data}}) \cdot h_x + E \frac{d}{dt}p \cdot h_x + (J^\top - R^\top)p \cdot h_x dt - p_1(T) \cdot h_x(T). \end{aligned}$$

In the second equation, we set $p_{1,0} := p_1(0)$ and used $E h_x(0) \cdot p(0) = h_x(0)^\top \text{diag}(I_r, 0)p(0) = h_{1,x}(0)^\top p_{1,0}$. As $d_x L(x, w, \mathbf{p})[h_x] = 0$ must hold for arbitrary h_x we identify the *adjoint equation* as

$$-E \frac{d}{dt}p = (J^\top - R^\top)p + B(y - y_{\text{data}}), \quad p_1(T) = 0. \quad (36)$$

Like the state equation, this adjoint equation is an pH-DAE of index-1, since $J^\top$ is a skew-symmetric matrix and $R^\top = R \succeq 0$. The index-1 condition follows directly from $\det((J_{22} - R_{22})^\top) = \det(J_{22} - R_{22}) \neq 0$. The difference is that the equation is solved backward in time with end condition $p_1(T) = 0$, and $p_{2,T}$ is calculated similarly to $x_{2,0}$.

The derivative $d_p L(x, w, \mathbf{p})[h_p] = 0$ yields the state equation (24). The optimality condition is derived as

$$\begin{aligned} d_w L(x, w, \mathbf{p})[h_w] &= d_w \mathcal{J}(x, w)[h_w] - d_w \langle e(x, w), \mathbf{p} \rangle[h_w] \\ &= d_w \mathcal{J}_1(x, w)[h_w] + d_w \mathcal{J}_2(w)[h_w] - d_w \langle e(x, w), \mathbf{p}[h_w] \rangle. \end{aligned} \quad (37)$$

For simplicity, we split the derivation into the different matrices and the initial condition. Moreover, we omit the term $\mathcal{J}_2$ during the calculations and will later add the term $\nabla \mathcal{J}_2$ to obtain the full gradient. For the derivative w.r.t. to $x_{1,0}$, we obtain

$$d_{x_{1,0}} \langle e(x, w), \mathbf{p} \rangle[h_{x_{1,0}}] = \langle p_{1,0}, h_{x_{1,0}} \rangle_{\mathbb{R}^r} = (p_{1,0}, h_{x_{1,0}})_{\mathbb{R}^r}. \quad (38)$$

Since $h_{x_{1,0}}$ was chosen arbitrarily, the Riesz representation theorem allows us to identify $\nabla_{x_{1,0}} \hat{\mathcal{J}}(w) = p_1(0) + \nabla_{x_{1,0}} \mathcal{J}_2(w)$. The gradients w.r.t. the matrices are derived analogously. Thus, for the derivative with respect to matrix J , we obtain with the Frobenius inner product

$$\begin{aligned} d_J \langle e(x, w), \mathbf{p} \rangle [h_J] &= - \int_0^T d_J \left(\left(E \frac{d}{dt} - (J - R)x - Bu \right) \cdot p \right) [h_J] dt \\ &= \int_0^T h_J x \cdot p dt = \int_0^T p^\top h_J x dt = \int_0^T \text{tr}(p^\top h_J x) dt = \int_0^T \text{tr}(x p^\top h_J) dt \\ &= \int_0^T \langle p x^\top, h_J \rangle_F dt = \left\langle \int_0^T p x^\top dt, h_J \right\rangle_F. \end{aligned}$$

From the last equation, we conclude that the gradient with respect to J is equal to $\int_0^T p x^\top dt$. The same calculation applies to matrix R , but with the opposite sign. The last directionally derivative with respect to matrix B is given by

$$\begin{aligned} d_B \mathcal{J}_1(x, w) [h_B] - d_B \langle e(x, w), \mathbf{p} \rangle [h_B] \\ &= \frac{1}{2} \int_0^T d_B |B^\top x - y_{\text{data}}|^2 [h_B] dt - \int_0^T d_B \left(\left(E \frac{d}{dt} - (J - R)x - Bu \right) \cdot p \right) [h_B] dt \\ &= \int_0^T x^\top h_B (B^\top x - y_{\text{data}}) + p^\top h_B u dt = \int_0^T \text{tr}(x^\top h_B (y - y_{\text{data}})) + \text{tr}(p^\top h_B u) dt \\ &= \int_0^T \langle x(y - y_{\text{data}})^\top, h_B \rangle_F + \langle p u^\top, h_B \rangle_F dt = \left\langle \int_0^T x(y - y_{\text{data}})^\top + p u^\top dt, h_B \right\rangle_F. \end{aligned}$$

The Riesz representation theorem once again allows us to identify the gradient

$$0 = \nabla_J \mathcal{J}_2(\bar{w}) + \int_0^T \bar{p}(t) \otimes \bar{x}(t) dt, \quad (39)$$

$$0 = \nabla_R \mathcal{J}_2(\bar{w}) - \int_0^T \bar{p}(t) \otimes \bar{x}(t) dt, \quad (40)$$

$$0 = \nabla_B \mathcal{J}_2(\bar{w}) + \int_0^T (\bar{p}(t) \otimes u(t) + (\bar{x}(t)) \otimes (\bar{y}(t) - y_{\text{data}}(t))) dt, \quad (41)$$

$$0 = \nabla_{x_{1,0}} \mathcal{J}_2(w) + \bar{p}_1(0), \quad (42)$$

where $\bar{p}$ satisfied the adjoint equation

$$-E \frac{d}{dt} \bar{p} = (\bar{J}^\top - \bar{R}^\top) \bar{p} + \bar{B}(\bar{y} - y_{\text{data}}), \quad \bar{p}_1(T) = 0, \quad (43)$$

and $\bar{x}$ the state equation with output $\bar{y}$ given by

$$\begin{aligned} E \frac{d}{dt} \bar{x} &= (\bar{J} - \bar{R}) \bar{x} + \bar{B}u, & \bar{x}_1(0) &= \bar{x}_{1,0}, \\ \bar{y} &= \bar{B}^\top \bar{x}, \end{aligned}$$

where $a \otimes b$ denotes the dyadic product $ab^\top$ for $a, b \in \mathbb{R}^n$.

Now, we apply the transformation we previously introduced and obtain for $J = K - K^\top$ and $R = LL^\top$ that it holds

$$\begin{aligned}\nabla_K \hat{\mathcal{J}}(w) &= \nabla_J \hat{\mathcal{J}}(w) - \nabla_J \hat{\mathcal{J}}(w)^\top, \\ \langle \hat{\mathcal{J}}'(w), h_L \rangle_F &= \text{tr}(\hat{\mathcal{J}}_R(w)^\top h_L L^\top) + \text{tr}(\hat{\mathcal{J}}_R(w)^\top L h_L^\top) = \langle \nabla_R \hat{\mathcal{J}}(w) L + \nabla_R \hat{\mathcal{J}}(w)^\top L, h_L \rangle_F \\ &\Rightarrow \nabla_L \hat{\mathcal{J}}(w) = (\nabla_R \hat{\mathcal{J}}(w) + \nabla_R \hat{\mathcal{J}}(w)^\top) L.\end{aligned}\tag{45}$$

Due to the transformation, the problem is unconstrained and we can summarize the derivation with the help of the following first-order optimality condition.

Theorem 5. *A stationary point $\bar{v} = (\bar{K}, \bar{L}, \bar{B}, \bar{x}_{1,0})$ of the transformed reduced problem satisfies the optimality condition*

$$0 = \nabla_K \hat{\mathcal{J}}(\bar{v}) = \nabla_K \mathcal{J}_2(\bar{v}) + \int_0^T \bar{p}(t) \otimes \bar{x}(t) dt - \int_0^T \bar{x}(t) \otimes \bar{p}(t) dt, \tag{46a}$$

$$0 = \nabla_L \hat{\mathcal{J}}(\bar{v}) = \nabla_L \mathcal{J}_2(\bar{v}) - \left(\int_0^T \bar{p}(t) \otimes \bar{x}(t) dt + \int_0^T \bar{x}(t) \otimes \bar{p}(t) dt \right) \bar{L}, \tag{46b}$$

$$0 = \nabla_B \mathcal{J}_2(\bar{w}) + \int_0^T (\bar{p}(t) \otimes u(t) + (\bar{x}(t)) \otimes (\bar{y}(t) - y_{\text{data}}(t))) dt, \tag{46c}$$

$$0 = \nabla_{x_{1,0}} \mathcal{J}_2(w) + \bar{p}_1(0), \tag{46d}$$

where $\bar{p}$ satisfied the adjoint equation

$$-E \frac{d}{dt} \bar{p} = (\bar{K}^\top - \bar{K} - \bar{L}^\top \bar{L}) \bar{p} + \bar{B}(\bar{y} - y_{\text{data}}), \quad \bar{p}_1(T) = 0, \tag{47}$$

and $\bar{x}$ the state equation with output $\bar{y}$ given by

$$\begin{aligned}E \frac{d}{dt} \bar{x} &= (\bar{K} - \bar{K}^\top - \bar{L} \bar{L}^\top) \bar{x} + \bar{B} u, & \bar{x}_1(0) &= \bar{x}_{1,0}, \\ \bar{y} &= \bar{B}^\top \bar{x},\end{aligned}$$

where $a \otimes b$ denotes the dyadic product $ab^\top$ for $a, b \in \mathbb{R}^n$.

Note that it is difficult to solve the first-order optimality system altogether due to its forward-backward structure. Instead, we proceed iteratively and first solve the state equation in the forward direction, followed by the adjoint equation in the backward direction. The gradient (46), which is computed based on the state and adjoint variables, then allows us to define a suitable descent direction. Finally, we formulate all steps in Algorithm 1.

Algorithm 1: Algorithm for the calibration process

Data: initial guess v , algorithmic parameters, stopping criterion

Result: optimized matrices and initial data $\bar{w} = (\bar{J}, \bar{R}, \bar{B}, \bar{x}_{1,0})$

- 1) set $J = K - K^\top$ and $R = LL^\top$;
 - 2) solve state equation (24) $\rightarrow x$;
 - 3) solve adjoint equation (36) $\rightarrow p$;
 - 4) evaluate the gradient (46) and (45) $\rightarrow \nabla \mathcal{J}(w)$;
 - 5) transform $\nabla \mathcal{J}(w)$ into a vector;
 - 5) compute the direction of descent via conjugate gradient method [22] $\rightarrow s$;
 - 6) find an admissible step size via Armijo rule [19] $\rightarrow \sigma$;
 - 7) update $v \mapsto v + \sigma s$;
 - 8) **if** stopping criterion is not fulfilled $\rightarrow$ go to 1);
else return optimal control $\bar{w}$;
-

Possible stopping criteria are $\|\nabla f(x)\| < \varepsilon$ for $0 < \varepsilon \ll 1$ or when the maximal number of iterations is reached. Now, we will formally prove that the calculated gradient corresponds to the gradient, obtained by directly differentiating the reduced cost functional $\hat{\mathcal{J}}$. Note that the state equation (24) is treated implicitly, and that x and y can be expressed in terms of $S(w)$ and $B^\top x$, respectively. To identify the gradient of the reduced cost functional, we first differentiate the equation $e(x, w) = e(S(w), w) = 0$ with respect to w in the direction $h \in W$ to obtain using the implicit function theorem

$$d_x e(S(w), w) S'(w)[h] + d_w e(S(w), w)[h] = 0 \Rightarrow S'(w)[h] = -d_x e(S(w), w)^{-1} d_w e(S(w), w)[h].$$

This allows us to find the directional derivative of $\hat{\mathcal{J}}(w) = \mathcal{J}(x, w)$:

$$\begin{aligned} d\hat{\mathcal{J}}(w)[h] &= \langle d_y \mathcal{J}(y, w), B^\top S'(w)[h] \rangle_{H^{-1}, H^1} + \langle d_w \mathcal{J}(y, w), h \rangle_{W^*, W} \\ &= \langle -d_w e(S(w), w)^* d_x e(S(w), w)^{-*} [B d_y \mathcal{J}(y, w)], h \rangle_{W^*, W} + \langle d_w \mathcal{J}(y, w), h \rangle_{W^*, W} \\ &= \langle d_w \mathcal{J}(y, w) - d_w e(S(w), w)^* d_x e(S(w), w)^{-*} [B d_y \mathcal{J}(y, w)], h \rangle_{W^*, W}. \end{aligned}$$

Since W is a Hilbert space and h was chosen arbitrarily, the Riesz theorem yields

$$(\nabla \hat{\mathcal{J}}(w), h)_W = \langle d_w \mathcal{J}(y, w) - d_w e(S(w), w)^* d_x e(S(w), w)^{-*} [B d_y \mathcal{J}(y, w)], h \rangle_{W^*, W}.$$

The gradient is therefore given by

$$\nabla \mathcal{J}(w) = d_w \mathcal{J}(y, w) - d_w e(S(w), w)^* \underbrace{d_x e(S(w), w)^{-*} [B d_y \mathcal{J}(y, w)]}_{=: p} \quad (49a)$$

$$= d_w \mathcal{J}(y, w) - d_w e(S(w), w)^* p. \quad (49b)$$

In this context, $d_x e(S(w), w)^* p = B d_y \mathcal{J}(y, w)$ describes the adjoint equation (36) and the equation (49b) is exactly the gradient calculated above, i.e. the left-hand side of system (46).

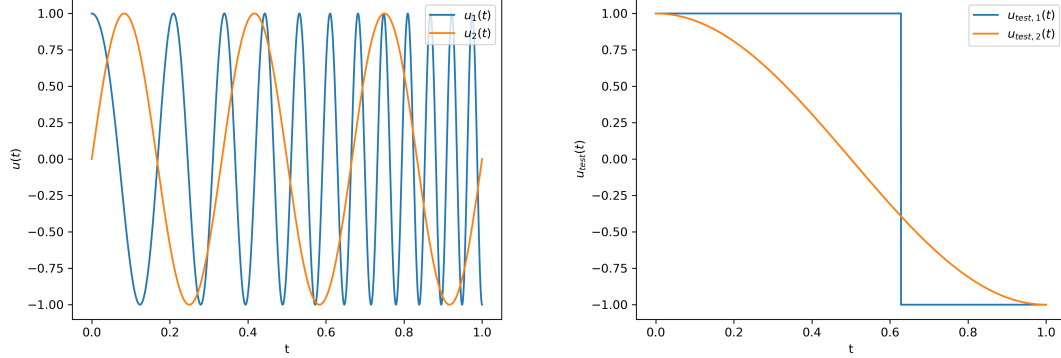


Figure 2: Signal u for the calibration (left) and signal u_{test} for cross-validation (right)

6. Numerical Results

For the numerical results we consider the interval $[0, T]$ with $T = 1$ and 1001 discretization points, resulting in a step size of $h = 0.0001$. We set $m = 2$ for all numerical test and used the same inputs as in [15] (see Figure 2)

$$u: [0, T] \rightarrow \mathbb{R}^2, \quad u(t) = \begin{pmatrix} \text{chirp}_{3,20}(t) \\ \sin(2\pi f_0 t) \end{pmatrix},$$

where chirp denotes the chirp-signal from `scipy.signal` toolbox with frequency 3 at time $t = 0$ and frequency 20 at time $t = 1$ and $f_0 = 3$. To check the robustness of the identified system, we perform a cross-validation using the test signal (see Figure 2)

$$u_{\text{test}}: [0, T] \rightarrow \mathbb{R}^2, \quad u_{\text{test}}(t) = \begin{pmatrix} \text{square}_{0.1}(t) \\ \cos(\pi t) \end{pmatrix},$$

where `square0.1` is the square signal of the `scipy.signal` toolbox with duty parameter 0.1.

6.1. Synthetic data

Below, we present the results for a synthetic dataset with $n = 5$ and $m = 2$. To this end, we select w_{data} and use it to compute y_{data} . We initialize B_{data} with independent uniformly distributed entries in $[-0.2, 0.2]$ and add $\text{diag}(4, 2)$ to the first $m \times m$ block. For the matrix R_{data} , we draw an $n \times n$ matrix L with independent uniformly distributed entries in $[-1, 1]$. Then, we set the strictly upper triangular part of L as well as the elements $l_{11} = l_{22}$ to zero. Finally, we define $R_{\text{data}} = L^\top L$, which is positive semidefinite. To generate a feasible initial J , we draw an $n \times n$ matrix with independent uniformly distributed entries in $[-1, 1]$. We take the upper part of this matrix, starting with the first off-diagonal, and fill the lower part of the matrix with the negative of the transpose of the upper part. Since we restrict ourselves to index-1 DAEs, we must check if J and R are admissible. The initial value of the differential component $x_{1,0,\text{data}}$ is a random vector with entries in $[-0.1, 1]$.

First, the number of differential variables r is 4, while in the second case, it is 2. This allows us to investigate how the number of differential variables affects the identification process. For all tests, we assume to have no information about the system, so, we set $\lambda = 0$.

Example 1 (System with four Differential Variables). In this test, we generated an output $y_{\text{data}} = y_{\text{ref}}$ with $r_{\text{ref}} = 4$. To investigate the influence of the choice of r , we solved a minimization problem for each $r \in \{1, 2, 3, 4, 5\}$. The same random initial values for the matrices and the first r components of the random vector $x_{1,0}$ were used for each minimization problem. Figure 3 shows the evolution of the cost for all five test cases. As can be seen, the cost functional for the system with one differential variable is the highest; in the other cases, the costs have decreased significantly. Table 1 lists the costs as well as the maximum norm error $\|y - y_{\text{ref}}\|_{\infty}$ for the first and last iterations. We observe that the costs and the maximum norm error are lowest for $r = 3$. For all test cases except $r = 1$, however, the reduction in the cost functional is above 99 %. Figures 4 and 5 show the two system outputs before and after the calibration together with the reference output (dashed yellow line) for $r = 1, \dots, 5$. As the error values already indicated, the fitted outputs match the reference output very well, except for the system with only one differential variable.

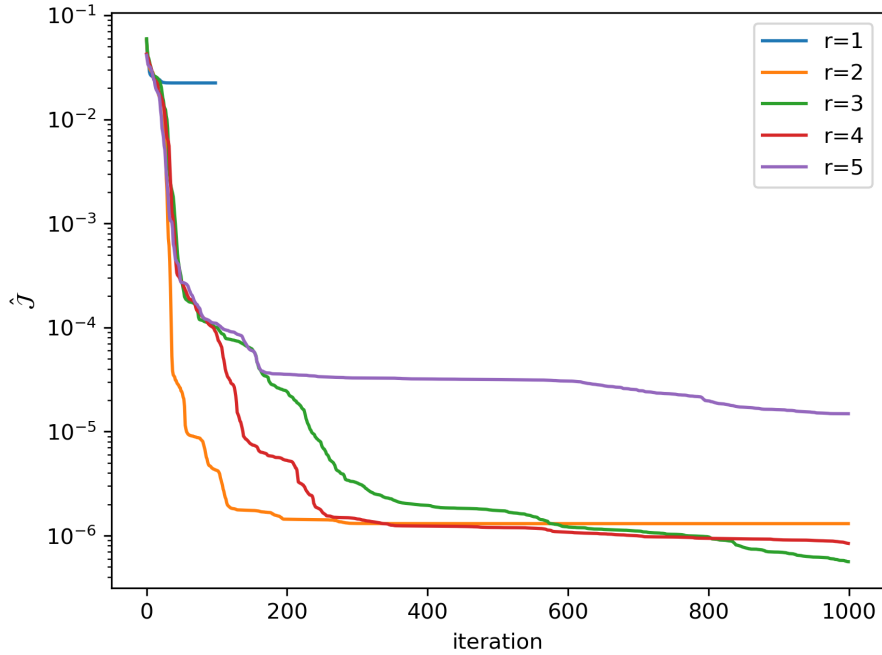


Figure 3: Evolution of the cost for various r over the iterations of the calibration in log plot for $r_{\text{ref}} = 4$.

Note that a single signal is insufficient for determining how well the fitted system represents the true system. Therefore, we conclude this example with a cross-validation test. We apply the input signal u_{test} to the fitted and reference systems, and we compare the corresponding outputs. Table 2 shows the L^2 -difference between y and y_{test} as well as the

number of differential variables r	1	2	3	4	5
cost in first iteration	5.75e-2	5.83e-2	5.95e-2	4.25e-2	4.14e-2
cost in last iteration	2.24e-2	1.30e-6	5.62e-7	8.42e-7	1.49e-5
$\ y - y_{\text{ref}}\ _{\infty}$ in first iteration	7.48e-1	7.70e-1	7.12e-1	7.86e-1	7.81e-1
$\ y - y_{\text{ref}}\ _{\infty}$ in last iteration	4.88e-1	2.95e-3	2.41e-3	3.45e-3	9.84e-3

Table 1: Initial and final cost values, the L^2 - and the maximum norm between the fitted output and reference output for various numbers of differential variables r for $r_{\text{ref}} = 4$.

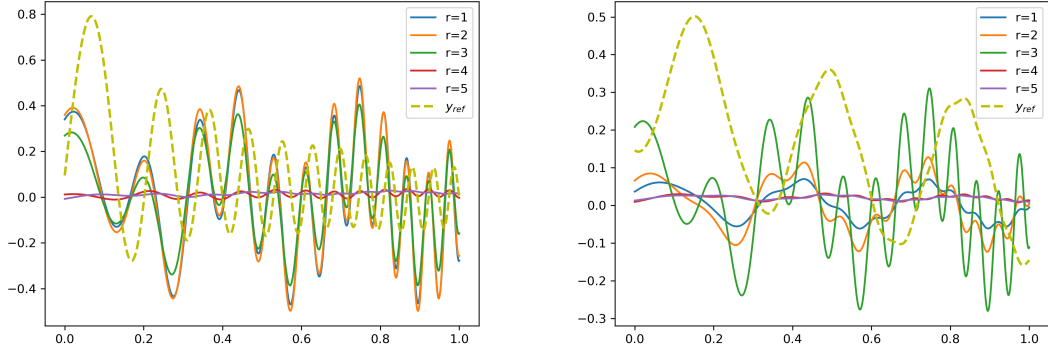


Figure 4: Left: First component of the output before the calibration for various r and for the reference system y_{ref} ; Right: Second component of the output before the calibration for various r and for the reference system y_{ref} for $r_{\text{ref}} = 4$.

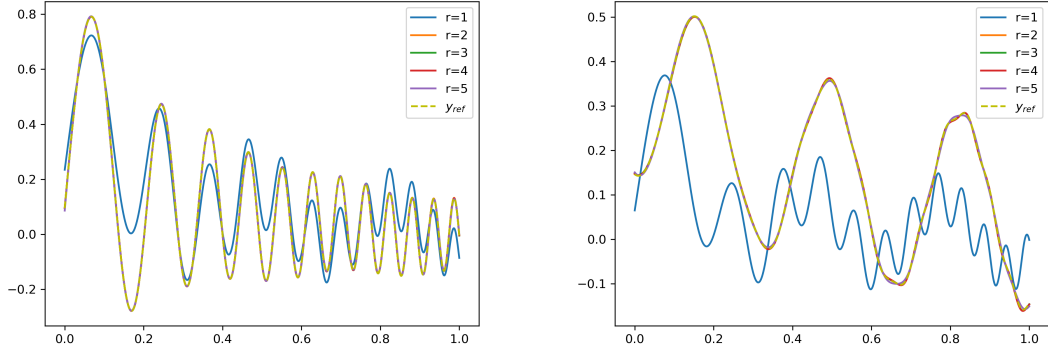


Figure 5: Left: First component of the output after the calibration for various r and for the reference system y_{ref} ; Right: Second component of the output after the calibration for various r and for the reference system y_{ref} for $r_{\text{ref}} = 4$.

maximum norm. As we can see, the ODE system ($r = 5$) has the largest error in the cross-validation test, while the error for $r = 4$ is the smallest and closest to the actual value. Figure 6 shows the results of this cross-validation test graphically and the outputs for the fitted systems $r = 2, 3$ also match well.

number of differential variables r	1	2	3	4	5
$\ y - y_{\text{test}}\ _{L^2}$	3.96	3.29e-1	4.42e-1	2.62e-1	5.11
$\ y - y_{\text{test}}\ _{\infty}$	5.73	6.25e-1	6.26e-1	4.60e-1	8.67

Table 2: The L^2 -difference and the maximum norm between the fitted output and reference output for various r in the cross-validation test and $r_{\text{ref}} = 4$.

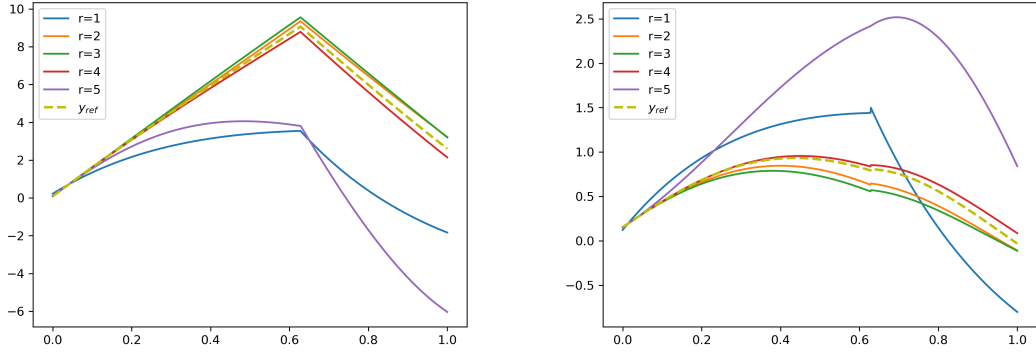


Figure 6: Left: First component of the output obtained in the cross-validation for various r . Right: Second component of the output obtained in the cross-validation for various r for $r_{\text{ref}} = 4$.

Example 2 (System with two Differential Variables). We used the same setup as in the first test. This time, we generated the y_{data} output using a pH-DAE system with $r_{\text{data}} = 2$ differential variables. As before, we solved the optimization problem for different values of r and compared the corresponding results. Figure 7 shows the evolution of the cost and Table 3 shows the costs and errors for the first and last iterations. As before, we observe that the cost for the system with $r = 1$ decreases only marginally. In contrast, the cost reduction is again greater than 99% for all other systems. Figure 8 shows the output obtained with the initial matrices before calibration in blue. The orange graph shows the output of the fitted system and the green dashed line shows the output obtained with the reference data. As the error values above indicated, the fitted output closely matches the reference output.

Table 4 and Figure 9 show the results for the cross-validation test. The fitted system with $r = 2$ provides the best results, yielding the lowest cost and the smallest errors.

In summary, we successfully identified pH-DAE systems using our approach. However, the results depend heavily on the choice of the hyperparameter r , at least in the cross-validation test. We note that our problem does not have a unique solution and can therefore depend heavily on the initial data v^0 . In other words, two different surrogate systems that minimize the cost functional may respond differently to the test signal.

6.2. Model order reduction test with a RCL Ladder Network

The last test case we consider is taken from the PHS benchmark collection [23] and generates pH-DAE models for electrical circuits consisting of ideal voltage sources, resistors,

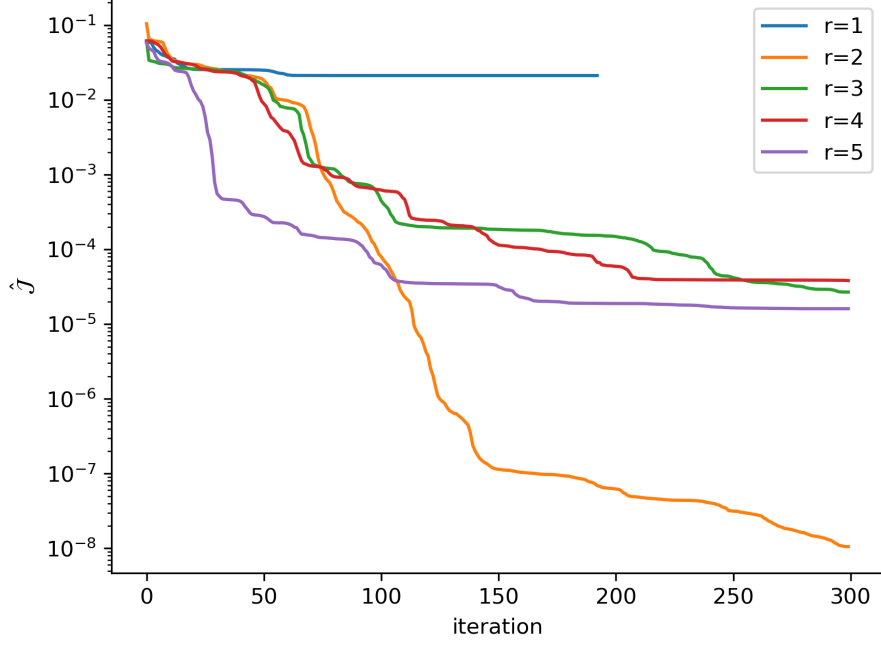


Figure 7: Evolution of the cost for various r over the iterations of the calibration in log plot for $r_{\text{ref}} = 2$

number of differential variables r	1	2	3	4	5
cost in first iteration	5.96e-2	1.04e-1	6.18e-2	6.17e-2	5.78e-2
cost in last iteration	2.12e-2	1.05e-8	2.68e-5	3.82e-5	1.61e-5
$\ y - y_{\text{ref}}\ _{\infty}$ in first iteration	7.84e-1	9.70e-1	8.74e-1	7.94e-1	7.96e-1
$\ y - y_{\text{ref}}\ _{\infty}$ in last iteration	4.81e-1	2.41e-4	1.20e-2	1.59e-2	9.43e-3

Table 3: Initial and final cost values and the maximum norm between the fitted output and reference output for various numbers of differential variables r for $r_{\text{ref}} = 2$.

inductors and capacitors. Modeling of such circuits with directed graphs (see Figure 10) leads to systems of the following form:

$$\begin{aligned} E\dot{x}(t) &= (J - R)x(t) + Bu(t), \\ y(t) &= B^{\top}x(t), \end{aligned}$$

where

$$E = \begin{pmatrix} \mathcal{A}_c C \mathcal{A}_c^T & 0 & 0 \\ 0 & L & 0 \\ 0 & 0 & 0 \end{pmatrix}, J = \begin{pmatrix} 0 & -\mathcal{A}_l & -\mathcal{A}_v \\ \mathcal{A}_l^{\top} & L & 0 \\ \mathcal{A}_v^{\top} & 0 & 0 \end{pmatrix}, R = \begin{pmatrix} \mathcal{A}_r R^{-1} \mathcal{A}_r^T & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 \\ 0 \\ -I_2 \end{pmatrix}.$$

Here, the matrices R , L , and C are positive definite diagonal matrices containing the resistances, inductances and capacitances as entries. The incidence matrices $\mathcal{A}_r$, $\mathcal{A}_l$, $\mathcal{A}_c$, and $\mathcal{A}_v$

number of differential variables r	1	2	3	4	5
$\ y - y_{\text{test}}\ _L^2$	3.96	3.68e-2	7.25e-1	2.84e-1	1.25
$\ y - y_{\text{test}}\ _\infty$	5.75	6.71e-2	1.06	3.80e-1	1.85

Table 4: Cross validation cost values and the maximum norm between the fitted output and reference output for various numbers of differential variables r for $r_{\text{ref}} = 2$.

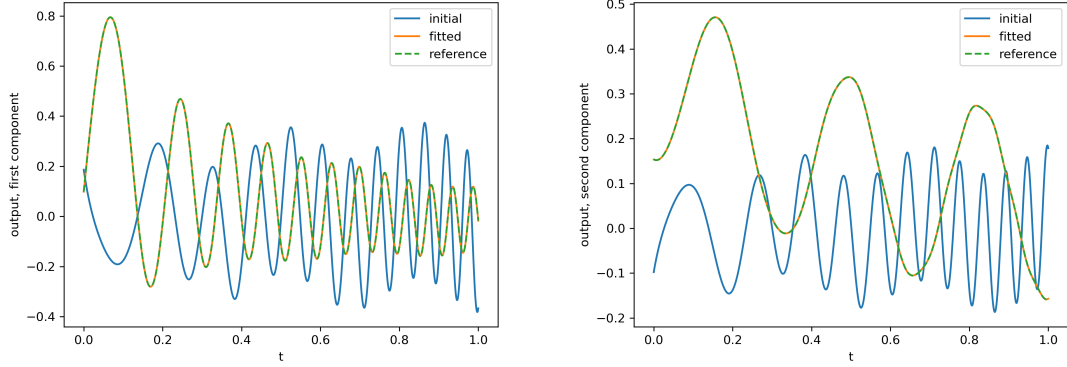


Figure 8: Left: First component of the output for the before the calibration (initial), after the calibration (fitted) and for the reference matrices (reference) used for the generation of the synthetic data. Right: Second component of the output for the before the calibration (initial), after the calibration (fitted) and for the reference matrices (reference) used for the generation of the synthetic data for $r_{\text{ref}} = 2$.

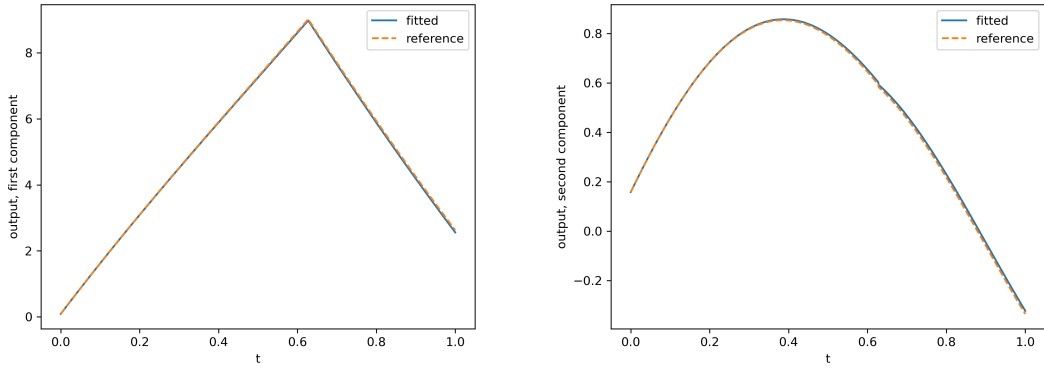


Figure 9: Left: First component of the output obtained in the cross-validation. Right: Second component of the output obtained in the cross-validation for $r_{\text{ref}} = 2$.

follow directly from the directed graph of the network and contain only entries in $\{-1, 0, 1\}$. The system's inputs $u(t)$ are the voltages $v_v(t)$ provided by the voltage sources and the outputs are the negative currents $i_v(t)$ through the voltage sources. The state vector is $x(t) := (v(t)^\top i_l(t)^\top i_v(t)^\top)^\top$ with node voltages $v(t)$ and inductor currents $i_l(t)$. Thus, the model dimension is $n = 3\bar{n} + 4$, where $\bar{n}$ denotes the number of loops. Here, we choose the value $\bar{n} = 32$ and set all resistances to 0.2, and capacitances and inductances to 1.

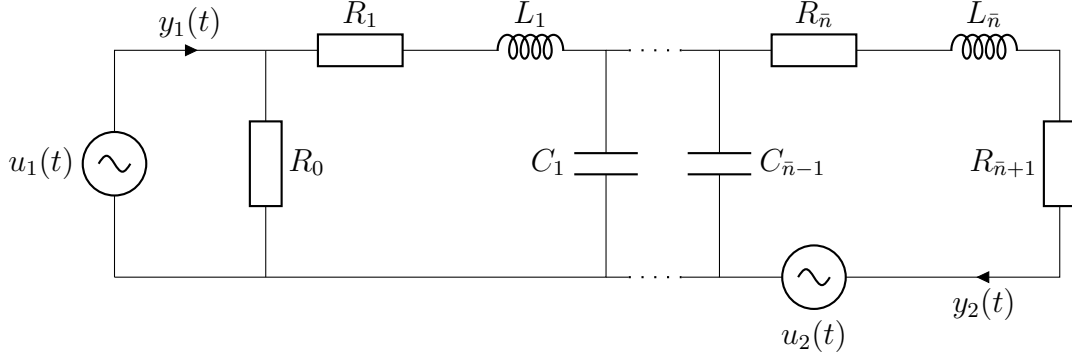


Figure 10: A schematic of the considered RCL Ladder Network, cf. [23].

Due to the simple choice of L and C , the system is already in semi-explicit form (21). Hence, the reference data y_{data} can be computed directly using the implicit midpoint rule, and the number of differential variables is given by $r = \text{rank } E$. We took a $2n^2 + mn + r \times 1$ vector with entries in $[-1, 1]$ as the initial value v^0 and a $r \times 1$ vector with entries in $[0, 1]$ as the initial value $x_{1,0}$. Our objective was to investigate whether a reduced system exhibiting the same behavior as the original system could be found. In doing so, we examined various ratios r/n . Figure 11 shows the evolution of the costs for 12 test cases, i.e., $N \in \{100, 50, 20\}$ and $r/n \in \{1, 0.9, 0.6, 0.3\}$. Table 5 shows that all costs are smaller than $1e-5$ and that the relative errors after the calibration are smaller than 1 %, except for the 3 cases where $N = r$. Figure 12 confirms that the outputs only match well for the real DAE cases, i.e., $r < n$. To conclude, we will now consider the cross-validation test. The results are shown in Table 6 and Figure 13. As can be seen here, our pH-DAE system cannot be represented by an ODE system. However, the outputs of the other systems are close to those of the original system.

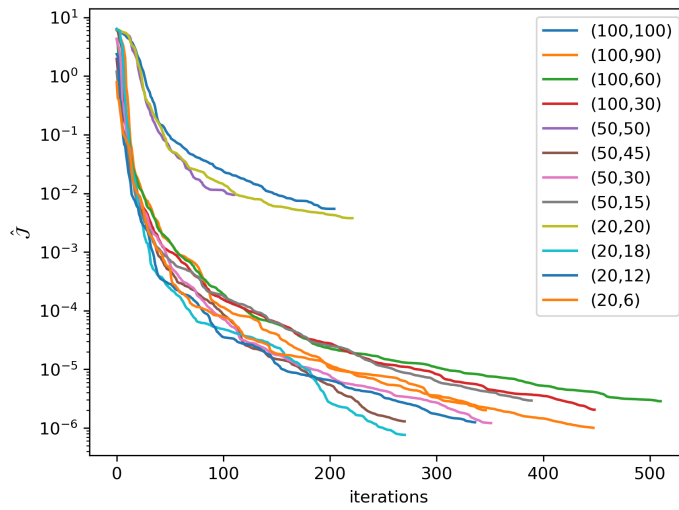


Figure 11: Evolution of the cost for various r and n for $N_{\text{ref}} = 100$ and $r_{\text{ref}} = 63$.

n	r	cost in last iter.	$\ y - y_{\text{ref}}\ _{\infty} / \ y_{\text{ref}}\ _{\infty}$ before	$\ y - y_{\text{ref}}\ _{\infty} / \ y_{\text{ref}}\ _{\infty}$ after
100	100	5.45e-3	115%	12.18%
100	90	1.01e-6	106%	0.10%
100	60	2.86e-6	135%	0.16%
100	30	2.05e-6	124%	0.08%
50	50	9.44e-3	100%	21.87%
50	45	1.30e-6	159%	0.11%
50	30	1.21e-6	118%	0.08%
50	15	2.93e-6	90%	0.08%
20	20	3.78e-3	128%	13.95%
20	18	7.61e-7	135%	0.08%
20	12	1.25e-6	105%	0.11%
20	6	2.01e-6	120%	0.09%

Table 5: Relative errors of the output measured before and after the calibration.

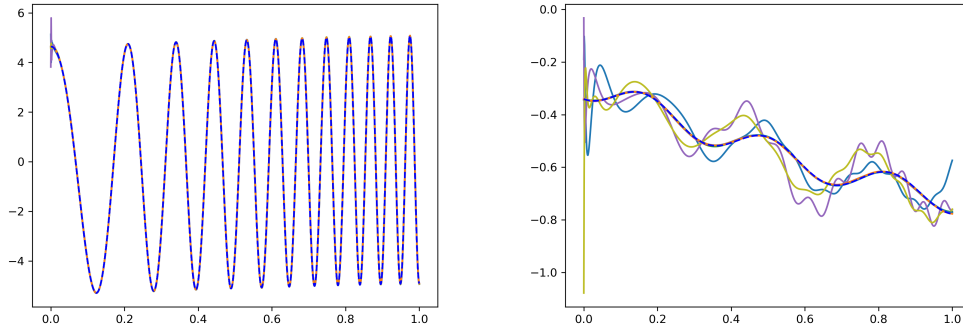


Figure 12: Left: first component of the fitted output for different n , r , and y_{ref} (blue dashed line). Right: second component of the fitted output for different n , r , and y_{ref} (blue dashed line).

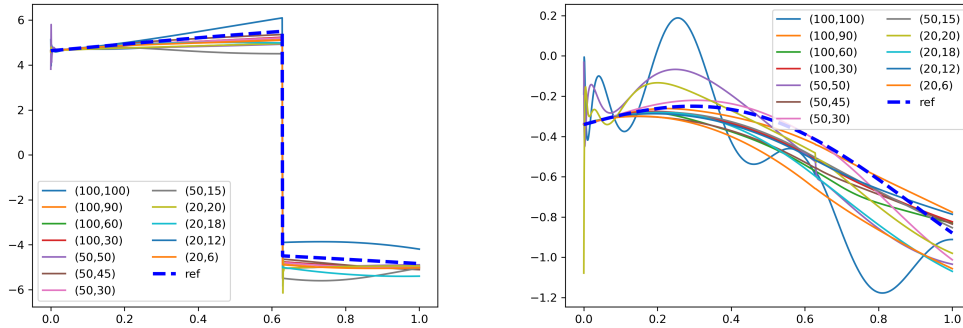


Figure 13: Left: first component of the fitted output for different n , r , and y_{ref} (blue dashed line). Right: second component of the fitted output for different n , r , and y_{ref} (blue dashed line).

n	r	$\ y - y_{\text{test}}\ _{L^2((0,1),\mathbb{R}^2)}$	$\ y - y_{\text{test}}\ _{\infty}$
100	100	0.46	3.36
100	90	0.23	0.33
100	60	0.22	0.29
100	30	0.21	0.30
50	50	0.40	3.74
50	45	0.18	0.28
50	30	0.22	0.33
50	15	0.66	1.03
20	20	0.36	2.81
20	18	0.45	0.67
20	12	0.50	0.77
20	6	0.33	0.41

Table 6: Difference of fitted output and reference output in maximum / L^2 norm in cross-validation test.

7. Conclusion

We presented a data-driven approach for identifying linear index-1 port-Hamiltonian differential-algebraic equation (pH-DAE) systems from input-output data. By incorporating the pH structure into the parameterization, the identified models retain the relevant structural properties. The index-1 structure is exploited to formulate the identification problem as an unconstrained optimization problem, while an adjoint formulation provides the gradient of the reduced cost functional efficiently.

Numerical experiments demonstrate that the proposed approach can construct pH-DAE surrogate models that accurately reproduce the input-output behaviour of the reference systems. The cross-validation experiments further indicate that the identified models can generalize to independent input signals. The experiments also show the potential for constructing lower-dimensional structure-preserving surrogate models.

Future work will focus on extending the approach to higher-index and nonlinear pH-DAE systems. Further directions include the treatment of noisy and incomplete measurements, uncertainty quantification, and the systematic combination of identification with structure-preserving model reduction for large-scale multiphysics applications.

Acknowledgements

M. Ehrhardt, M. Günther and C. Totzeck acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 531152215 – CRC 1701.

Author contributions

TZ Investigation, Software, Validation, Visualization, Writing – original draft

ME Investigation, Project administration, Supervision, Writing – review and editing

MG Conceptualization, Methodology, Writing – review and editing

CT Methodology, Supervision, Writing – review and editing

References

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, 2008. [doi:10.1515/9781400830244](https://doi.org/10.1515/9781400830244).
- [2] N. Andrei, *Modern Numerical Nonlinear Optimization*, Springer Optimization and Its Applications, Volume 195. Springer International Publishing, Cham, 2022. [doi:10.1007/978-3-031-08720-2](https://doi.org/10.1007/978-3-031-08720-2).
- [3] A. Bartel, M. Clemens, M. Günther, B. Jacob, and T. Reis, *Port-Hamiltonian systems' modelling in electrical engineering*, In: M. van Beurden, N.V. Budko, G. Ciuprina, W. Schilders, H. Bansal, and R. Barbulescu, (eds.), Scientific Computing in Electrical Engineering. SCEE 2022. Mathematics in Industry vol 43. Springer, Cham, 2024, pages 133–143. [doi:10.1007/978-3-031-54517-7_15](https://doi.org/10.1007/978-3-031-54517-7_15).
- [4] A. Bartel, M. Diab, A. Frommer, M. Günther, and N. Marheineke, *Splitting techniques for DAEs with port-Hamiltonian applications*, Appl. Numer. Math. 214 (2025), 28–53. [doi:10.1016/j.apnum.2025.03.004](https://doi.org/10.1016/j.apnum.2025.03.004).
- [5] A. Bartel, M. Günther, B. Jacob, and T. Reis, *Operator splitting based dynamic iteration for linear differential-algebraic port-Hamiltonian systems*, Numer. Math. 155(1-2) (2023), 1–34. [doi:10.1007/s00211-023-01369-5](https://doi.org/10.1007/s00211-023-01369-5).
- [6] C. Beattie, V. Mehrmann, H. Xu, and H. Zwart, *Linear port-Hamiltonian descriptor systems*, Math. Contr. Sign. Syst. 30(4) (2018) 17. [doi:10.1007/s00498-018-0223-3](https://doi.org/10.1007/s00498-018-0223-3).
- [7] P. Benner, P. Goyal, and P. van Dooren, *Identification of port-Hamiltonian systems from frequency response data*, Syst. Contr. Lett. 143 (2020), 104741. [doi:10.1016/j.sysconle.2020.104741](https://doi.org/10.1016/j.sysconle.2020.104741).
- [8] R. Byers, P. Kunkel, and V. Mehrmann, *Regularization of Linear Descriptor Systems with Variable Coefficients*, SIAM J. Contr. Optim. 35(1) (1997), 117–133. [doi:10.1137/S0363012994278936](https://doi.org/10.1137/S0363012994278936).
- [9] K. Cherifi, P.K. Goyal, and P. Benner, *A non-intrusive method to inferring linear port-Hamiltonian realizations using time-domain data*, Electr. Trans. Numer. Anal. 56 (2022), 102–116. [doi:10.1553/etna_vol56s102](https://doi.org/10.1553/etna_vol56s102).
- [10] K. Cherifi, A. El Messaoudi, H. Gernandt, and M. Roschkowski, *Nonlinear port-Hamiltonian system identification from input-state-output data*, 2025. [arXiv:2501.06118](https://arxiv.org/abs/2501.06118), [doi:10.48550/arXiv.2501.06118](https://doi.org/10.48550/arXiv.2501.06118).

- [11] M. Ehrhardt, M. Günther, and D. Ševčovič, *Structure-preserving coupling and decoupling of port-Hamiltonian systems*, Appl. Math. Lett. 177 (2026), 109894. doi:[10.1016/j.aml.2026.109894](https://doi.org/10.1016/j.aml.2026.109894).
- [12] A. Frommer, M. Günther, B. Liljegren-Sailer, and N. Marheineke, *Operator splitting for port-Hamiltonian systems*, 2023. [arXiv:2304.01766](https://arxiv.org/abs/2304.01766), doi:[10.48550/arXiv.2304.01766](https://doi.org/10.48550/arXiv.2304.01766).
- [13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, 2013. doi:[10.56021/9781421407944](https://doi.org/10.56021/9781421407944).
- [14] O. Gonzalez, *Time integration and discrete Hamiltonian systems*, J. Nonlin. Sci. 6(5) (1996), 449–467. doi:[10.1007/BF02440162](https://doi.org/10.1007/BF02440162).
- [15] M. Günther, B. Jacob, and C. Totzeck, *Data-driven adjoint-based calibration of port-Hamiltonian systems in time domain*, Math. Contr. Sign. Syst. 36(4) (2024), 957–977. doi:[10.1007/s00498-024-00389-2](https://doi.org/10.1007/s00498-024-00389-2).
- [16] M. Günther, B. Jacob, and C. Totzeck, *Structure-preserving identification of port-Hamiltonian systems — a sensitivity-based approach*, In: M. van Beurden, N.V. Budko, G. Ciuprina, W. Schilders, H. Bansal, and R. Barbulescu, (eds.), *Scientific Computing in Electrical Engineering. SCEE 2022*. Mathematics in Industry vol 43. Springer, Cham, 2024, pages 167–174. doi:[10.1007/978-3-031-54517-7_19](https://doi.org/10.1007/978-3-031-54517-7_19).
- [17] E. Hairer and C. Lubich, *Energy-diminishing integration of gradient systems*, IMA J. Numer. Anal., 34(2) (2014), 452–461. doi:[10.1093/imanum/drt031](https://doi.org/10.1093/imanum/drt031).
- [18] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II*, Springer Series in Computational Mathematics, Volume 14, Springer Berlin Heidelberg, Berlin, Heidelberg, 1996. doi:[10.1007/978-3-642-05221-7](https://doi.org/10.1007/978-3-642-05221-7).
- [19] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, Mathematical Modelling – Theory and Applications, Volume 23, Springer, Dordrecht, 2009. doi:[10.1007/978-1-4020-8839-1](https://doi.org/10.1007/978-1-4020-8839-1).
- [20] B. Jacob and H.J. Zwart, *Linear Port-Hamiltonian Systems on Infinite-dimensional Spaces*, Springer Basel, Basel, 2012. doi:[10.1007/978-3-0348-0399-1](https://doi.org/10.1007/978-3-0348-0399-1).
- [21] J. Lorenz, T. Zwerschke, M. Günther, and K. Schäfers, *Operator splitting for coupled linear port-Hamiltonian systems*, Appl. Math. Lett. 160 (2025), 109309. doi:[10.1016/j.aml.2024.109309](https://doi.org/10.1016/j.aml.2024.109309).
- [22] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering. Springer, New York, 2. edition, 2006. doi:[10.1007/978-0-387-40065-5](https://doi.org/10.1007/978-0-387-40065-5).

- [23] P. Schwerdtner, `Algopaul/PortHamiltonianBenchmarkSystems.jl`. URL: <https://github.com/Algopaul/PortHamiltonianBenchmarkSystems.jl>.
- [24] K. Strehmel, R. Weiner, and H. Podhaisky, *Numerik gewöhnlicher Differentialgleichungen: Nichtsteife, steife und differential-algebraische Gleichungen*, Vieweg+Teubner Verlag, Wiesbaden, 2012. doi:10.1007/978-3-8348-2263-5.
- [25] C. Totzeck, `PHScalibration`. URL: <https://github.com/ctotzeck/PHScalibration>.
- [26] A. J. van der Schaft, *Port-Hamiltonian Systems: Network Modeling and Control of Nonlinear Physical Systems*, In: H. Irschik and K. Schlacher, (eds.), *Advanced Dynamics and Control of Structures and Machines*, pages 127–167. Springer Vienna, Vienna, 2004. doi:10.1007/978-3-7091-2774-2_9.
- [27] A. van der Schaft, *L2-Gain and Passivity Techniques in Nonlinear Control*, Communications and Control Engineering, Springer International Publishing, Cham, 2017. doi:10.1007/978-3-319-49992-5.
- [28] P. Zaspel and M. Günther, *Data-driven identification of port-Hamiltonian DAE systems by Gaussian processes*, 2024. arXiv:2406.18726. doi:10.48550/arXiv.2406.18726.