IEEE CONTROL SYSTEMS LETTERS, VOL. 8, 2024

Parameter Identifiability and Reduction for Smooth and Nonsmooth Differential–Algebraic Equation Systems

Hesham Abdelfattah, *Graduate Student Member, IEEE*, Peter Stechlinski, and Sameh A. Eisa, *Member, IEEE*

***Abstract*—We extend the sensitivity rank condition (SERC), which tests for identifiability of smooth input-output systems, to a broader class of systems. Particularly, we build on our recently developed lexicographic SERC (L-SERC) theory and methods to achieve an identifiability test for differential-algebraic equation (DAE) systems for the first time, including nonsmooth systems. Additionally, we develop a method to determine the identifiable and non-identifiable parameter sets. We show how this new theory can be used to establish a (non-local) parameter reduction procedure and we show how parameter estimation problems can be solved. We apply the new methods to problems in wind turbine power systems and glucose-insulin kinetics.**

***Index Terms*—Differential algebraic systems, nonlinear systems identification, hybrid systems.**

I. Introduction

Smooth differential–algebraic equations (DAEs) are important in modeling a wide range of physical phenomena and applications [1]. Nonsmooth DAE systems appear in many fields of applications such as, but not limited to, power systems [2], process systems [3], and multibody mechanical systems [4]. We are interested in identification of a DAE system's parameters from the measurement of outputs. In particular, we are concerned with local structural identifiability in the sense of [5]: A dynamical and control input-output system is said to be locally structurally identifiable [5] at reference parameters $\boldsymbol{\theta}^*$ if, for any admissible control input $\mathbf{u}^*$ and time t and parameters $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ in a neighborhood of $\boldsymbol{\theta}^*$, $\mathbf{y}(t;\mathbf{u}^*,\boldsymbol{\theta}_1) = \mathbf{y}(t;\mathbf{u}^*,\boldsymbol{\theta}_2)$ implies that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ (i.e., the parameters are uniquely identifiable from the outputs). Methods for studying local structural identifiability of ODEs are typically computationally expensive (e.g., geometric-based methods that require successive Lie derivative computations) [5]. To overcome such limitations, the "sensitivity rank condition" (SERC) method [5], [6], [7] was introduced as a sensitivity-based method for characterizing local identifiability that is computationally effective. The authors of this letter introduced a novel nonsmooth analog of the smooth SERC test, called the lexicographic SERC (L-SERC) test in [8]. However, this new nonsmooth test, as well as the SERC test, were derived for ODE systems, not DAE systems.

Received 16 September 2024; revised 25 November 2024; accepted 11 December 2024. Date of publication 23 December 2024; date of current version 6 January 2025. This work was supported by the National Science Foundation under Award 2318772 and Award 2318773. Recommended by Senior Editor M. Guay. *(Corresponding author: Sameh A. Eisa.)*

Hesham Abdelfattah is with the Department of Aerospace Engineering and Engineering Mechanics and the Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45220 USA (e-mail: abdelfhm@mail.uc.edu).

Peter Stechlinski is with the Department of Mathematics and Statistics, University of Maine, Orono, ME 04469 USA (e-mail: peter.stechlinski@maine.edu).

Sameh A. Eisa is with the Department of Aerospace Engineering and Engineering Mechanics, University of Cincinnati, Cincinnati, OH 45220 USA (e-mail: eisash@ucmail.uc.edu).

Digital Object Identifier 10.1109/LCSYS.2024.3521427

Contributions: First, we extend the SERC and L-SERC methods to smooth and nonsmooth DAEs. Second, we provide a novel algorithm for extracting the parameters judged to be responsible for rank deficiency in the SERC and L-SERC test (i.e., non-identifiable parameters). Third, we establish a protocol for testing identifiability "non-locally" (which we refer to as repeated-sampling). Fourth, we provide a parameter reduction algorithm, based on the repeated-sampling SERC and L-SERC test mentioned above, which is useful in system design, control and optimization [9]. Lastly, we show how the sensitivities used to build the SERC and L-SERC test can be used to conduct parameter estimation.

II. Preliminaries

For use later, we briefly review the lexicographic directional derivative (LD-derivative) [10] of a lexicographically smooth (L-smooth) [11] function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ at $\mathbf{x}_0$ in the directions $\mathbf{M} = [\mathbf{m}_1 \quad \cdots \quad \mathbf{m}_k] \in \mathbb{R}^{n\times k}$ is denoted by $\mathbf{f}'(\mathbf{x}_0;\mathbf{M})$. We adopt the notation that a well-defined vertical block matrix $\begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix}$ can be written as $(\mathbf{X},\mathbf{Y})$. Examples of L-smooth functions, which are locally Lipschitz continuous, include piecewise continuously differentiable (PC^1) [12] (e.g., min, max, abs-value) and convex functions (e.g., p-norms). The lexicographic derivative (L-derivative) $\mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{x}_0;\mathbf{M})$ [11] is a computationally relevant Jacobian-like object (e.g., it can be supplied to nonsmooth optimization methods [13]), which can be obtained using the LD-derivative when $\mathbf{M}$ has full row rank (and is thus right invertible) using the following relation:

$$\underbrace{\mathbf{f}'(\mathbf{x}_0;\mathbf{M})}_{m\times k} = \underbrace{\mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{x}_0;\mathbf{M})}_{m\times n}\underbrace{\mathbf{M}}_{n\times k}. \tag{1}$$

In the case of a C^1 function, $\mathbf{f}'(\mathbf{x}_0;\mathbf{M}) = \mathbf{Jf}(\mathbf{x}_0)\mathbf{M}$, i.e., the L-derivative equals the Jacobian. The main advantage of the LD-derivative approach is that it satisfies sharp calculus rules, and therefore possesses a strong theoretical and numerical toolkit (see [14] for more details). For the absolute-value function with $\mathbf{M} = \begin{bmatrix} m_1 & m_2 & \dots & m_k \end{bmatrix} \in \mathbb{R}^{1\times k}$:

$$\text{abs}'(x_0;\mathbf{M}) = \text{fsign}(x_0, m_1, m_2, \dots, m_k)\mathbf{M}, \tag{2}$$

where fsign returns the sign of the first nonzero element, or zero otherwise. For max with $\mathbf{M} = \begin{bmatrix} \mathbf{M}_{1,\bullet} \\ \mathbf{M}_{2,\bullet} \end{bmatrix} \in \mathbb{R}^{2\times k}$:

$$\max{}'(x_0, y_0;\mathbf{M}) = \mathbf{slmax}\big(\begin{bmatrix} x_0 & \mathbf{M}_{1,\bullet}\end{bmatrix}, \begin{bmatrix} y_0 & \mathbf{M}_{2,\bullet}\end{bmatrix}\big)$$
$$:= \begin{cases} \mathbf{M}_{1,\bullet}, \text{ if } \text{fsign}\big(x_0 - y_0, \mathbf{M}_{1,\bullet} - \mathbf{M}_{2,\bullet}\big) \geq 0, \\ \mathbf{M}_{2,\bullet}, \text{ otherwise}, \end{cases}$$

i.e., the left-shifted "lexicographic maximum" of the two arguments.

III. Main Results

A. Sensitivity Theory for Input-Output DAE Systems

In this subsection, we extend the L-SERC method to nonsmooth DAEs (and thus SERC to smooth DAEs). Consider an input-output DAE system in semi-explicit form:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{w}(t), \mathbf{u}(t), \boldsymbol{\theta}), \quad \mathbf{x}(t_0) = \mathbf{f}_0(\boldsymbol{\theta}), \tag{3a}$$
$$\mathbf{0} = \mathbf{g}(\mathbf{x}(t), \mathbf{w}(t), \mathbf{v}(t), \boldsymbol{\theta}), \tag{3b}$$
$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{w}(t), \mathbf{u}(t), \mathbf{v}(t), \boldsymbol{\theta}), \tag{3c}$$

where $\mathbf{x}(t) \in \mathbb{R}^{n_x}$ are the differential variables, $\mathbf{w}(t) \in \mathbb{R}^{n_w}$ are the algebraic variables, $\mathbf{u}(t) \in \mathbb{R}^{n_u}$ and $\mathbf{v}(t) \in \mathbb{R}^{n_v}$ are the admissible control inputs in the differential and algebraic equations, respectively, $\boldsymbol{\theta} \in \mathbb{R}^{n_p}$ are the parameters, $\mathbf{y}(t) \in \mathbb{R}^{n_y}$ are the outputs, and t is the independent variable. The right-hand side (RHS) functions $\mathbf{f} : D_x \times D_w \times D_u \times \Theta \to \mathbb{R}^{n_x}$, $\mathbf{g} : D_x \times D_w \times D_v \times \Theta \to \mathbb{R}^{n_w}$ and $\mathbf{h} : D_x \times D_w \times D_u \times D_v \times \Theta \to \mathbb{R}^{n_y}$ are defined on the open and connected sets $D_x \subseteq \mathbb{R}^{n_x}$, $D_u \subseteq \mathbb{R}^{n_u}$, $D_v \subseteq \mathbb{R}^{n_v}$, $\Theta \subseteq \mathbb{R}^{n_p}$, and $\mathbf{f}_0 : \Theta \to D_x$ is the initial state function (for example, $\mathbf{f}_0(\boldsymbol{\theta}) = \mathbf{x}_0$ for a typical initial value problem). In this letter, we are concerned with the structural identifiability of the DAE system in (3).

Definition 1 [8]: The system in (3) is **locally partially identifiable** at $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$ if there exist a neighborhood $N \subseteq \boldsymbol{\Theta}$ of $\boldsymbol{\theta}^*$ and a connected set $V \subseteq \boldsymbol{\Theta}$ containing $\boldsymbol{\theta}^*$ such that for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in N \cap V$, we have that $\mathbf{y}(t;\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\theta}_1) = \mathbf{y}(t;\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\theta}_2)$, where $\mathbf{u}^*, \mathbf{v}^*$ are the reference control inputs, for all $t \in [t_0, t_f]$ if and only if $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$.

In this work, we study identifiability using a sensitivity-based tool. More precisely, we calculate the so-called lexicographic sensitivity (L-sensitivity) function using the LD-derivative of the output function. Accordingly, we need a sensitivity theory for nonsmooth DAE input-output systems that have generalized differentiation index one, which is implied by the solution of the DAE system being regular, where a solution of a nonsmooth DAE system is called regular if, roughly speaking, certain partial Clarke generalized derivative elements [15] are all nonsingular (see [16, Definitions 4.1 and 4.2]), which is analogous to the notion of classical differentiation of index one (i.e., $\det(\frac{\partial \mathbf{g}}{\partial \mathbf{w}}) \neq 0$) [1].

Proposition 1: Suppose that the RHS functions $\mathbf{f}$, $\mathbf{f}_0$, $\mathbf{g}$ and $\mathbf{h}$ in (3) are L-smooth on their respective domains. Assume that the system in (3) admits a regular solution $\mathbf{z}^*(t) := \mathbf{z}(t;\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\theta}^*) = (\mathbf{x}^*(t), \mathbf{w}^*(t))$ on $[t_0, t_f]$ through $\{(\mathbf{f}_0(\boldsymbol{\theta}^*), \mathbf{w}_0, \mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\theta}^*)\}$, for the given reference parameters $\boldsymbol{\theta}^* \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{n_p}$ and reference control inputs $\mathbf{u}^* \in \mathcal{U} := L^1([t_0, t_f], D_u)$, which is the class of Lebesgue-integrable functions mapping $[t_0, t_f]$ into D_u, and $\mathbf{v}^* \in \mathcal{V}$, which is the class of L-smooth functions mapping $[t_0, t_f]$ into D_v. Given a directions matrix $\mathbf{M} \in \mathbb{R}^{n_p \times n_k}$, the following system

$$\dot{\mathbf{X}} = \mathbf{f}'\big(\mathbf{x}^*, \mathbf{w}^*, \mathbf{u}^*, \boldsymbol{\theta}^*; (\mathbf{X}, \mathbf{W}, \mathbf{0}_u, \mathbf{M})\big), \tag{4a}$$
$$\mathbf{0}_{n_y \times n_k} = \mathbf{g}'\big(\mathbf{x}^*, \mathbf{w}^*, \mathbf{v}^*, \boldsymbol{\theta}^*; (\mathbf{X}, \mathbf{W}, \mathbf{0}_v, \mathbf{M})\big), \tag{4b}$$
$$\mathbf{Y} = \mathbf{h}'\big(\mathbf{x}^*, \mathbf{w}^*, \mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\theta}^*; (\mathbf{X}, \mathbf{W}, \mathbf{0}_u, \mathbf{0}_v, \mathbf{M})\big), \tag{4c}$$
$$\mathbf{X}(t_0) = \mathbf{f}_0'\big(\boldsymbol{\theta}^*;\mathbf{M}\big), \tag{4d}$$

where the (t) arguments are omitted and $\mathbf{0}_u := \mathbf{0}_{n_u \times n_k}$ and $\mathbf{0}_v := \mathbf{0}_{n_v \times n_k}$, is the nonsmooth forward sensitivity system associated with (3) on $[t_0, t_f]$, and is uniquely solved by the LD-derivative mappings

$$\mathbf{X}^*(t) = \big[\mathbf{x}^*\big(t;\mathbf{u}^*, \mathbf{v}^*, \cdot\big)\big]'\big(\boldsymbol{\theta}^*;\mathbf{M}\big) \in \mathbb{R}^{n_x \times n_k}, \tag{5}$$
$$\mathbf{W}^*(t) = \big[\mathbf{w}^*\big(t;\mathbf{u}^*, \mathbf{v}^*, \cdot\big)\big]'\big(\boldsymbol{\theta}^*;\mathbf{M}\big) \in \mathbb{R}^{n_w \times n_k}, \tag{6}$$
$$\mathbf{Y}^*(t) = \big[\mathbf{y}\big(t;\mathbf{u}^*, \mathbf{v}^*, \cdot\big)\big]'\big(\boldsymbol{\theta}^*;\mathbf{M}\big) \in \mathbb{R}^{n_y \times n_k}, \tag{7}$$

where $\mathbf{X}^*(t)$ is an absolutely continuous function, and $\mathbf{W}^*(t)$ and $\mathbf{Y}^*(t)$ are Lebesgue integrable functions.

Proof: Following a similar proof as in [8], note that (3a)-(3b) can be rewritten as

$$\dot{\mathbf{x}}(t) = \bar{\mathbf{f}}(t, \mathbf{x}(t), \mathbf{w}(t), \boldsymbol{\theta}), \quad \mathbf{x}(t_0) = \mathbf{f}_0(\boldsymbol{\theta}), \tag{8}$$
$$\mathbf{0}_{n_y} = \bar{\mathbf{g}}(t, \mathbf{x}(t), \mathbf{w}(t), \boldsymbol{\theta}), \tag{9}$$

with $\bar{\mathbf{f}} : (t, \mathbf{x}, \mathbf{w}, \boldsymbol{\theta}) \mapsto \mathbf{f}(\mathbf{x}, \mathbf{w}, \mathbf{u}^*(t), \boldsymbol{\theta})$ and $\bar{\mathbf{g}} : (t, \mathbf{x}, \mathbf{w}, \boldsymbol{\theta}) \mapsto \mathbf{g}(\mathbf{x}, \mathbf{w}, \mathbf{v}^*(t), \boldsymbol{\theta})$. There exists an open set V, satisfying $X := \{(\mathbf{x}^*(t), \mathbf{w}^*(t), \boldsymbol{\theta}^*) : t \in [t_0, t_f]\} \subset V \subseteq D_x \times D_w \times \Theta$, on which: (i) $\bar{\mathbf{f}}(t, \cdot, \cdot, \cdot)$ and $\bar{\mathbf{g}}(t, \cdot, \cdot, \cdot)$ are L-smooth for each $t \in [t_0, t_f]$ (from L-smoothness of $\mathbf{f}$ and $\mathbf{g}$); (ii) $\bar{\mathbf{f}}(\cdot, \mathbf{x}, \mathbf{w}, \boldsymbol{\theta})$ and $\bar{\mathbf{g}}(\cdot, \mathbf{x}, \mathbf{w}, \boldsymbol{\theta})$ are measurable for each $(\mathbf{x}, \mathbf{w}, \boldsymbol{\theta}) \in V$ (from Lebesgue integrability, and hence measurability, of $\mathbf{u}^*$, and from the continuity of $\mathbf{v}^*$); and (iii) $\mathbf{f}$ and $\mathbf{g}$ are bounded on $[t_0, t_f] \times N$ (since X is compact and V can be as small as needed). Hence, [16, Assumptions 4.1 and 4.2] are satisfied, and thus [16, Th. 4.1] may be applied to (8)-(9), which gives that $\mathbf{z}_t^* := \mathbf{z}(t;\mathbf{u}^*, \mathbf{v}^*, \cdot)$ is L-smooth on a neighborhood of $\boldsymbol{\theta}^*$ with $\mathbf{x}_t^*$ and $\mathbf{y}_t^*$ being also L-smooth, with $\mathbf{z} = (\mathbf{x}, \mathbf{w})$. Additionally, given any directions matrix $\mathbf{M} \in \mathbb{R}^{n_p \times n_k}$, the LD-derivative mappings (5)-(7) exist, with $\mathbf{X}^*(t)$ being absolutely continuous, $\mathbf{W}^*(t)$ Lebesgue integrable (from [16, Th. 4.1]), and $\mathbf{Y}^*(t)$ Lebesgue integrable (from [17, Lemma 1] noting that the Lebesgue integrable functions, $\mathbf{X}^*(t)$ and $\mathbf{W}^*(t)$, are measurable). Additionally, noting that

$$\big[\bar{\mathbf{f}}(t, \cdot)\big]'(\mathbf{z}, \boldsymbol{\theta}; (\mathbf{Z}, \mathbf{M})) = \bar{\mathbf{f}}'(t, \mathbf{z}, \boldsymbol{\theta}; (\mathbf{0}, \mathbf{Z}, \mathbf{M})) \tag{10}$$
$$\big[\bar{\mathbf{g}}(t, \cdot)\big]'(\mathbf{z}, \boldsymbol{\theta}; (\mathbf{Z}, \mathbf{M})) = \bar{\mathbf{g}}'(t, \mathbf{z}, \boldsymbol{\theta}; (\mathbf{0}, \mathbf{Z}, \mathbf{M})) \tag{11}$$

for any directions matrix $\mathbf{M}$, with $\mathbf{z} = (\mathbf{x}, \mathbf{w})$, $\mathbf{Z} = (\mathbf{X}, \mathbf{W})$, then we have from [16, Th. 4.1] that the LD-derivative mapping $t \mapsto [\mathbf{z}_t^*]'(\boldsymbol{\theta}^*;\mathbf{M})$ uniquely solves (4a) and (4b) on $[t_0, t_f]$, with initial condition given by (4d). Lastly, since $\mathbf{h}$ is L-smooth on its domain and $\mathbf{y}^*$ satisfies (3c), it follows that $\mathbf{y}_t^* := \mathbf{y}(t;\mathbf{u}^*, \mathbf{v}^*, \cdot)$ is L-smooth on a neighborhood of $\boldsymbol{\theta}^*$

as a composition of L-smooth functions and, by applying the LD-derivative chain rule as in (10), (4c) follows. ■

We define the L-SERC matrix $\boldsymbol{\Upsilon}_{\mathbf{d}}(\boldsymbol{\theta}^*) \in \mathbb{R}^{(N+1)n_y \times n_p}$ corresponding to the direction $\mathbf{d} \in \mathbb{R}^{n_p}$ by stacking output L-sensitivities at different time samples as in [8]:

$$\boldsymbol{\Upsilon}_{\mathbf{d}}(\boldsymbol{\theta}^*) := \left(\mathbf{S}^{\mathrm{L}}_{\mathbf{y}}(t_0), \mathbf{S}^{\mathrm{L}}_{\mathbf{y}}(t_1), \ldots, \mathbf{S}^{\mathrm{L}}_{\mathbf{y}}(t_N)\right), \tag{12}$$

where the output L-sensitivity functions are given by

$$\mathbf{S}^{\mathrm{L}}_{\mathbf{y}}(t) := \mathbf{Y}^*(t)\begin{bmatrix}\mathbf{d} & \mathbf{I}_{n_p}\end{bmatrix}^{-1} = \mathbf{Y}^*(t)\begin{bmatrix}\mathbf{0}_{1\times n_p} \\ \mathbf{I}_{n_p}\end{bmatrix}, \tag{13}$$

where $\mathbf{Y}^*$ solves (4c) with $\mathbf{M} = [\mathbf{d} \quad \mathbf{I}_{n_p}]$.

Definition 2 [8]: The system (3) is **L-SERC identifiable** at $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$ in the direction $\mathbf{d} \in \mathbb{R}^{n_p}$ if $\mathrm{rank}(\boldsymbol{\Upsilon}_{\mathbf{d}}) = n_p$, and **L-SERC non-identifiable** if $\mathrm{rank}(\boldsymbol{\Upsilon}_{\mathbf{d}}) < n_p$.

Hence, if (3) is L-SERC identifiable at $\boldsymbol{\theta}^*$ in the direction $\mathbf{d}$, then it is locally partially identifiable with $V = \{\alpha\mathbf{d} : \alpha \geq 0\}$ and some neighborhood N of $\boldsymbol{\theta}^*$.

Theorem 1: Assume the setting of Proposition 1. If the system in (3) is L-SERC identifiable at $\boldsymbol{\theta}^*$ in the direction $\mathbf{d} \in \mathbb{R}^{n_p}$, then it is locally partially identifiable at $\boldsymbol{\theta}^*$.

Proof: This follows immediately from the second part of the proof of [8, Th. 4.5], using the sensitivities obtained via Proposition 1 with $\mathbf{M} = \begin{bmatrix}\mathbf{d}\ \mathbf{I}_{n_p}\end{bmatrix}$. ■

Note that a system that is L-SERC non-identifiable may actually be structurally identifiable (i.e., a false negative). This type of false negative can also occur in the smooth case (see [18]). On the other hand, a system that is L-SERC identifiable is not necessarily structurally identifiable (i.e., a false positive). This type of false positive is unique to the nonsmooth case.

B. Determining Identifiable and Non-Identifiable Parameters

In order to determine which of the system's parameters are locally partially non-identifiable (i.e., those for which it is impossible to determine the parameters' values from the system's output), we introduce the following definition.

Definition 3: The parameters in $\Theta_r \subseteq \{\theta_1, \ldots, \theta_{n_p}\}$ are **locally partially non-identifiable** at $\boldsymbol{\theta}^* \in \Theta$ if there exists a neighborhood $N \subseteq \boldsymbol{\Theta}$ of $\boldsymbol{\theta}^*$ and a connected set $V \subseteq \boldsymbol{\Theta}$ containing $\boldsymbol{\theta}^*$ such that the parameters in Θ_r are locally non-identifiable in $N \cap V$.

If system (3) is L-SERC non-identifiable, then using the L-SERC matrix $\boldsymbol{\Upsilon}_{\mathbf{d}}$, we can hypothesize with confidence which are locally partially non-identifiable at $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$. Let $n_s = (N+1)n_y$, and let the SVD of $\boldsymbol{\Upsilon}_{\mathbf{d}}$ be

$$\boldsymbol{\Upsilon}_{\mathbf{d}} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}} \tag{14}$$

with orthogonal $\mathbf{U} \in \mathbb{R}^{n_s \times n_s}$, $\mathbf{V} \in \mathbb{R}^{n_p \times n_p}$ containing the left ($\mathbf{u}_{(i)}$) and right ($\mathbf{v}_{(i)}$) singular vectors, respectively, and $\Sigma \in \mathbb{R}^{n_s \times n_p}$ containing the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{n_p}$. Suppose there are n_r nonzero singular values and $n_p - n_r$ zero singular values, i.e., $\sigma_{n_r+1} = \cdots = \sigma_{n_p} = 0$. The $n_p - n_r$ columns in $\mathbf{V}$ associated with zero singular values in Σ form a basis of the right null space of $\boldsymbol{\Upsilon}_{\mathbf{d}}$, $\mathrm{Null}(\boldsymbol{\Upsilon}_{\mathbf{d}})$. Let $\mathbf{V}_r$ be the submatrix of $\mathbf{V}$ with these columns:

$$\mathbf{V}_r = \begin{bmatrix}\mathbf{v}_{(n_r+1)} & \cdots & \mathbf{v}_{(n_p)}\end{bmatrix}. \tag{15}$$

If the system in (3) is L-SERC non-identifiable at $\boldsymbol{\theta}^*$ in the direction $\mathbf{d} \in \mathbb{R}^{n_p}$, then the rank of $\mathbf{V}_r$ in (15) is equal to the number of locally partially non-identifiable parameters at $\boldsymbol{\theta}^*$. This can be seen from the construction of $\mathbf{V}_r$; since the columns of $\mathbf{V}_r$ form a basis of $\mathrm{Null}(\boldsymbol{\Upsilon}_{\mathbf{d}})$, then the rank of $\mathbf{V}_r$ determines the dimension of $\mathrm{Null}(\boldsymbol{\Upsilon}_{\mathbf{d}})$. We can categorize locally partially non-identifiable parameters into two categories. The first category of locally partially non-identifiable parameters, which we call "isolated non-identifiable" parameters are those that are non-identifiable in an isolated way independent of other parameters. The second category, which we call "pairwise non-identifiable" parameters, are for two (or more) parameters that cannot be determined from output measurements in a unique way because of how they are coupled together (e.g., if θ_1 and θ_2 always appear in the form $\theta_1\theta_2$).

We claim that a parameter $\theta_i \in \{\theta_1, \ldots, \theta_{n_p}\}$ is locally partially non-identifiable at $\boldsymbol{\theta}^*$ if column i of the row echelon form of $\mathbf{V}_r^{\mathrm{T}}$, $\mathrm{rref}(\mathbf{V}_r^{\mathrm{T}})$, is a pivot column. In addition, looking at the rows of $\mathrm{rref}(\mathbf{V}_r^{\mathrm{T}})$, if row j contains the pivot from pivot column i but no other nonzero elements, then parameter θ_i is an isolated non-identifiable. If row j contains the pivot from pivot column i and also contains other nonzero elements, e.g., in columns indexed by $\mathcal{I} \subseteq \{n_r+1, \ldots, n_p\}$, then θ_i and $\{\theta_j : j \in \mathcal{I}\}$ are pairwise non-identifiable parameters. See Example 1 for an illustration.

The DAE L-SERC test is implemented in Algorithm 1 (with MATLAB code implementation in [19]), which builds on [8, Algorithm 1] by specifying which parameters are identifiable/non-identifiable. The algorithm has three stages (marked by S1, S2, S3): The first stage, S1, is the primary probing stage in which we probe along the so-called primary directions that, to a high degree of certainty, determine the identifiability of the system. The second stage, S2, is the "twin probing" which aims to decrease the number of false positive L-SERC tests. The third stage, S3, is the singularity probing stage which aims to decrease the number of false negative L-SERC tests. In Algorithm 1, we provide as inputs the reference parameters vector, $\boldsymbol{\theta}^*$, the reference controls $\mathbf{u}^*, \mathbf{v}^*$, the set of directions, $D = \{\mathbf{d}_i\}$, scalars $\epsilon_{\mathrm{twin}}, \epsilon_{\mathrm{sing}}$ corresponding to the twin and singularity probing stages, respectively, and a natural number q corresponding to the required number of iterations for the singularity probing stage. As for the output of Algorithm 1, we return the set Θ_{lni} of parameters that are locally partially non-identifiable within some neighborhood of $\boldsymbol{\theta}^*$, as well as the set of locally partially identifiable parameters $\Theta_{li} = \{\theta_1, \ldots, \theta_{n_p}\} \setminus \Theta_{lni}$.

Remark 1: For the smooth case, i.e., if the RHS functions in (3) are C^1, the forward sensitivity system in (4) becomes the classical forward sensitivity system:

$$\begin{aligned}
\dot{\mathbf{S}}_{\mathbf{x}}(t) &= \tfrac{\partial \mathbf{f}}{\partial \mathbf{x}}\mathbf{S}_{\mathbf{x}}(t) + \tfrac{\partial \mathbf{f}}{\partial \mathbf{w}}\mathbf{S}_{\mathbf{w}}(t) + \tfrac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}, \quad \mathbf{S}_{\mathbf{x}}(t_0) = \mathbf{J}\mathbf{f}_0(\boldsymbol{\theta}^*),\\
\mathbf{0} &= \tfrac{\partial \mathbf{g}}{\partial \mathbf{x}}\mathbf{S}_{\mathbf{x}}(t) + \tfrac{\partial \mathbf{g}}{\partial \mathbf{w}}\mathbf{S}_{\mathbf{w}}(t) + \tfrac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}},\\
\mathbf{S}_{\mathbf{y}}(t) &= \tfrac{\partial \mathbf{h}}{\partial \mathbf{x}}\mathbf{S}_{\mathbf{x}}(t) + \tfrac{\partial \mathbf{h}}{\partial \mathbf{w}}\mathbf{S}_{\mathbf{w}}(t) + \tfrac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}},
\end{aligned} \tag{16}$$

with omitted partial derivative arguments, which is uniquely on $[t_0, t_f]$ by the classical forward sensitivity functions

Algorithm 1 L-SERC Algorithm

Input: $\boldsymbol{\theta}^*, \mathbf{u}^*, \mathbf{v}^*, D = \{\mathbf{d}_i\}, \epsilon_{\text{twin}}, \epsilon_{\text{sing}} \geq 0, q \in \mathbb{N}$

1: Set $D_{\text{sing}} \leftarrow \emptyset$, $\Theta_{\text{sing}} \leftarrow \emptyset$, $\Theta_{\text{lni}} \leftarrow \emptyset$
2: **for** $i = 1, 2, \ldots, |D|$ **do**
3: Compute SVD $\boldsymbol{\Upsilon}_{\mathbf{d}_i} = \mathbf{U}\Sigma\mathbf{V}^{\text{T}}$, $n_i = \text{rank}(\Sigma)$ ▷ S1
4: **if** $n_i < n_p$ **then**
5: $\mathbf{V}_r = [\mathbf{v}_{(k)\,:\,\sigma_k=0}]$, $\text{rref}(\mathbf{V}_r^{\text{T}}) = [\tilde{\mathbf{v}}_1 \ \cdots \ \tilde{\mathbf{v}}_{n_p}]$
6: Set $\Theta_{\text{lni}} \leftarrow \Theta_{\text{lni}} \cup \{\theta_j\}$: $\tilde{\mathbf{v}}_j$ is pivot column
7: **if** $\epsilon_{\text{sing}} > 0$ **then**
8: Set $D_{\text{sing}} \leftarrow D_{\text{sing}} \cup \{\mathbf{d}_i\}$
9: **if** $\epsilon_{\text{twin}} > 0$ **then** ▷ S2
10: Set $\widehat{\mathbf{d}}_i \leftarrow \mathbf{d}_i + \epsilon_{\text{twin}}\mathbf{e}_j$
11: Compute SVD $\boldsymbol{\Upsilon}_{\widehat{\mathbf{d}}_i} = \mathbf{U}\Sigma\mathbf{V}^{\text{T}}$, $\widehat{n}_i = \text{rank}(\Sigma)$
12: **if** $\widehat{n}_i < n_p$ **then**
13: $\mathbf{V}_r = [\mathbf{v}_{(k)\,:\,\sigma_k=0}]$, $\text{rref}(\mathbf{V}_r^{\text{T}}) = [\tilde{\mathbf{v}}_1 \ \cdots \ \tilde{\mathbf{v}}_{n_p}]$
14: Set $\Theta_{\text{lni}} \leftarrow \Theta_{\text{lni}} \cup \{\theta_j\}$: $\tilde{\mathbf{v}}_j$ is pivot column
15: **if** $\epsilon_{\text{sing}} > 0$ **then**
16: Set $D_{\text{sing}} \leftarrow D_{\text{sing}} \cup \{\widehat{\mathbf{d}}_i\}$
17: **if** $\epsilon_{\text{sing}} > 0$ and $q > 0$ and $D_{\text{sing}} \neq \emptyset$ **then** ▷ S3
18: **for all** $\mathbf{d}_j \in D_{\text{sing}}$ **do**
19: Compute SVD $\boldsymbol{\Upsilon}_{\mathbf{d}_j} = \mathbf{U}\Sigma\mathbf{V}^{\text{T}}$
20: Set $\boldsymbol{\theta}^*_{\pm} \leftarrow \boldsymbol{\theta}^* \pm \epsilon_{\text{sing}} \sum_{k\,:\,\sigma_k=0} \mathbf{v}_{(k)}$
21: Set $\Theta_{\text{sing}} \leftarrow \Theta_{\text{sing}} \cup \{\boldsymbol{\theta}^*_+\} \cup \{\boldsymbol{\theta}^*_-\}$
22: Set $q \leftarrow q - 1$ and **go to 1** for some $\boldsymbol{\theta}^* \in \Theta_{\text{sing}}$
23: **return** Θ_{lni}, $\Theta_{\text{li}} := \{\theta_1, \ldots, \theta_{n_p}\} \setminus \Theta_{lni}$ for each $\boldsymbol{\theta}^*$.

$$\mathbf{S}_{\mathbf{x}} = \tfrac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}}, \quad \mathbf{S}_{\mathbf{w}} = \tfrac{\partial \mathbf{w}}{\partial \boldsymbol{\theta}}, \quad \mathbf{S}_{\mathbf{y}} = \tfrac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}}. \tag{17}$$

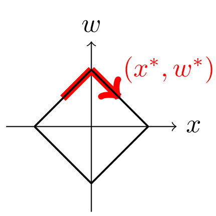
And the smooth version of the L-SERC matrix in (12) is the SERC matrix $\boldsymbol{\Upsilon}(\boldsymbol{\theta}^*) \in \mathbb{R}^{(N+1)n_y \times n_p}$, which is given by

$$\boldsymbol{\Upsilon}(\boldsymbol{\theta}^*) := \big(\mathbf{S}_{\mathbf{y}}(t_0), \mathbf{S}_{\mathbf{y}}(t_1), \ldots, \mathbf{S}_{\mathbf{y}}(t_N)\big). \tag{18}$$

Hence, the system (3) with smooth RHS functions is **SERC identifiable** at $\boldsymbol{\theta}^*$ if $\text{rank}(\boldsymbol{\Upsilon}) = n_p$, and **SERC non-identifiable** if $\text{rank}(\boldsymbol{\Upsilon}) < n_p$. Note that the algorithms presented in this work can also be applied to the smooth case (i.e., it simplifies to a SERC test for smooth DAEs).

Example 1: Consider the following nonsmooth DAE system (MATLAB code implementation in [19]):

$$\begin{aligned} \dot{x}(t) &= \max(0, \theta_1)w(t), \ \ x(0) = \theta_3\theta_4, \\ 0 &= |x(t)| + |w(t)| - \theta_2, \\ y(t) &= x(t). \end{aligned} \tag{19}$$



Consider $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^4 : \theta_2 > 0, -\theta_2 < \theta_3\theta_4 < 0\}$. Suppose that $w_0 = w(t_0) \geq 0$ is chosen as the consistent initialization (i.e., w_0 such that $|\theta_3\theta_4| + w_0 = \theta_2$) and $t_f > 0$ is chosen so that the solution $w^*(t)$ satisfies $w^*(t) > 0$ for all $t \in [t_0, t_f] = [0, t_f]$. It follows that the solution of (19) is

$$x^*(t) = \begin{cases} \gamma(t), & t \in [0, t^*], \theta_1 > 0, \\ \alpha(t - t^*), & t \in (t^*, t_f], \theta_1 > 0 \\ \theta_3\theta_4, & t \in [0, t_f], \theta_1 \leq 0, \end{cases} \tag{20}$$

from which $w^*(t) = \theta_2 - |x^*(t)|$ and $y^*(t) = x^*(t)$ can be easily obtained, with $t^* = \ln(\theta_2/(\theta_2 + \theta_3\theta_4))/\theta_1$ and

$$\gamma(t) = (\theta_3\theta_4 + \theta_2)e^{\theta_1 t} - \theta_2, \alpha(t) = (\theta_3\theta_4 - \theta_2)e^{-\theta_1 t} + \theta_2.$$

(See the figure below (19) for a typical trajectory with $\theta_1 > 0$. Note there is a loss of regularity, and thus index one, if $w^*(t) = 0$.) Given a direction $\mathbf{d} \in \mathbb{R}^4$, the sensitivity system in (4) with $\mathbf{M} = [\mathbf{d} \quad \mathbf{I}_4]$ takes the following form

$$\begin{aligned} \dot{\mathbf{X}} &= \mathbf{slmax}\big([0 \quad \mathbf{0}_{1\times 5}], [\theta_1^* \quad d_1 \quad \mathbf{e}_1^{\text{T}}]\big)w^* + \theta_1^*\mathbf{W}, \\ \mathbf{0} &= \text{fsign}(x^*, \mathbf{X})\mathbf{X} + \text{fsign}(w^*, \mathbf{W})\mathbf{W} - [d_2 \quad \mathbf{e}_2^{\text{T}}], \\ \mathbf{Y} &= \mathbf{X}, \\ \mathbf{X}(t_0) &= \theta_4^*[d_3 \quad \mathbf{e}_3^{\text{T}}] + \theta_3^*[d_4 \quad \mathbf{e}_4^{\text{T}}], \end{aligned} \tag{21}$$

with $\mathbf{e}_i \in \mathbb{R}^4$ the standard i^{th} basis vector. The solutions of (21) can be found, with $\mathbf{Y}^*(t) = \mathbf{X}^*(t)$ on $[t_0, t_f]$ given as

$$\mathbf{Y}^*(t) = \begin{cases} \Big[t(\theta_3^*\theta_4^* + \theta_2^*)e^{\theta_1^* t}, e^{\theta_1^* t} - 1, \theta_4^* e^{\theta_1^* t}, \theta_3^* e^{\theta_1^* t}\Big], & \theta_1^* > 0, \\ [0 \quad 0 \quad \theta_4^* \quad \theta_3^*], & \theta_{*1} \leq 0. \end{cases}$$

(Note that $\mathbf{W}^*(t)$ experiences a jump at t^*, because the sign of $x^*(t)$ switches from -1 to $+1$, but $\mathbf{Y}^*(t) = \mathbf{X}^*(t)$ are continuous.)

Let $\boldsymbol{\theta}^* = (0, 3, 1, -2)$. Let $\mathbf{d} = \mathbf{e}_1$, then the rank of the L-SERC matrix $\text{rank}(\boldsymbol{\Upsilon}_{\mathbf{d}}) = 2$ (with zero singular values σ_3 and σ_4). Let $\mathbf{V}_r = [\mathbf{v}_3 \ \mathbf{v}_4]$, then

$$\text{rref}(\mathbf{V}_r^{\text{T}}) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}. \tag{22}$$

We see in (22) that column 1 is a non-pivot column, hence, θ_1 is locally partially identifiable. Column 2 is a pivot column with pivot in row 1, and since row 1 has no other nonzero elements, then θ_2 is an isolated non-identifiable parameter. Column 3 is a pivot column, with pivot in row 2. Since row 2 has another nonzero element (in column 4), then θ_3 and θ_4 are pairwise non-identifiable parameters (matching expectations in (20) as θ_1 and θ_2 are multiplied by each other). Repeating the process with the same $\boldsymbol{\theta}^*$ and $\mathbf{d} = -\mathbf{e}_1$, we get $\text{rank}(\boldsymbol{\Upsilon}_{\mathbf{d}}) = 1$ (with zero singular values $\sigma_2, \sigma_3, \sigma_4$). Then $\mathbf{V}_r = [\mathbf{v}_2 \quad \mathbf{v}_3 \quad \mathbf{v}_4]$ such that

$$\text{rref}(\mathbf{V}_r^{\text{T}}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}. \tag{23}$$

In (23), we see that columns 1 and 2 are pivot columns, with no other nonzero elements in the rows where the pivots exist. Hence, θ_1 and θ_2 are isolated non-identifiable parameters. We can also see (using the same reasoning as above) that θ_3 and θ_4 are pairwise non-identifiable parameters.

Algorithm 2 Parameter Reduction Algorithm

Input: $\{\boldsymbol{\theta}^{(s)}\}$, $\mathbf{u}^*$, $\mathbf{v}^*$

1: Set $\boldsymbol{N} = \{N_{\theta_1}, \ldots, N_{\theta_{n_p}}\} \leftarrow \{0, \ldots, 0\}$
2: **for** $s = 1, \ldots, n_{smp}$ **do**
3: Perform L-SERC Algorithm 1 with $\boldsymbol{\theta}^{(s)}$, $\mathbf{u}^*$, $\mathbf{v}^*$,
4: $D = \{\pm\mathbf{e}_i\}$, $\epsilon_{\text{twin}} = \epsilon_{\text{sing}} = 0.01$, $q = 1$ to get $\Theta_{lni}^{(s)}$
5: **if** $\Theta_{lni} = \emptyset$ **then**
6: Set $\Theta_{nlni} \leftarrow \emptyset$ and go to 13
7: **for all** $\theta_j \in \Theta_{lni}^{(s)}$ **do**
8: $N_{\theta_j} = N_{\theta_j} + 1$
9: Set $\Theta_{nlni} \leftarrow \emptyset$
10: **for all** $N_{\theta_j} \in \boldsymbol{N}$ **do**
11: **if** $N_{\theta_j} = |\Theta|$ **then**
12: Set $\Theta_{nlni} \leftarrow \Theta_{nlni} \cup \{\theta_j\}$
13: **return** Θ_{nlni}, $\Theta_{nli} := \{\theta_1, \ldots, \theta_{n_p}\} \setminus \Theta_{nlni}$

C. Parameter Reduction Method

In this subsection, we introduce a parameter reduction algorithm (Algorithm 2) for the nonsmooth DAEs in (3) (and hence smooth DAEs as well). The intuition behind the parameter reduction algorithm proposed is that we use the L-SERC test introduced in Section III-A, which is a local test, to judge the identifiability of system (3) non-locally. This is done by a "repeated-sampling" test: we perform the L-SERC test at $n_{smp} \in \mathbb{N}$ different reference parameters $\boldsymbol{\theta}^{(s)} := (\theta_1^{(s)}, \ldots, \theta_{n_p}^{(s)})$, $s = 1, \ldots, n_{smp}$ sampled from a parameter space Ω to get the set of locally partially non-identifiable parameters, $\Theta_{lni}^{(s)}$, at every sample tested. A parameter $\theta_j, j = 1, \ldots, n_p$ that is judged to be locally partially non-identifiable from the L-SERC test performed at every sample, i.e., $\theta_j \in \Theta_{lni}^{(s)}, \forall s = 1, \ldots, n_{smp}$, will be added to the set of non-locally non-identifiable parameters, denoted by Θ_{nlni}. Non-locally non-identifiable parameters are "removed" from the model by fixing them to some nominal values, leading to the reduced system. Here we use a Latin hypercube scheme (LHS) to generate the set $\{\boldsymbol{\theta}^{(s)}: s = 1, \ldots, n_{smp}\}$ of reference parameters from a user-chosen region of parameter space $\Omega = \prod_{i=1}^{n_p} [\theta_i^{(L)}, \theta_i^{(U)}]$, where $\theta_i^{(L)}$ and $\theta_i^{(U)}$ are some predefined lower and upper bounds for the reference parameters that are selected based on domain knowledge of the problem, and where the reference values are chosen uniformly from these intervals $[\theta_i^{(L)}, \theta_i^{(U)}]$ (using an LHS approach).

Example 2: Recall the DAE system (19) in Example 1 and consider 25 different reference parameters $\boldsymbol{\theta}^{(s)}$ selected from the region of parameter space (MATLAB code implementation in [19])

$$\Omega = [0.1, 0.2] \times [30, 60] \times [1, 2] \times [-4, -2] \subset \Theta,$$

to get the set of reference parameters samples, $\{\boldsymbol{\theta}^{(s)}\}$ using the LHS scheme outlined above. Consider the time horizon $[t_0, t_f] = [0, 1]$ with 17 time samples t_i. Applying Algorithm 2 using $\{\boldsymbol{\theta}^{(s)}\}$ (no input for $\mathbf{u}^*$ and $\mathbf{v}^*$), we get $\Theta_{nlni} = \{\theta_3\}$ and $\Theta_{nli} = \{\theta_1, \theta_2, \theta_4\}$. Hence, the reduced version of system (3) is the DAE system with parameters θ_1, θ_2 and θ_4, with parameter θ_3 fixed to some nominal value (which we choose to equal the average of all values of θ_3 in $\{\boldsymbol{\theta}^{(s)}\}$).

D. Parameter Estimation Procedure

After performing the parameter reduction in Algorithm 2 and obtaining the set of non-identifiable parameters Θ_{nlni}, one may proceed in building the reduced DAE system in (3) by fixing the parameters in Θ_{nlni} to some nominal values (e.g., midpoint between $\theta_j^{(L)}$ and $\theta_j^{(U)}$). With the above reduction scheme in place, it is natural to next consider how to perform a parameter estimation of the reduced model, using the sensitivity information already obtained. In particular, we provide a parameter estimation scheme for the non-locally identifiable parameters $\Theta_{nli} := \{\theta_1, \ldots, \theta_{n_p}\} \setminus \Theta_{nlni}$, which form a reduced vector of parameters $\boldsymbol{\theta}_r$ composed of the parameters in Θ_{nli}:

$$\min_{\boldsymbol{\theta}_r} \phi(\boldsymbol{\theta}_r) = \sum_{i=0}^{N} \|\mathbf{y}(t_i, \boldsymbol{\theta}_r) - \mathbf{y}_i\|^2 \quad \text{subject to the reduced DAE system in (3),} \tag{24}$$

where $\mathbf{y}(t, \boldsymbol{\theta}_r)$ is the output function of the DAE system in (3), and $\mathbf{y}_i, i = 0, \ldots, N$ is a data set of measurements obtained at the sample times $t_1 \leq t_2 \leq \cdots \leq t_N$, which match the L-SERC matrix sample times in (12). Multiple nonsmooth optimization methods exist [13] which require generalized derivative information in order to successfully solve a nonsmooth dynamic optimization problem such as (24). We use the optimal control theory of [20, Th. 3.2] to obtain a Clarke generalized gradient [15].

Proposition 2: Assume the setting of Proposition 1. Then a Clarke generalized gradient $\boldsymbol{\mu}$ of ϕ at $\boldsymbol{\theta}_r^*$ is given by

$$\boldsymbol{\mu} = 2\sum_{i=0}^{N} \left(\mathbf{S}_{\mathbf{y}}^{\mathrm{L}}(t_i)\right)^{\mathrm{T}} \left(\mathbf{y}(t_i, \boldsymbol{\theta}_r^*) - \mathbf{y}_i\right). \tag{25}$$

Proof: The proof follows immediately from [20, Th. 3.2] by setting $\mathbf{p_0} = \boldsymbol{\theta}_r^*$, $n_a = n_b = 1$, $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{\theta}_r$, $\boldsymbol{\psi}_i = \boldsymbol{\xi}_i = 1$, $\hat{\mathbf{u}}(t, \boldsymbol{a}) = \hat{\mathbf{v}}(t, \boldsymbol{b}) = \boldsymbol{\theta}_r$, and adjusting for the sampling times to get $\mathbf{J}_{\mathrm{L}}\phi(\boldsymbol{\theta}_r^*; [\mathbf{d} \quad \mathbf{I}_{n_p}]) = 2\sum_{i=0}^{N}(\mathbf{S}_{\mathbf{y}}^{\mathrm{L}}(t_i))^{\mathrm{T}}(\mathbf{y}(t_i, \boldsymbol{\theta}_r^*) - \mathbf{y}_i)$, which is a Clarke generalized gradient because ϕ is scalar [11]. ■

Example 3: Recall the DAE system (19) in Example 1 and consider now estimating the "true" (but unknown) reduced parameters $\boldsymbol{\theta}_{true} = (0.150, 45.0, 1.33, -3.00)$ using the generalized derivative information from (25). Using a gradient-descent like method (MATLAB code implementation in [19]), we obtained $\boldsymbol{\theta}_{est}$, which matches $\boldsymbol{\theta}_{true}$ to eight significant digits using an initial guess $\boldsymbol{\theta}_{initial} = (0.1, 30, 1, -2)$ (which averages the values $\{\boldsymbol{\theta}^{(s)}\}$ from the LHS scheme to get $\boldsymbol{\theta}_{avg}$).

IV. Applications

Example 4 (Wind Turbine Power System): In this example, we consider a wind turbine power system under a constant wind speed (hence, the active power is constant and only the reactive power changes with time). Consider the following input-output DAE system [21, Sec. 3.2]:

$$\begin{aligned} \dot{V}_{ref} &= K_{Q_i}(Q_{cmd} - Q), \\ \dot{E}_q'' &= K_{V_i}\left(V_{ref} - V\right), \\ 0 &= V^4 - \left[2(PR + QX) + E^2\right]V^2 + \left(R^2 + X^2\right)\left(P^2 + Q^2\right) \end{aligned} \tag{26}$$

with differential variables $\mathbf{x} = (V_{ref}, E_q'')$, algebraic variable $w = V$, output $y = V$, and system parameters $\boldsymbol{\theta} =$

$(K_{Q_i}, K_{V_i}, R, X, E)$, where V is the terminal voltage, V_{ref} is the reference terminal voltage, E''_q is an equivalent voltage that controls the reactive current injection, V is the terminal voltage (at the connection between the wind turbine and the grid), K_{Q_i}, K_{V_i} are integral control gains, $Q := V\frac{E''_q - V}{X_{eq}}$ is the injected reactive power, $P = 1$ is the constant/rated active power, $Q_{cmd} = 0.6484$ is the constant reactive power command, and $X_{eq} = 0.8$ is the equivalent Norton reactance. The power grid is represented by the known infinite bus model [21, Fig. 8] with one net resistance R, net reactance X, and with E as the infinite bus voltage. The DAE system in (26) is smooth, so we use (6) to calculate its classical sensitivity system. Applying Algorithm 1 to system (26) for reference parameters $\boldsymbol{\theta}^* = (0.1, 40, 0.02, 0.02987, 1.0164)$, we obtain that system (6) is locally identifiable at $\boldsymbol{\theta}^*$. Applying Algorithm 2 to system (6), we get $\Theta_{nli} = \{K_{Q_i}, K_{V_i}, R, X, E\}$, i.e., all parameters are identifiable. Now, we estimate the loads R and X that represent best the infinite bus (grid) model based on V measurements taken at the connection between the wind turbine and the grid. We solved the parameter estimation problem in (24) to estimate the "true" (but unknown) parameter values $\boldsymbol{\theta}_{true} = (R_{true}, X_{true})$ and the results, $\boldsymbol{\theta}_{est}$, matched the values reported in [21], $(R, X) = (0.02, 0.02987)$, up to four significant digits when using an initial guess $\boldsymbol{\theta}_{initial} = (0.0202, 0.302)$. The reader is referred to [19] for the MATLAB code for this example.

Example 5 (Glucose-Insulin Kinetics): In this example, we consider a nonsmooth ODE glucose-insulin kinetics model called the Intravenous Glucose Tolerance Test (IVGTT), which measures a subject's insulin response to glucose over time for characterizing diabetes [22]:

$$\begin{aligned}\dot{G} &= p_2(G - p_9) - GX + p_{11}e^{-p_{12}t}, G(0) = p_1,\\ \dot{X} &= p_3X + p_4(I - p_{10}), X(0) = 0\\ \dot{I} &= u - p_7(I - p_{10}), I(0) = p_8 + p_{10},\end{aligned} \tag{27}$$

where G is the glucose concentration in the bloodstream, I is the insulin concentration in the bloodstream, X is the net effect of insulin on glucose disappearance, and the input $u := \text{mid}(0, \frac{\gamma(t-p_6)}{p_5-p_6}, \gamma)$ (which returns the median of its three inputs) models an external device (artificial pancreas) that infuses insulin for diabetic patients. Hence the parameters are $\boldsymbol{\theta} = (p_1, \ldots, p_{12}, \gamma)$, with values and physical interpretations given in [22, Table I]. We implemented Algorithm 2 by selecting 25 reference parameters from the region of parameter space according to the LHS scheme (MATLAB code implementation in [19]) to get the set of reference parameters samples, $\{\boldsymbol{\theta}^{(s)}\}$. We considered the time horizon $[t_0, t_f] = [0, 180]$ and applied Algorithm 2 using $\{\boldsymbol{\theta}^{(s)}\}$ to get $\Theta_{nlni} = \{p_1, p_2, p_3, p_4, p_9, p_{11}, p_{12}\}$ and $\Theta_{nli} = \{p_5, p_6, p_7, p_8, \gamma\}$, which is what we expected given the sensitivities of the system (see [22, Fig. 3(b)]).

V. Conclusion

For the first time in the literature, we provided a sensitivity-based identifiability test (L-SERC) for DAE systems, including smooth and nonsmooth ones. We offered an algorithm which extends the local nature of the L-SERC test to judge identifiability in a non-local sense. Additionally, we provided a protocol that capitalizes on identifiability to enable parameter reduction in smooth and nonsmooth DAE systems (and ODE systems, as shown in Example 5).

References

[1] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. Philadelphia, PA, USA: Soc. Ind. Appl. Math., 1995.

[2] S. A. Eisa and P. Stechlinski, "Sensitivity analysis of nonsmooth power control systems with an example of wind turbines," *Commun. Nonlin. Sci. Numer. Simul.*, vol. 95, Apr. 2021, Art. no. 105633.

[3] P. Stechlinski, M. Patrascu, and P. I. Barton, "Nonsmooth differential-algebraic equations in chemical engineering," *Comput. Chem. Eng.*, vol. 114, pp. 52–68, Jun. 2018.

[4] J.-S. Pang and J. Shen, "Strongly regular differential variational systems," *IEEE Trans. Autom. Control*, vol. 52, no. 2, pp. 242–255, Feb. 2007.

[5] L. G. van Willigenburg, J. D. Stigter, and J. Molenaar, "Sensitivity matrices as keys to local structural system properties of large-scale nonlinear systems," *Nonlin. Dyn.*, vol. 107, no. 3, pp. 2599–2618, 2022.

[6] J. D. Stigter and J. Molenaar, "A fast algorithm to assess local structural identifiability," *Automatica*, vol. 58, pp. 118–124, Aug. 2015.

[7] H. Miao, X. Xia, A. S. Perelson, and H. Wu, "On identifiability of nonlinear ODE models and applications in viral dynamics," *SIAM Rev.*, vol. 53, no. 1, pp. 3–39, 2011.

[8] P. Stechlinski, S. Eisa, and H. Abdelfattah, "Identifiability and observability of nonsmooth systems via Taylor-like approximations," 2024, *arXiv:2403.12930*.

[9] P. Benner, S. Gugercin, and K. Willcox, "A survey of projection-based model reduction methods for parametric dynamical systems," *SIAM Rev.*, vol. 57, no. 4, pp. 483–531, 2015.

[10] K. A. Khan and P. I. Barton, "A vector forward mode of automatic differentiation for generalized derivative evaluation," *Optim. Methods Softw.*, vol. 30, no. 6, pp. 1185–1212, 2015.

[11] Y. Nesterov, "Lexicographic differentiation of nonsmooth functions," *Math. Program.*, vol. 104, pp. 669–700, Nov. 2005.

[12] S. Scholtes, *Introduction to Piecewise Differentiable Equations*. New York, NY, USA: Springer, 2012.

[13] A. M. Bagirov, M. Gaudioso, N. Karmitsa, M. M. Mäkelä, and S. Taheri, Eds., *Numerical Nonsmooth Optimization: State of the Art Algorithms*. Cham, Switzerland: Springer, 2020.

[14] P. I. Barton, K. A. Khan, P. Stechlinski, and H. A. J. Watson, "Computationally relevant generalized derivatives: Theory, evaluation and applications," *Optim. Methods Softw.*, vol. 33, no. 4–6, pp. 1030–1072, 2018.

[15] F. H. Clarke, *Optimization and Nonsmooth Analysis*. Philadelphia, PA, USA: Soc. Ind. Appl. Math., 1990.

[16] P. Stechlinski and P. I. Barton, "Generalized derivatives of differential–algebraic equations," *J. Optim. Theory Appl.*, vol. 171, no. 1, pp. 1–26, 2016.

[17] A. Filippov, *Differential Equations with Discontinuous Righthand Sides: Control Systems*, vol. 18. Dordrecht, The Netherlands: Springer, 1988.

[18] L. G. van Willigenburg, J. D. Stigter, and J. Molenaar, "Extending a sensitivity based algorithm to detect local structural identifiability," *IFAC-PapersOnLine*, vol. 55, no. 20, pp. 343–348, 2022.

[19] H. Abdelfattah, P. Stechlinski, and S. Eisa. "DAE L-SERC-algorithm." GitHub. 2024. [Online]. Available: https://github.com/MDCL-UC/DAE_LSERC/

[20] P. Stechlinski and P. I. Barton, "Generalized derivatives of optimal control problems with Nonsmooth differential-algebraic equations embedded," in *Proc. 55th IEEE Conf. Decision Control*, 2016, pp. 592–597.

[21] G. Tsourakis, B. Nomikos, and C. Vournas, "Effect of wind parks with doubly fed asynchronous generators on small-signal stability," *Electr. Power Syst. Res.*, vol. 79, no. 1, pp. 190–200, 2009.

[22] M. Ackley and P. Stechlinski, "Lexicographic derivatives of nonsmooth glucose-insulin kinetics under normal and artificial pancreatic responses," *Appl. Math. Comput.*, vol. 395, Apr. 2021, Art. no. 125876.