

MIT Open Access Articles

Generalized Derivatives of Differential–Algebraic Equations

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Stechlinski, Peter G., and Paul I. Barton. “Generalized Derivatives of Differential–Algebraic Equations.” *Journal of Optimization Theory and Applications* 171.1 (2016): 1–26.

As Published: <http://dx.doi.org/10.1007/s10957-016-0988-9>

Publisher: Springer US

Persistent URL: <http://hdl.handle.net/1721.1/105419>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



Generalized Derivatives of Differential-Algebraic Equations

Peter G. Stechlinski · Paul I. Barton

Received: date / Accepted: date

Abstract Nonsmooth equation-solving and optimization algorithms which require local sensitivity information are extended to systems with nonsmooth parametric differential-algebraic equations embedded. Nonsmooth differential-algebraic equations refers here to semi-explicit differential-algebraic equations with algebraic equations satisfying local Lipschitz continuity and differential right-hand side functions satisfying Carathéodory-like conditions. Using lexicographic differentiation, an auxiliary nonsmooth differential-algebraic equation system is obtained whose unique solution furnishes the desired parametric sensitivities. More specifically, lexicographic

Peter G. Stechlinski

Process Systems Engineering Laboratory, Massachusetts Institute of Technology

Cambridge, MA, USA

pstechli@mit.edu

Paul I. Barton, Corresponding author

Process Systems Engineering Laboratory, Massachusetts Institute of Technology

Cambridge, MA, USA

pib@mit.edu

derivatives of solutions of nonsmooth parametric differential-algebraic equations are obtained. Lexicographic derivatives have been shown to be elements of the plenary hull of the (Clarke) generalized Jacobian and thus computationally relevant in the aforementioned algorithms. To accomplish this goal, the lexicographic smoothness of an extended implicit function is proved. Moreover, these generalized derivative elements can be calculated in tractable ways thanks to recent advancements in nonsmooth analysis. Forward sensitivity functions for nonsmooth parametric differential-algebraic equations are therefore characterized, extending the classical sensitivity results for smooth parametric differential-algebraic equations.

Keywords Generalized Jacobians · Sensitivity analysis · Nonsmooth analysis · Optimization · Parametric uncertainty

Mathematics Subject Classification (2000) 49J52 · 34A09 · 90C31

1 Introduction

Algorithms for nonsmooth equation-solving (e.g., semismooth Newton methods [1, 2] and LP-Newton methods [3]) and nonsmooth optimization (e.g., bundle methods for local optimization [4–6]) require sensitivity information for which many current theoretical and computational approaches are lacking. Recent progress has been made in tractable algorithms [7] for obtaining elements of a class of generalized derivative, using lexicographic differentiation [8] to calculate lexicographic directional derivatives, as introduced in [7]. Applicable to lexicographically smooth functions (which includes all differentiable functions, convex functions, and PC^1 functions in the sense

of Scholtes [9]), this approach has been used to furnish computationally relevant generalized derivatives for parametric ordinary-differential equations (ODEs) with nonsmooth right-hand sides [10]; hybrid systems, inverse functions, and implicit functions [11]; ODEs with linear programs embedded [12]; and nonsmooth optimal control problems with nonsmooth ODEs embedded [13].

With applications in mechanical, electrical, and chemical engineering, differential-algebraic equations (DAEs, also called singular or descriptor systems) have become a widely applied modeling tool [14]. Narrowing the focus more, nonsmooth DAEs provide a natural modeling framework for a number of physical phenomena found in engineering and applied mathematics such as campaign continuous pharmaceutical manufacturing (see, e.g., [15–17]). In this paper, generalized derivative notions from nonsmooth analysis are used (for background, the reader is referred to [9, 18–20] and the references therein). Elements of the plenary hull of Clarke’s generalized Jacobian comprise the desired sensitivity information for the nonsmooth algorithms described earlier. As DAEs pose a number of theoretical and numerical difficulties over ODEs (see, e.g., [14, 21–24] and the references therein), the extension of the aforementioned lexicographic differentiation theory to nonsmooth DAEs requires careful consideration.

Numerous studies have been completed on forward and adjoint sensitivities of smooth DAEs (see, e.g., [25, 26] and the references therein), hybrid and discontinuous systems (see, e.g., [27–29]), and oscillating systems [30, 31]. However, the theoretical tools and findings in these works are not applicable here due to incompatible assumptions. Clarke first derived a result on generalized Jacobians of solutions of

nonsmooth parametric ODEs [18, Theorem 7.4.1]. Pang and Stewart [32, Theorem 11, Corollary 12] proved that such generalized Jacobian supersets are linear Newton approximations (LNAs, see [19] for details) when the ODE right-hand side functions are semismooth in the sense of Qi [2]. Pang and Stewart [32] then applied their ODE sensitivity results to differential variational inequalities (DVI), as introduced in [33], with differentiable ODE right-hand side functions and differentiable variational condition functions; the authors calculated directional derivatives of local solutions of DVIs and obtained LNAs of the solution map about an initial data point. As DVIs can be expressed as a class of DAEs with specialized structure, the results in [32] are restricted to a subclass of nonsmooth DAEs with differentiable ODE right-hand side functions and nonsmooth algebraic equations. Furthermore, LNAs have been shown to not necessarily satisfy desirable properties which are satisfied by generalized Jacobians, such as LNAs of differentiable functions containing elements other than the Jacobian evaluated at said point (see [10, Example 4.2] and [12, Example 1.1]) and LNAs of convex scalar-valued functions including elements that are not subgradients [10].

Khan and Barton [10] derived a method for obtaining lexicographic derivatives of the unique solution of parametric Carathéodory ODEs from the unique solution of an auxiliary ODE system obtained via the lexicographic directional derivative chain rule [7]. The findings in [10] are a natural extension of the classical sensitivity results for smooth parametric ODE systems obtained via the classical chain rule (see, e.g., [34, Chap. V]). As a subset of the plenary Jacobian, elements of the lexicographic subdifferential have been shown to be computationally relevant in many applications

[10], including the nonsmooth algorithms detailed earlier. Moreover, as a key property of the lexicographic directional derivative is that it satisfies strict calculus rules, the implementation of a vector forward mode of automatic differentiation to calculate elements of the plenary Jacobian is therefore possible [7].

The main contribution of the current article is the development of a suitable theory for obtaining generalized derivative elements of solutions of nonsmooth parametric DAEs. In the spirit of [10], lexicographic derivatives (and therefore elements of the plenary Jacobian) of unique solutions of Carathéodory index-1 semi-explicit DAEs are obtained from the unique solution of an auxiliary nonsmooth DAE system via the lexicographic directional derivative chain rule. First, we derive the lexicographic smoothness of the extended implicit function constructed in [35] inherited from lexicographic smoothness of the participating functions. In doing so, it is possible to formulate the nonsmooth DAEs as equivalent parametric Carathéodory ODEs on an open and connected set containing the unique solution. The sensitivity theory developed here applies to DAEs for which existing methods fail and, thanks to the strict calculus rules of the lexicographic directional derivative, lays the theoretical groundwork upon which efficient numerical implementations can be designed. Methods for nonsmooth equation-solving and nonsmooth optimization are thus extended to systems with nonsmooth parametric DAEs embedded.

The rest of this article is organized as follows. Necessary background in nonsmooth analysis is presented in Section 2. Lexicographic smoothness of extended implicit functions is proved in Section 3. Generalized derivatives of nonsmooth DAEs

are calculated in Section 4. Examples are given in Section 5 and concluding remarks are provided in Section 6.

2 Preliminaries

The notational conventions here largely echo those set out in [7, 10]. The set of positive integers is denoted by \mathbb{N} and the set of nonnegative real numbers is denoted by \mathbb{R}_+ . The vector space \mathbb{R}^n is equipped with the Euclidean norm $\|\cdot\|$ and the vector space $\mathbb{R}^{m \times n}$ is equipped with the corresponding induced norm. Sets are denoted by uppercase letters (e.g., H), matrices in $\mathbb{R}^{m \times n}$ and matrix-valued functions are denoted by uppercase boldface letters (e.g., \mathbf{H}), elements of \mathbb{R} and scalar-valued functions are denoted by lowercase letters (e.g., h), and vectors in \mathbb{R}^n and vector-valued functions are denoted by lowercase boldface letters (e.g., \mathbf{h}). The zero vector in \mathbb{R}^n is denoted by $\mathbf{0}_n$, the $m \times n$ zero matrix is denoted by $\mathbf{0}_{m \times n}$, and the $n \times n$ identity matrix is denoted by \mathbf{I}_n . A well-defined vertical block matrix (or vector):

$$\begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix}$$

can be written as $(\mathbf{H}_1, \mathbf{H}_2)$. The i^{th} component of a vector \mathbf{h} is denoted by h_i . Parenthetical subscripts may be used to indicate the column vector of a matrix (e.g., the matrix \mathbf{H} has the k^{th} column $\mathbf{h}_{(k)}$), or to indicate a sequence of vectors or vector-valued functions. Parenthetical superscripts (e.g., $\mathbf{h}^{(k)}$) are used for lexicographic differentiation.

The *open and closed balls of radius $r > 0$ centered at $\mathbf{h} \in \mathbb{R}^n$* are denoted by $B_r(\mathbf{h})$ and $\bar{B}_r(\mathbf{h})$, respectively. A *neighborhood of $\mathbf{h} \in \mathbb{R}^n$* is a set of points $B_\delta(\mathbf{h})$ for

some $\delta > 0$. A *neighborhood of $H \subset \mathbb{R}^n$* is given by $B_\delta(H) := \cup_{\mathbf{h} \in H} B_\delta(\mathbf{h})$ for some $\delta > 0$. Given a set $H \subset \mathbb{R}^n$, its convex hull is denoted by $\text{conv}H$. A set of matrices $H \subset \mathbb{R}^{n \times n}$ is said to be of *maximal rank* if it contains no singular matrices. Given $n_x, n_y, n_z \in \mathbb{N}$ and $W \subset \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_z}$, the *projections of W onto \mathbb{R}^{n_x} and $\mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$* are given by, respectively,

$$\begin{aligned}\pi_x W &:= \{\boldsymbol{\eta}_x \in \mathbb{R}^{n_x} : \exists (\boldsymbol{\eta}_x, \boldsymbol{\eta}_y, \boldsymbol{\eta}_z) \in W\} \subset \mathbb{R}^{n_x}, \\ \pi_{x,y} W &:= \{(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} : \exists (\boldsymbol{\eta}_x, \boldsymbol{\eta}_y, \boldsymbol{\eta}_z) \in W\} \subset \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}.\end{aligned}$$

The *shadows of W at $\mathbf{y} \in \pi_y W$ with respect to \mathbb{R}^{n_x} and $\mathbb{R}^{n_x} \times \mathbb{R}^{n_z}$* are given by, respectively,

$$\begin{aligned}\pi_x(W; \mathbf{y}) &:= \pi_x \{(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y, \boldsymbol{\eta}_z) \in W : \boldsymbol{\eta}_y = \mathbf{y}\} \subset \mathbb{R}^{n_x}, \\ \pi_{x,z}(W; \mathbf{y}) &:= \pi_{x,z} \{(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y, \boldsymbol{\eta}_z) \in W : \boldsymbol{\eta}_y = \mathbf{y}\} \subset \mathbb{R}^{n_x} \times \mathbb{R}^{n_z}.\end{aligned}$$

The *shadow of W at $(\mathbf{x}, \mathbf{y}) \in \pi_{x,y} W$ with respect to \mathbb{R}^{n_z}* is given by

$$\pi_z(W; (\mathbf{x}, \mathbf{y})) := \pi_z \{(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y, \boldsymbol{\eta}_z) \in W : (\boldsymbol{\eta}_x, \boldsymbol{\eta}_y) = (\mathbf{x}, \mathbf{y})\} \subset \mathbb{R}^{n_z}.$$

Given $n_q \in \mathbb{N}$, $W_x \subset \pi_x W$, $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in W$, and $\mathbf{f} : W \rightarrow \mathbb{R}^{n_q}$, the *cross-section of \mathbf{f} at $\mathbf{x} \in \pi_x W$* is given by

$$\mathbf{f}_{\mathbf{x}} : \pi_{y,z}(W; \mathbf{x}) \rightarrow \mathbb{R}^{n_q} : (\boldsymbol{\eta}_y, \boldsymbol{\eta}_z) \mapsto \mathbf{f}(\mathbf{x}, \boldsymbol{\eta}_y, \boldsymbol{\eta}_z).$$

The *W_x -blind cross-section of \mathbf{f} at \mathbf{x}* is given by

$$\mathbf{f}_{\mathbf{x} \setminus W_x} : \pi_{y,z}(W; \mathbf{x}) \rightarrow \mathbb{R}^{n_q} : (\boldsymbol{\eta}_y, \boldsymbol{\eta}_z) \mapsto \begin{cases} \mathbf{f}(\mathbf{x}, \boldsymbol{\eta}_y, \boldsymbol{\eta}_z), & \mathbf{x} \in \pi_x W \setminus W_x, \\ \mathbf{0}_{n_q}, & \mathbf{x} \in W_x. \end{cases}$$

The other non-vacuous projections, shadows, cross-sections and blind cross-sections are defined similarly.

2.1 Generalized Derivatives

Let $X \subset \mathbb{R}^n$ be an open set and $\mathbf{f} : X \rightarrow \mathbb{R}^m$. Given that \mathbf{f} is (Fréchet) differentiable at $\mathbf{x} \in X$, its (Fréchet) derivative is called the *Jacobian matrix* and is denoted by $\mathbf{Jf}(\mathbf{x}) \in \mathbb{R}^{m \times n}$. The function \mathbf{f} is said to be *differentiable on X* if \mathbf{f} is differentiable at each point $\mathbf{x} \in X$. The function \mathbf{f} is said to be *continuously differentiable (C^1) at $\mathbf{x} \in X$* if \mathbf{f} is differentiable on a neighborhood $N(\mathbf{x}) \subset X$ of \mathbf{x} and $\mathbf{Jf} : N(\mathbf{x}) \rightarrow \mathbb{R}^{m \times n}$ is continuous on $N(\mathbf{x})$. The function \mathbf{f} is said to be *C^1 on X* if \mathbf{f} is C^1 at each point $\mathbf{x} \in X$. As defined by Scholtes [9], \mathbf{f} is said to be *piecewise differentiable (PC^1) at $\mathbf{x} \in X$* if there exist a neighborhood $N(\mathbf{x}) \subset X$ of \mathbf{x} and a finite collection of C^1 functions on $N(\mathbf{x})$, $\{\mathbf{f}_{(1)}, \dots, \mathbf{f}_{(k)}\}$, such that \mathbf{f} is continuous on $N(\mathbf{x})$ and

$$\mathbf{f}(\boldsymbol{\eta}) \in \{\mathbf{f}_{(i)}(\boldsymbol{\eta}) : i \in \{1, \dots, k\}\}, \quad \forall \boldsymbol{\eta} \in N(\mathbf{x}).$$

\mathbf{f} is said to be *PC^1 on X* if \mathbf{f} is PC^1 at each point $\mathbf{x} \in X$.

Let \mathbf{f} be locally Lipschitz continuous on X . It follows that \mathbf{f} is differentiable at each point $\mathbf{x} \in X \setminus Z_{\mathbf{f}}$, where $Z_{\mathbf{f}} \subset X$ has zero (Lebesgue) measure, by Rademacher's Theorem. Clarke [18] established the following definitions and results concerning generalized derivatives. The *B -subdifferential* of \mathbf{f} at $\mathbf{x} \in X$ is defined as

$$\partial_B \mathbf{f}(\mathbf{x}) := \left\{ \lim_{i \rightarrow \infty} \mathbf{Jf}(\mathbf{x}_{(i)}) : \lim_{i \rightarrow \infty} \mathbf{x}_{(i)} = \mathbf{x}, \quad \mathbf{x}_{(i)} \in X \setminus Z_{\mathbf{f}}, \forall i \in \mathbb{N} \right\}.$$

The Clarke (generalized) *Jacobian* of \mathbf{f} at $\mathbf{x} \in X$ is defined as

$$\partial \mathbf{f}(\mathbf{x}) := \text{conv } \partial_B \mathbf{f}(\mathbf{x}).$$

For a point $\mathbf{x} \in X$, $\partial_{\mathbf{B}}\mathbf{f}(\mathbf{x})$ is necessarily nonempty and compact, while $\partial\mathbf{f}(\mathbf{x})$ is necessarily nonempty, compact, and convex. If \mathbf{f} is differentiable at \mathbf{x} then $\mathbf{J}\mathbf{f}(\mathbf{x}) \in \partial\mathbf{f}(\mathbf{x})$. If \mathbf{f} is C^1 at \mathbf{x} then $\partial\mathbf{f}(\mathbf{x}) = \partial_{\mathbf{B}}\mathbf{f}(\mathbf{x}) = \{\mathbf{J}\mathbf{f}(\mathbf{x})\}$.

The *plenary Jacobian* of \mathbf{f} at $\mathbf{x} \in X$ [36] is defined as

$$\partial_{\mathbf{P}}\mathbf{f}(\mathbf{x}) := \{\mathbf{M} \in \mathbb{R}^{m \times n} : \forall \mathbf{d} \in \mathbb{R}^n, \exists \mathbf{H} \in \partial\mathbf{f}(\mathbf{x}) \text{ s.t. } \mathbf{M}\mathbf{d} = \mathbf{H}\mathbf{d}\}.$$

As the name suggests, the plenary Jacobian of \mathbf{f} at \mathbf{x} is the plenary hull of its Clarke Jacobian at \mathbf{x} (see [36] for details on plenary sets and plenary hulls); it is the intersection of all plenary supersets of $\partial\mathbf{f}(\mathbf{x})$, which includes all linear transformations for which images are indistinguishable. As demonstrated by Imbert [37], $\partial_{\mathbf{P}}\mathbf{f}(\mathbf{x})$ is nonempty, compact, convex, and satisfies

$$\partial\mathbf{f}(\mathbf{x}) \subset \partial_{\mathbf{P}}\mathbf{f}(\mathbf{x}) \subset \prod_{i=1}^m \partial f_i(\mathbf{x}).$$

As pointed out in [10], if $\min\{m, n\} = 1$ then $\partial\mathbf{f}(\mathbf{x}) = \partial_{\mathbf{P}}\mathbf{f}(\mathbf{x})$. Moreover, if $m = n$ and if $\partial\mathbf{f}(\mathbf{x})$ is of maximal rank then a similar relationship holds between images of inverses of elements of $\partial\mathbf{f}(\mathbf{x})$ and $\partial_{\mathbf{P}}\mathbf{f}(\mathbf{x})$:

$$\{\mathbf{H}^{-1}\mathbf{d} \in \mathbb{R}^n : \mathbf{H} \in \partial_{\mathbf{P}}\mathbf{f}(\mathbf{x})\} = \{\mathbf{H}^{-1}\mathbf{d} \in \mathbb{R}^n : \mathbf{H} \in \partial\mathbf{f}(\mathbf{x})\}, \quad \forall \mathbf{d} \in \mathbb{R}^n.$$

As a consequence of these observations, elements of the plenary Jacobian are no less useful than elements of the Clarke Jacobian in any of the following: bundle methods for finding local minima for nonsmooth nonlinear programs (since the objective function is scalar-valued), semismooth Newton methods, Clarke's mean value theorem [18, Proposition 2.6.5], and Clarke's inverse function theorem [18, Theorem 7.1.1].

Given $n_x, n_y, n_z, n_q \in \mathbb{N}$, $W \subset \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_z}$ open, and $\mathbf{g} : W \rightarrow \mathbb{R}^{n_q}$ Lipschitz continuous on a neighborhood of $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in W$, the Clarke (generalized) *Jacobian projections* of \mathbf{g} at $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ are defined as

$$\pi_1 \partial \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{z}) := \left\{ \mathbf{M} \in \mathbb{R}^{n_q \times n_x} : \exists [\mathbf{M} \quad \mathbf{N}_1 \quad \mathbf{N}_2] \in \partial \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \right\},$$

$$\pi_2 \partial \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{z}) := \left\{ \mathbf{M} \in \mathbb{R}^{n_q \times n_y} : \exists [\mathbf{N}_1 \quad \mathbf{M} \quad \mathbf{N}_2] \in \partial \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \right\},$$

$$\pi_{2,3} \partial \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{z}) := \left\{ [\mathbf{M}_1 \quad \mathbf{M}_2] \in \mathbb{R}^{n_q \times (n_y + n_z)} : \exists [\mathbf{N} \quad \mathbf{M}_1 \quad \mathbf{M}_2] \in \partial \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \right\},$$

with $\pi_3 \partial \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{z})$, $\pi_{1,2} \partial \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{z})$, and $\pi_{1,3} \partial \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{z})$ defined similarly. If \mathbf{g} is C^1 at $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ then

$$\pi_{2,3} \partial \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \left\{ \left[\begin{array}{c} \frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \\ \frac{\partial \mathbf{g}}{\partial \mathbf{z}}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \end{array} \right] \right\}.$$

2.2 Lexicographic Differentiation

Nesterov [8] introduced lexicographically smooth functions and the lexicographic (generalized) derivative. Given $X \subset \mathbb{R}^n$ open and $\mathbf{f} : X \rightarrow \mathbb{R}^m$, the *directional derivative* of \mathbf{f} at $\mathbf{x} \in X$ in the direction $\mathbf{d} \in \mathbb{R}^n$ is given by

$$\mathbf{f}'(\mathbf{x}; \mathbf{d}) := \lim_{\alpha \downarrow 0} \frac{\mathbf{f}(\mathbf{x} + \alpha \mathbf{d}) - \mathbf{f}(\mathbf{x})}{\alpha},$$

if it exists. The function \mathbf{f} is said to be *directionally differentiable at \mathbf{x}* if $\mathbf{f}'(\mathbf{x}; \mathbf{d})$ exists and is finite for all $\mathbf{d} \in \mathbb{R}^n$. Given that \mathbf{f} is locally Lipschitz continuous on X , \mathbf{f} is said to be *lexicographically smooth (L-smooth) at $\mathbf{x} \in X$* if for any $k \in \mathbb{N}$ and any $\mathbf{M} := [\mathbf{m}_{(1)} \cdots \mathbf{m}_{(k)}] \in \mathbb{R}^{n \times k}$, the following higher-order directional derivatives are

well-defined:

$$\begin{aligned}
\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(0)} &: \mathbb{R}^n \rightarrow \mathbb{R}^m : \mathbf{d} \mapsto \mathbf{f}'(\mathbf{x}; \mathbf{d}), \\
\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(1)} &: \mathbb{R}^n \rightarrow \mathbb{R}^m : \mathbf{d} \mapsto [\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(0)}]'(\mathbf{m}_{(1)}; \mathbf{d}), \\
\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(2)} &: \mathbb{R}^n \rightarrow \mathbb{R}^m : \mathbf{d} \mapsto [\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(1)}]'(\mathbf{m}_{(2)}; \mathbf{d}), \\
&\vdots \\
\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(k)} &: \mathbb{R}^n \rightarrow \mathbb{R}^m : \mathbf{d} \mapsto [\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(k-1)}]'(\mathbf{m}_{(k)}; \mathbf{d}).
\end{aligned}$$

The function \mathbf{f} is said to be *lexicographically smooth (L-smooth) on X* if it is L-smooth at each point $\mathbf{x} \in X$. The class of L-smooth functions is closed under composition, and includes all C^1 functions, convex functions [8], and PC^1 functions [7] in the sense of Scholtes [9]. Given any nonsingular matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$ and $\mathbf{f} : X \rightarrow \mathbb{R}^m$ L-smooth at $\mathbf{x} \in X$, the mapping $\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(n)} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is linear and the *lexicographic (L-)derivative* of \mathbf{f} at \mathbf{x} in the directions \mathbf{M} is

$$\mathbf{J}_L \mathbf{f}(\mathbf{x}; \mathbf{M}) := \mathbf{J} \mathbf{f}_{\mathbf{x},\mathbf{M}}^{(n)}(\mathbf{0}_n) \in \mathbb{R}^{m \times n}.$$

The *lexicographic subdifferential* of \mathbf{f} at \mathbf{x} is defined as

$$\partial_L \mathbf{f}(\mathbf{x}) := \{ \mathbf{J}_L \mathbf{f}(\mathbf{x}; \mathbf{N}) : \mathbf{N} \in \mathbb{R}^{n \times n}, \det \mathbf{N} \neq 0 \}.$$

If \mathbf{f} is differentiable at \mathbf{x} then $\partial_L \mathbf{f}(\mathbf{x}) = \{ \mathbf{J} \mathbf{f}(\mathbf{x}) \}$ and if $m = 1$ then $\partial_L f(\mathbf{x}) \subset \partial f(\mathbf{x})$.

The lexicographic directional derivative was introduced by Khan and Barton [7]: given any $k \in \mathbb{N}$, any $\mathbf{M} := [\mathbf{m}_{(1)} \cdots \mathbf{m}_{(k)}] \in \mathbb{R}^{n \times k}$, and $\mathbf{f} : X \rightarrow \mathbb{R}^m$ L-smooth at $\mathbf{x} \in X$, the *lexicographic directional (LD-)derivative* of \mathbf{f} at \mathbf{x} in the directions \mathbf{M} is defined as

$$\mathbf{f}'(\mathbf{x}; \mathbf{M}) := \left[\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(0)}(\mathbf{m}_{(1)}) \quad \mathbf{f}_{\mathbf{x},\mathbf{M}}^{(1)}(\mathbf{m}_{(2)}) \quad \cdots \quad \mathbf{f}_{\mathbf{x},\mathbf{M}}^{(k-1)}(\mathbf{m}_{(k)}) \right].$$

Note that $\mathbf{f}'(\mathbf{x};\mathbf{M})$ is uniquely defined for all $\mathbf{M} \in \mathbb{R}^{n \times k}$ and all $k \in \mathbb{N}$. The LD-derivative adopts its name because if \mathbf{M} is square and nonsingular then it follows that $\mathbf{f}'(\mathbf{x};\mathbf{M}) = \mathbf{J}_L \mathbf{f}(\mathbf{x};\mathbf{M})\mathbf{M}$. If \mathbf{f} is differentiable at \mathbf{x} then $\mathbf{f}'(\mathbf{x};\mathbf{M}) = \mathbf{J}\mathbf{f}(\mathbf{x})\mathbf{M}$. If \mathbf{M} has one column, the LD-derivative is equivalent to the directional derivative. Unlike the generalized Jacobian, the LD-derivative obeys a strict chain rule [7].

Theorem 2.1 *Let $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$ be open and $\mathbf{h} : X \rightarrow Y$ and $\mathbf{g} : Y \rightarrow \mathbb{R}^q$ be locally Lipschitz functions on X and Y , respectively. Let \mathbf{h} and \mathbf{g} be L -smooth at $\mathbf{x} \in X$ and $\mathbf{h}(\mathbf{x}) \in Y$, respectively. Then the composition $\mathbf{g} \circ \mathbf{h}$ is L -smooth at \mathbf{x} ; for any $k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n \times k}$, the chain rule for LD-derivatives is given as:*

$$[\mathbf{g} \circ \mathbf{h}]'(\mathbf{x};\mathbf{M}) = \mathbf{g}'(\mathbf{h}(\mathbf{x});\mathbf{h}'(\mathbf{x};\mathbf{M})). \quad (1)$$

Theorem 2.1 reduces to Nesterov's chain rule [8, Theorem 5] when the matrix \mathbf{M} is square and nonsingular, and reduces to the classical chain rule when \mathbf{g} and \mathbf{h} are both differentiable. Significantly, the strict chain rule of Theorem 2.1 allows for the development of a vector forward mode of automatic differentiation to calculate LD-derivatives [7].

2.3 Generalized Derivatives of Ordinary Differential Equations

Given an open set $X \subset \mathbb{R}^n$ and $\mathbf{f} : X \rightarrow \mathbb{R}^m$ that is L -smooth at $\mathbf{x} \in X$, $\partial_L \mathbf{f}(\mathbf{x}) \subset \partial_P \mathbf{f}(\mathbf{x})$ [10]. If \mathbf{f} is PC^1 at \mathbf{x} then \mathbf{f} is L -smooth at \mathbf{x} and $\partial_L \mathbf{f}(\mathbf{x}) \subset \partial_B \mathbf{f}(\mathbf{x})$ [7]. Prompted by these relations and the discussions in Section 2.1 on the usefulness of elements of the plenary Jacobian, obtaining an element of $\partial_L \mathbf{f}(\mathbf{x})$ is therefore just as useful as an element of the Clarke Jacobian in a variety of applications, and can be furnished

via computing $\mathbf{f}'(\mathbf{x}; \mathbf{M})$ for a square and nonsingular matrix \mathbf{M} and solving the linear equation system $\mathbf{f}'(\mathbf{x}; \mathbf{M}) = \mathbf{J}_L \mathbf{f}(\mathbf{x}; \mathbf{M}) \mathbf{M}$ [7]. Motivated by these observations, Khan and Barton found LD-derivatives of nonsmooth parametric ODEs [10, Theorem 4.2], which is restated here for parametric ODEs whose right-hand side functions depend explicitly on parameters by virtue of its proof and the remarks following Example 4.2 in [10].

Theorem 2.2 *Let $n_p, n_x \in \mathbb{N}$, $n_t = 1$, $D \subset \mathbb{R}^{n_t} \times \mathbb{R}^{n_p} \times \mathbb{R}^{n_x}$ be open and connected, $t_0, t_f \in \pi_t D$ satisfy $t_0 < t_f$, and $Z_{\mathbf{f}}$ be a zero-measure subset of $[t_0, t_f]$. Let $\mathbf{f} : D \rightarrow \mathbb{R}^{n_x}$ and $\mathbf{f}_0 : \pi_p D \rightarrow \pi_x D$. Assume that the following conditions are satisfied:*

- (i) $\mathbf{f}(\cdot, \mathbf{p}, \boldsymbol{\eta})$ is measurable on $[t_0, t_f]$ for each $(\mathbf{p}, \boldsymbol{\eta}) \in \pi_{p,x} D$;
- (ii) $\mathbf{f}(t, \cdot, \cdot)$ is L -smooth on $\pi_{p,x}(D; t)$ for each $t \in [t_0, t_f] \setminus Z_{\mathbf{f}}$;
- (iii) there exist Lebesgue integrable functions $k_{\mathbf{f}}, m_{\mathbf{f}} : [t_0, t_f] \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ for which:
 - (a) $\|\mathbf{f}(t, \mathbf{p}, \boldsymbol{\eta})\| \leq m_{\mathbf{f}}(t)$, $\forall t \in [t_0, t_f]$, $\forall (\mathbf{p}, \boldsymbol{\eta}) \in \pi_{p,x}(D; t)$;
 - (b) $\|\mathbf{f}(t, \mathbf{p}_1, \boldsymbol{\eta}_1) - \mathbf{f}(t, \mathbf{p}_2, \boldsymbol{\eta}_2)\| \leq k_{\mathbf{f}}(t) \|(\mathbf{p}_1, \boldsymbol{\eta}_1) - (\mathbf{p}_2, \boldsymbol{\eta}_2)\|$, $\forall t \in [t_0, t_f]$,
 $\forall (\mathbf{p}_1, \boldsymbol{\eta}_1), (\mathbf{p}_2, \boldsymbol{\eta}_2) \in \pi_{p,x}(D; t)$;
- (iv) \mathbf{f}_0 is L -smooth on $\pi_p D$;
- (v) for some $\mathbf{p}_0 \in \pi_p D$, there exists a solution $\mathbf{x}(\cdot, \mathbf{p}_0)$ of the following parametric ODE system at $\mathbf{p} := \mathbf{p}_0$:

$$\dot{\mathbf{x}}(t, \mathbf{p}) = \mathbf{f}(t, \mathbf{x}(t, \mathbf{p})), \quad \text{a.e. } t \in [t_0, t_f],$$

$$\mathbf{x}(t_0, \mathbf{p}) = \mathbf{f}_0(\mathbf{p}),$$

which satisfies $\{(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\} \subset D$. Then, for each $t \in [t_0, t_f]$, the mapping $\mathbf{x}_t \equiv \mathbf{x}(t, \cdot)$ is Lipschitz continuous on a neighborhood of \mathbf{p}_0 , with a Lipschitz

constant that is independent of t . Moreover, \mathbf{x}_t is L -smooth at \mathbf{p}_0 ; for any $k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n_p \times k}$, the LD-derivative mapping $\tilde{\mathbf{X}} : [t_0, t_f] \rightarrow \mathbb{R}^{n_x \times k} : t \mapsto [\mathbf{x}_t]'$ ($\mathbf{p}_0; \mathbf{M}$) is the unique solution on $[t_0, t_f]$ of the following ODE system:

$$\begin{aligned} \dot{\mathbf{X}}(t) &= [\mathbf{f}_{t \setminus Z_t}]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, \mathbf{X}(t))), \\ \mathbf{X}(t_0) &= [\mathbf{f}_0]'(\mathbf{p}_0; \mathbf{M}). \end{aligned} \tag{2}$$

Remark 2.1 The right-hand side function $(t, \mathbf{A}) \mapsto [\mathbf{f}_{t \setminus Z_t}]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, \mathbf{A}))$ in (2) is measurable with respect to t but not necessarily continuous with respect to \mathbf{A} at almost every $t \in [t_0, t_f]$ (see [10, Example 4.1]). However, the columns of (2) can be decoupled to yield a sequence of k Carathéodory ODEs [10, Corollary 4.2]. Consequently, the k columns of the matrix-valued function $t \mapsto [\mathbf{x}_t]'$ ($\mathbf{p}_0; \mathbf{M}$) are absolutely continuous vector-valued functions mapping $[t_0, t_f]$ to \mathbb{R}^{n_x} .

3 Lexicographic Smoothness of Extended Implicit Functions

Clarke provided local inverse and implicit function theorems for locally Lipschitz continuous functions [18], but without generalized derivative descriptions for said nonsmooth inverse and implicit functions. Levy and Mordukhovich [38] derived an implicit function theorem for coderivatives. Extending the results of Scholtes [9, Theorem 3.2.3] concerning directional derivative information, Khan and Barton [11] established results on the lexicographic smoothness of local inverse and implicit functions and their corresponding LD-derivatives. For congruence with the present article, the L -smooth implicit function result in [11] is restated with a stricter sufficient condition concerning projections of Clarke Jacobians (see the discussion following Theorem 2 in [11]).

Theorem 3.1 *Let $W \subset \mathbb{R}^n \times \mathbb{R}^m$ be open. Suppose that $\mathbf{g} : W \rightarrow \mathbb{R}^m$ is L -smooth at $(\mathbf{x}^*, \mathbf{y}^*) \in W$, $\mathbf{g}(\mathbf{x}^*, \mathbf{y}^*) = \mathbf{0}_m$, and $\pi_2 \partial \mathbf{g}(\mathbf{x}^*, \mathbf{y}^*)$ is of maximal rank. Then there exist neighborhoods $N(\mathbf{x}^*) \subset \pi_x W$ and $N(\mathbf{x}^*, \mathbf{y}^*) \subset W$ of \mathbf{x}^* and $(\mathbf{x}^*, \mathbf{y}^*)$, respectively, and a function $\mathbf{r} : N(\mathbf{x}^*) \rightarrow \mathbb{R}^m$ that is Lipschitz continuous on $N(\mathbf{x}^*)$ such that, for each $\mathbf{x} \in N(\mathbf{x}^*)$, $(\mathbf{x}, \mathbf{r}(\mathbf{x}))$ is the unique vector in $N(\mathbf{x}^*, \mathbf{y}^*)$ satisfying $\mathbf{g}(\mathbf{x}, \mathbf{r}(\mathbf{x})) = \mathbf{0}_m$. Moreover, \mathbf{r} is L -smooth at \mathbf{x}^* ; for any $k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n \times k}$, the LD-derivative $\mathbf{r}'(\mathbf{x}^*; \mathbf{M})$ is the unique solution $\mathbf{N} \in \mathbb{R}^{m \times k}$ of the equation system*

$$\mathbf{g}'(\mathbf{x}^*, \mathbf{y}^*; (\mathbf{M}, \mathbf{N})) = \mathbf{0}_{m \times k}. \quad (3)$$

In [35], an extended implicit function theorem was provided for locally Lipschitz continuous functions. The lexicographic smoothness of such an extended implicit function is detailed in the next result.

Theorem 3.2 *Let $W \subset \mathbb{R}^n \times \mathbb{R}^m$ be open and $\Omega \subset W$ be a compact set such that each $\mathbf{x} \in \pi_x \Omega$ is the projection of only one point $(\mathbf{x}, \mathbf{y}) \in \Omega$. Suppose that $\mathbf{g} : W \rightarrow \mathbb{R}^m$ is L -smooth on W , $\mathbf{g}(\Omega) = \{\mathbf{0}_m\}$, and $\pi_2 \partial \mathbf{g}(\mathbf{x}, \mathbf{y})$ is of maximal rank for each $(\mathbf{x}, \mathbf{y}) \in \Omega$. Then there exist $\delta, \rho > 0$ and a function $\mathbf{r} : B_\delta(\pi_x \Omega) \subset \pi_x W \rightarrow \mathbb{R}^m$ that is Lipschitz continuous and L -smooth on $B_\delta(\pi_x \Omega)$ such that $\pi_2 \partial \mathbf{g}(\mathbf{x}, \mathbf{y})$ is of maximal rank for all $(\mathbf{x}, \mathbf{y}) \in B_\rho(\Omega) \subset W$ and, for each $\mathbf{x} \in B_\delta(\pi_x \Omega)$, $(\mathbf{x}, \mathbf{r}(\mathbf{x}))$ is the unique vector in $B_\rho(\Omega)$ satisfying $\mathbf{g}(\mathbf{x}, \mathbf{r}(\mathbf{x})) = \mathbf{0}_m$. Moreover, for any $\mathbf{x} \in B_\delta(\pi_x \Omega)$, any $k \in \mathbb{N}$, and any $\mathbf{M} \in \mathbb{R}^{n \times k}$, $\mathbf{r}'(\mathbf{x}; \mathbf{M})$ is the unique solution $\mathbf{N} \in \mathbb{R}^{m \times k}$ of the equation system*

$$\mathbf{g}'(\mathbf{x}, \mathbf{r}(\mathbf{x}); (\mathbf{M}, \mathbf{N})) = \mathbf{0}_{m \times k}. \quad (4)$$

Proof By [35, Theorem 3.6], there exist $\delta_1, \rho_1 > 0$ and a function

$$\mathbf{r}_1 : B_{\delta_1}(\pi_x \Omega) \subset \pi_x W \rightarrow \mathbb{R}^m$$

that is Lipschitz continuous on $B_{\delta_1}(\pi_x \Omega)$ such that $\pi_2 \partial \mathbf{g}(\mathbf{x}, \mathbf{y})$ is of maximal rank for all $(\mathbf{x}, \mathbf{y}) \in B_{\rho_1}(\Omega) \subset W$ and, for each $\mathbf{x} \in B_{\delta_1}(\pi_x \Omega)$, $(\mathbf{x}, \mathbf{r}_1(\mathbf{x}))$ is the unique vector in $B_{\rho_1}(\Omega)$ satisfying $\mathbf{g}(\mathbf{x}, \mathbf{r}_1(\mathbf{x})) = \mathbf{0}_m$.

Let $\tilde{\Omega} := \{(\mathbf{x}, \mathbf{r}_1(\mathbf{x})) : \mathbf{x} \in \bar{B}_{0.5\delta_1}(\pi_x \Omega)\} \subset B_{\rho_1}(\Omega) \subset W$, which is a compact set such that each point $\mathbf{x} \in \pi_x \tilde{\Omega} = \bar{B}_{0.5\delta_1}(\pi_x \Omega) \subset B_{\delta_1}(\pi_x \Omega)$ is the projection of only one point in $\tilde{\Omega}$ (namely, $(\mathbf{x}, \mathbf{r}_1(\mathbf{x}))$). Moreover, $\pi_2 \mathbf{g}(\mathbf{x}, \mathbf{y})$ is of maximal rank for all $(\mathbf{x}, \mathbf{y}) \in \tilde{\Omega} \subset B_{\rho_1}(\Omega)$ and $\mathbf{g}(\tilde{\Omega}) = \{\mathbf{0}_m\}$. Therefore, [35, Theorem 3.6] can be applied once more to yield the existence of $\delta_2, \rho_2 > 0$ and a function

$$\mathbf{r}_2 : B_{\delta_2}(\pi_x \tilde{\Omega}) \subset \pi_x W \rightarrow \mathbb{R}^m$$

that is Lipschitz continuous on $B_{\delta_2}(\pi_x \tilde{\Omega})$ such that $\pi_2 \partial \mathbf{g}(\mathbf{x}, \mathbf{y})$ is of maximal rank for all $(\mathbf{x}, \mathbf{y}) \in B_{\rho_2}(\tilde{\Omega}) \subset W$ and, for each $\mathbf{x} \in B_{\delta_2}(\pi_x \tilde{\Omega})$, $(\mathbf{x}, \mathbf{r}_2(\mathbf{x}))$ is the unique vector in $B_{\rho_2}(\tilde{\Omega})$ satisfying $\mathbf{g}(\mathbf{x}, \mathbf{r}_2(\mathbf{x})) = \mathbf{0}_m$.

Choose any $\hat{\mathbf{x}} \in \pi_x \tilde{\Omega}$. By virtue of the proof of [35, Theorem 3.6], there exist a neighborhood $N(\hat{\mathbf{x}}) \subset \pi_x W$ of $\hat{\mathbf{x}}$ and a Lipschitz continuous function

$$\mathbf{r}_{\hat{\mathbf{x}}} : N(\hat{\mathbf{x}}) \rightarrow \mathbb{R}^m$$

such that $\mathbf{g}(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \mathbf{0}_m$, where $\hat{\mathbf{y}} := \mathbf{r}_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})$, and $\pi_2 \partial \mathbf{g}(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ is of maximal rank. Moreover, $\mathbf{r}_{\hat{\mathbf{x}}} = \mathbf{r}_2$ on $N(\hat{\mathbf{x}}) \cap \pi_x \tilde{\Omega}$. By Theorem 3.1, $\mathbf{r}_{\hat{\mathbf{x}}}$ is L-smooth at $\hat{\mathbf{x}}$; for any $k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n \times k}$, $[\mathbf{r}_{\hat{\mathbf{x}}}]'(\hat{\mathbf{x}}; \mathbf{M}) = [\mathbf{r}_2]'(\hat{\mathbf{x}}; \mathbf{M})$ is the unique solution $\mathbf{N} \in \mathbb{R}^{m \times k}$ of the equation system

$$\mathbf{0}_{m \times k} = \mathbf{g}'(\hat{\mathbf{x}}, \hat{\mathbf{y}}; (\mathbf{M}, \mathbf{N})) = \mathbf{g}'(\hat{\mathbf{x}}, \mathbf{r}_2(\hat{\mathbf{x}}); (\mathbf{M}, \mathbf{N})).$$

Let $\delta := 0.5\delta_1$, $\rho := \rho_1$, and

$$\mathbf{r} : B_{\delta}(\pi_x \Omega) \subset \pi_x W \rightarrow \mathbb{R}^m : \boldsymbol{\eta} \mapsto \mathbf{r}_2(\boldsymbol{\eta}).$$

\mathbf{r} is Lipschitz continuous on $B_\delta(\pi_x\Omega) \subset \pi_x\tilde{\Omega} \subset B_{\delta_2}(\pi_x\tilde{\Omega})$ and $\pi_2\partial\mathbf{g}(\mathbf{x},\mathbf{y})$ is of maximal rank for all $(\mathbf{x},\mathbf{y}) \in B_\rho(\tilde{\Omega}) = B_{\rho_1}(\tilde{\Omega}) \subset W$. By uniqueness, $\mathbf{r}_1 = \mathbf{r}_2$ on the set $B_{\delta_1}(\pi_x\Omega) \cap B_{\delta_2}(\pi_x\tilde{\Omega}) \supset B_\delta(\pi_x\Omega)$; for each $\mathbf{x} \in B_\delta(\pi_x\Omega)$,

$$(\mathbf{x}, \mathbf{r}(\mathbf{x})) = (\mathbf{x}, \mathbf{r}_1(\mathbf{x})) = (\mathbf{x}, \mathbf{r}_2(\mathbf{x}))$$

is the unique vector in $B_{\rho_1}(\tilde{\Omega})$ satisfying $\mathbf{g}(\mathbf{x}, \mathbf{r}(\mathbf{x})) = \mathbf{0}_m$. Moreover, \mathbf{r} is L-smooth on $B_\delta(\pi_x\Omega) \subset \pi_x\tilde{\Omega}$; for any $\mathbf{x} \in B_\delta(\pi_x\Omega)$, any $k \in \mathbb{N}$, and any $\mathbf{M} \in \mathbb{R}^{n \times k}$, the LD-derivative $\mathbf{r}'(\mathbf{x}; \mathbf{M}) = [\mathbf{r}_2]'\!(\mathbf{x}; \mathbf{M})$ is the unique solution $\mathbf{N} \in \mathbb{R}^{m \times k}$ of the equation system

$$\mathbf{0}_{m \times k} = \mathbf{g}'(\mathbf{x}, \mathbf{r}_2(\mathbf{x}); (\mathbf{M}, \mathbf{N})) = \mathbf{g}'(\mathbf{x}, \mathbf{r}(\mathbf{x}); (\mathbf{M}, \mathbf{N})).$$

□

Remark 3.1 The implicit function \mathbf{r} outlined in the statement of Theorem 3.2 is L-smooth on its open domain $B_\delta(\pi_x\Omega) \supset \pi_x\Omega$, which is needed for the higher-order directional derivatives outlined earlier to be well-defined and is essential for the analysis to follow. The fact that \mathbf{r} is also Lipschitz continuous on $B_\delta(\pi_x\Omega)$ is not immediately implied by its L-smoothness. Moreover, when Ω is a singleton, Theorem 3.1 is recovered.

4 Forward Sensitivity Functions for Nonsmooth Differential-Algebraic

Equations

Let $n_p, n_x, n_y \in \mathbb{N}$. Let $D_t \subset \mathbb{R}$, $D_p \subset \mathbb{R}^{n_p}$, $D_y \subset \mathbb{R}^{n_y}$, and $D_x \subset \mathbb{R}^{n_x}$ be open and connected. Let $D := D_t \times D_p \times D_x \times D_y$, $\mathbf{f} : D \rightarrow \mathbb{R}^{n_x}$, $\mathbf{g} : D \rightarrow \mathbb{R}^{n_y}$, and $\mathbf{f}_0 : D_p \rightarrow D_x$.

Given $t_0 \in D_t$, consider the following initial-value problem (IVP) in semi-explicit DAEs:

$$\dot{\mathbf{x}}(t, \mathbf{p}) = \mathbf{f}(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p}), \mathbf{y}(t, \mathbf{p})), \quad (5a)$$

$$\mathbf{0}_{n_y} = \mathbf{g}(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p}), \mathbf{y}(t, \mathbf{p})), \quad (5b)$$

$$\mathbf{x}(t_0, \mathbf{p}) = \mathbf{f}_0(\mathbf{p}), \quad (5c)$$

where t is the independent variable and $\mathbf{p} \in D_p$ is a vector of the problem parameters.

The following assumption is made regarding the right-hand side functions in (5).

Assumption 4.1 Let $t_f \in D_t$ satisfy $t_0 < t_f$ and $Z_{\mathbf{f}}$ be a zero-measure subset of $[t_0, t_f]$.

Suppose that the following conditions hold:

- (i) $\mathbf{f}(\cdot, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y)$ is measurable on $[t_0, t_f]$ for each $(\mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y) \in D_p \times D_x \times D_y$;
- (ii) $\mathbf{f}(t, \cdot, \cdot, \cdot)$ is L-smooth on $D_p \times D_x \times D_y$ for each $t \in [t_0, t_f] \setminus Z_{\mathbf{f}}$;
- (iii) there exist Lebesgue integrable functions $k_{\mathbf{f}}, m_{\mathbf{f}} : [t_0, t_f] \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ for which:
 - (a) $\|\mathbf{f}(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y)\| \leq m_{\mathbf{f}}(t), \quad \forall t \in [t_0, t_f], \quad \forall (\mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y) \in D_p \times D_x \times D_y$;
 - (b) $\|\mathbf{f}(t, \mathbf{p}_1, \boldsymbol{\eta}_{x_1}, \boldsymbol{\eta}_{y_1}) - \mathbf{f}(t, \mathbf{p}_2, \boldsymbol{\eta}_{x_2}, \boldsymbol{\eta}_{y_2})\| \leq k_{\mathbf{f}}(t) \|(\mathbf{p}_1, \boldsymbol{\eta}_{x_1}, \boldsymbol{\eta}_{y_1}) - (\mathbf{p}_2, \boldsymbol{\eta}_{x_2}, \boldsymbol{\eta}_{y_2})\|,$
 $\forall t \in [t_0, t_f], \quad \forall (\mathbf{p}_1, \boldsymbol{\eta}_{x_1}, \boldsymbol{\eta}_{y_1}), (\mathbf{p}_2, \boldsymbol{\eta}_{x_2}, \boldsymbol{\eta}_{y_2}) \in D_p \times D_x \times D_y$;
- (iv) \mathbf{g} and \mathbf{f}_0 are L-smooth on D and D_p , respectively.

Notions of consistent initialization, regularity, and solutions of (5) from [35] are reproduced here for the reader's convenience.

Definition 4.1 The consistency set, initial consistency set, and regularity set of (5)

are given by, respectively,

$$G_C := \{(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y) \in D : \mathbf{g}(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y) = \mathbf{0}_{n_y}\},$$

$$G_{C,0} := \{(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y) \in G_C : t = t_0, \boldsymbol{\eta}_x = \mathbf{f}_0(\mathbf{p})\},$$

$$G_R := \{(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y) \in D : \pi_4 \partial \mathbf{g}(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y) \text{ is of maximal rank}\}.$$

Definition 4.2 Let $P \subset D_p$, $\Omega_0 \subset G_{C,0}$, and $T \subset D_t$ be a connected set containing t_0 .

A mapping $\mathbf{z} \equiv (\mathbf{x}, \mathbf{y}) : T \times P \rightarrow D_x \times D_y$ is called a solution of (5) on $T \times P$ through

Ω_0 if, for each $\mathbf{p} \in P$, $\mathbf{z}(\cdot, \mathbf{p})$ is an absolutely continuous function on T which satisfies

(5a) for almost every $t \in T$, (5b) for every $t \in T$, (5c) at $t = t_0$, and

$$\{(t_0, \mathbf{p}, \mathbf{x}(t_0, \mathbf{p}), \mathbf{y}(t_0, \mathbf{p})) : \mathbf{p} \in P\} = \Omega_0.$$

If, in addition,

$$\{(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p}), \mathbf{y}(t, \mathbf{p})) : (t, \mathbf{p}) \in T \times P\} \subset G_R,$$

then \mathbf{z} is called a regular solution of (5) on $T \times P$ through Ω_0 .

Definition 4.3 Let \mathbf{z} be a solution of (5) on $T \times P$ through Ω_0 . Then \mathbf{z} is said to

be unique if, given any other solution \mathbf{z}^* of (5) on $T^* \times P^*$ through Ω_0^* satisfying

$T \cap T^* \neq \{t_0\}$, $P \cap P^* \neq \emptyset$, and

$$\{(t_0, \mathbf{p}, \mathbf{z}(t_0, \mathbf{p})) : \mathbf{p} \in P \cap P^*\} = \{(t_0, \mathbf{p}, \mathbf{z}^*(t_0, \mathbf{p})) : \mathbf{p} \in P \cap P^*\},$$

$\mathbf{z}(t, \mathbf{p}) = \mathbf{z}^*(t, \mathbf{p})$ for all $(t, \mathbf{p}) \in (T \cap T^*) \times (P \cap P^*)$.

A generalization of the notion that (5) has differential index equal to one (see [14, 22]) for all $(t, \mathbf{p}) \in T \times P$ is implied by regularity. The following assumption regarding the existence of a solution is made.

Assumption 4.2 Suppose that for some $(\mathbf{p}_0, \mathbf{x}_0, \mathbf{y}_0) \in D_p \times D_x \times D_y$, there exists a regular solution \mathbf{z} of (5) on $[t_0, t_f] \times \{\mathbf{p}_0\}$ through $\{(t_0, \mathbf{p}_0, \mathbf{x}_0, \mathbf{y}_0)\}$.

Before proceeding to the main result, a proposition is proved concerning uniqueness and parametric dependence of solutions of (5), as well as its equivalence to parametric Carathéodory ODEs via an extended implicit function.

Proposition 4.1 *Let Assumptions 4.1 and 4.2 hold. Then there exists a neighborhood $N(\mathbf{p}_0) \subset D_p$ of \mathbf{p}_0 , a set $\Omega_0 \subset G_{C,0}$ containing $(t_0, \mathbf{p}_0, \mathbf{x}_0, \mathbf{y}_0)$, and a unique regular solution \mathbf{z} of (5) on $[t_0, t_f] \times N(\mathbf{p}_0)$ through Ω_0 . Furthermore, there exist $\delta, \rho > 0$ and a function*

$$\mathbf{r} : B_\delta(\{(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\}) \subset D_t \times D_p \times D_x \rightarrow \mathbb{R}^{n_y}$$

that is Lipschitz continuous and L -smooth on its open and connected domain which satisfy $\mathbf{y}(t, \mathbf{p}) = \mathbf{r}(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p}))$ for all $(t, \mathbf{p}) \in [t_0, t_f] \times N(\mathbf{p}_0)$ and

$$\{(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p})) : (t, \mathbf{p}) \in [t_0, t_f] \times N(\mathbf{p}_0)\} \subset B_\delta(\{(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\}),$$

$$\{(t, \mathbf{p}, \mathbf{z}(t, \mathbf{p})) : (t, \mathbf{p}) \in [t_0, t_f] \times N(\mathbf{p}_0)\} \subset B_\rho(\{(t, \mathbf{p}_0, \mathbf{z}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\}) \subset D.$$

Proof Define the following compact sets:

$$\Lambda := \{(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\},$$

$$\Omega := \{(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0), \mathbf{y}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\}.$$

$\pi_4 \partial \mathbf{g}(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y)$ is of maximal rank for all $(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y) \in \Omega$ by regularity and $\mathbf{g}(\Omega) = \{\mathbf{0}_{n_y}\}$ by consistency. Each point in Λ is the projection of a unique point in Ω . By Theorem 3.2, there exist $\delta_1, \rho_1 > 0$ and a function

$$\mathbf{r} : B_{\delta_1}(\Lambda) \subset D_t \times D_p \times D_x \rightarrow \mathbb{R}^{n_y}$$

that is Lipschitz continuous and L-smooth on $B_{\delta_1}(\Lambda)$ such that $\pi_4 \partial \mathbf{g}(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y)$ is of maximal rank for all $(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y) \in B_{\rho_1}(\Omega) \subset D$. Moreover, for each vector $(t, \mathbf{p}, \boldsymbol{\eta}_x) \in B_{\delta_1}(\Lambda)$, $(t, \mathbf{p}, \boldsymbol{\eta}_x, \mathbf{r}(t, \mathbf{p}, \boldsymbol{\eta}_x))$ is the unique vector in $B_{\rho_1}(\Omega)$ which satisfies the equation $\mathbf{g}(t, \mathbf{p}, \boldsymbol{\eta}_x, \mathbf{r}(t, \mathbf{p}, \boldsymbol{\eta}_x)) = \mathbf{0}_{n_y}$.

By proceeding as in the proof of [35, Theorem 4.34] using the inherited L-smoothness of the implicit function in place of the Lipschitzian construction, the following conclusions are immediately furnished: there exist $\xi, \beta > 0$ satisfying $\beta < \xi$ and a regular solution \mathbf{z} of (5) on $[t_0, t_f] \times B_\beta(\mathbf{p}_0) \subset D_t \times D_p$ through

$$\Omega_0 := \{(t, \mathbf{p}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_y) : t = t_0, \mathbf{p} \in B_\beta(\mathbf{p}_0), \boldsymbol{\eta}_x = \mathbf{f}_0(\mathbf{p}), \boldsymbol{\eta}_y = \mathbf{r}(t_0, \mathbf{p}, \mathbf{f}_0(\mathbf{p}))\} \subset G_{C,0}$$

such that $\mathbf{y}(t, \mathbf{p}) = \mathbf{r}(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p}))$ for all $\mathbf{p} \in B_\beta(\mathbf{p}_0)$ and

$$\{(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p}), \mathbf{y}(t, \mathbf{p})) : (t, \mathbf{p}) \in [t_0, t_f] \times B_\beta(\mathbf{p}_0)\} \subset B_{\rho_1}(\Omega).$$

Moreover, the intermediate construction \mathbf{u} in the proof of [35, Theorem 4.34] satisfies

$$\begin{aligned} & \{(t, \mathbf{u}(t, \mathbf{c})) : (t, \mathbf{c}) \in [t_0, t_f] \times B_\xi(\mathbf{p}_0, \mathbf{f}_0(\mathbf{p}_0))\} \subset B_{0.5\delta_1}(\Lambda), \\ & \begin{bmatrix} \mathbf{p} \\ \mathbf{x}(t, \mathbf{p}) \end{bmatrix} = \mathbf{u}(t, (\mathbf{p}, \mathbf{f}_0(\mathbf{p}))), \quad \forall (t, \mathbf{p}) \in [t_0, t_f] \times B_\beta(\mathbf{p}_0), \end{aligned}$$

and $\{(\mathbf{p}, \mathbf{f}_0(\mathbf{p})) : \mathbf{p} \in B_\beta(\mathbf{p}_0)\} \subset B_\xi(\mathbf{p}_0, \mathbf{f}_0(\mathbf{p}_0))$. Thus,

$$\begin{aligned} & \{(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p})) : (t, \mathbf{p}) \in [t_0, t_f] \times B_\beta(\mathbf{p}_0)\} \\ & = \{(t, \mathbf{u}(t, (\mathbf{p}, \mathbf{f}_0(\mathbf{p}))) : (t, \mathbf{p}) \in [t_0, t_f] \times B_\beta(\mathbf{p}_0)\}, \\ & \subset \{(t, \mathbf{u}(t, \mathbf{c})) : (t, \mathbf{c}) \in [t_0, t_f] \times B_\xi(\mathbf{p}_0, \mathbf{f}_0(\mathbf{p}_0))\}, \\ & \subset B_{0.5\delta_1}(\Lambda). \end{aligned}$$

Then [35, Theorem 4.22] implies that \mathbf{z} is the unique regular solution of (5) on $[t_0, t_f] \times B_\beta(\mathbf{p}_0)$ through Ω_0 and the result holds with $N(\mathbf{p}_0) := B_\beta(\mathbf{p}_0)$, $\delta := 0.5\delta_1$, and $\rho := \rho_1$. \square

Using lexicographic differentiation, forward sensitivity functions for (5) are given.

Theorem 4.1 *Let Assumptions 4.1 and 4.2 hold. Then, for each $t \in [t_0, t_f]$, the mapping $\mathbf{z}_t \equiv \mathbf{z}(t, \cdot)$ is L -smooth at \mathbf{p}_0 ; for any $k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n_p \times k}$, the LD-derivative mapping*

$$\tilde{\mathbf{Z}} \equiv (\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) : [t_0, t_f] \rightarrow \mathbb{R}^{(n_x+n_y) \times k} : t \mapsto [\mathbf{z}_t]'(\mathbf{p}_0; \mathbf{M})$$

is such that $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ are absolutely continuous and Lebesgue integrable on $[t_0, t_f]$, respectively. Furthermore, $\tilde{\mathbf{Z}}$ uniquely (in the sense of Definition 4.3) satisfies the following DAE system:

$$\begin{aligned} \dot{\mathbf{X}}(t) &= [\mathbf{f}_{t \setminus Z_t}]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0), \mathbf{y}(t, \mathbf{p}_0); (\mathbf{M}, \mathbf{X}(t), \mathbf{Y}(t))), \quad \text{a.e. } t \in [t_0, t_f], \\ \mathbf{0}_{n_y \times k} &= [\mathbf{g}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0), \mathbf{y}(t, \mathbf{p}_0); (\mathbf{M}, \mathbf{X}(t), \mathbf{Y}(t))), \quad \forall t \in [t_0, t_f], \quad (6) \\ \mathbf{X}(t_0) &= [\mathbf{f}_0]'(\mathbf{p}_0; \mathbf{M}), \end{aligned}$$

on $[t_0, t_f]$ through $\{(t_0, \mathbf{X}_0, \mathbf{Y}_0)\}$, where $\mathbf{X}_0 := [\mathbf{f}_0]'(\mathbf{p}_0; \mathbf{M})$ and $\mathbf{Y}_0 \in \mathbb{R}^{n_y \times k}$ is the unique solution of the equation system

$$\mathbf{0}_{n_y \times k} = [\mathbf{g}_{t_0}]'(\mathbf{p}_0, \mathbf{x}_0, \mathbf{y}_0; (\mathbf{M}, \mathbf{X}_0, \mathbf{Y}_0)).$$

Proof Let $\delta, \rho > 0$, \mathbf{r} , and $N(\mathbf{p}_0)$ be given as in the statement of Proposition 4.1.

Define the sets

$$D_\delta := B_\delta(\{(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\}) \subset D_t \times D_p \times D_x,$$

$$D_\rho := B_\rho(\{(t, \mathbf{p}_0, \mathbf{z}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\}) \subset D,$$

and the following mappings:

$$\begin{aligned}\mathbf{q} : D_\delta &\rightarrow D_p \times D_x \times D_y : (t, \mathbf{p}, \boldsymbol{\eta}_x) \mapsto (\mathbf{p}, \boldsymbol{\eta}_x, \mathbf{r}(t, \mathbf{p}, \boldsymbol{\eta}_x)), \\ \bar{\mathbf{f}} : D_\delta &\rightarrow \mathbb{R}^{n_x} : (t, \mathbf{p}, \boldsymbol{\eta}_x) \mapsto \mathbf{f}(t, \mathbf{q}(t, \mathbf{p}, \boldsymbol{\eta}_x)).\end{aligned}$$

For each $(\mathbf{p}, \boldsymbol{\eta}_x) \in \pi_{p,x} D_\delta \subset D_p \times D_x$,

$$\bar{\mathbf{f}}(\cdot, \mathbf{p}, \boldsymbol{\eta}_x) \equiv \mathbf{f}(\cdot, \mathbf{q}(\cdot, \mathbf{p}, \boldsymbol{\eta}_x)) : \pi_t(D_\delta; (\mathbf{p}, \boldsymbol{\eta}_x)) \subset D_t \rightarrow \mathbb{R}^{n_x}$$

is measurable on $[t_0, t_f]$ by [39, Chap. 1, Sect. 1, Lemma 1] because the mapping

$t \mapsto (\mathbf{p}, \boldsymbol{\eta}_x, \mathbf{r}(t, \mathbf{p}, \boldsymbol{\eta}_x))$ is continuous on $\pi_t(D_\delta; (\mathbf{p}, \boldsymbol{\eta}_x)) \supset [t_0, t_f]$,

$$[t_0, t_f] \times \mathbf{q}(D_\delta) \subset [t_0, t_f] \times D_p \times D_x \times D_y, \quad (7)$$

and \mathbf{f} satisfies the Carathéodory conditions (see, e.g., [39]) on $[t_0, t_f] \times D_p \times D_x \times D_y$ by assumption.

L-smoothness of $\bar{\mathbf{f}}(t, \cdot, \cdot)$ is demonstrated as follows: for each $t \in [t_0, t_f]$, the mapping $(\mathbf{p}, \boldsymbol{\eta}_x) \mapsto (\mathbf{p}, \boldsymbol{\eta}_x, \mathbf{r}(t, \mathbf{p}, \boldsymbol{\eta}_x))$ is L-smooth on $\pi_{p,x}(D_\delta; t)$ by L-smoothness of \mathbf{r} on D_δ (and hence $\mathbf{r}_t \equiv \mathbf{r}(t, \cdot, \cdot)$ on $\pi_{p,x}(D_\delta; t)$). Thus, $\mathbf{q}_t \equiv \mathbf{q}(t, \cdot, \cdot)$ is L-smooth on $\pi_{p,x}(D_\delta; t)$. Since

$$\mathbf{q}_t(\pi_{p,x}(D_\delta; t)) \subset \pi_{p,x,y}(D_\delta; t) \subset D_p \times D_x \times D_y, \quad \forall t \in [t_0, t_f] \setminus Z_{\mathbf{f}},$$

and the composition of L-smooth functions is L-smooth, it follows that

$$\bar{\mathbf{f}}(t, \cdot, \cdot) \equiv \mathbf{f}(t, \mathbf{q}(t, \cdot, \cdot)) : \pi_{p,x}(D_\delta; t) \subset D_p \times D_x \rightarrow \mathbb{R}^{n_x}$$

is L-smooth on $\pi_{p,x}(D_\delta; t)$ for each $t \in [t_0, t_f] \setminus Z_{\mathbf{f}}$.

For any $t \in [t_0, t_f]$ and any $(\mathbf{p}, \boldsymbol{\eta}_x) \in \pi_{p,x}(D_\delta; t)$,

$$\|\bar{\mathbf{f}}(t, \mathbf{p}, \boldsymbol{\eta}_x)\| = \|\mathbf{f}(t, \mathbf{p}, \boldsymbol{\eta}_x, \mathbf{r}(t, \mathbf{p}, \boldsymbol{\eta}_x))\| \leq m_{\mathbf{f}}(t),$$

by (7) and the Carathéodory conditions of \mathbf{f} . By Lipschitz continuity of \mathbf{r} on D_δ , there exists $k_r \geq 0$ such that

$$\|\mathbf{r}(t, \mathbf{p}_1, \boldsymbol{\eta}_{x_1}) - \mathbf{r}(t, \mathbf{p}_2, \boldsymbol{\eta}_{x_2})\| \leq k_r \|(\mathbf{p}_1, \boldsymbol{\eta}_{x_1}) - (\mathbf{p}_2, \boldsymbol{\eta}_{x_2})\|, \quad (8)$$

for any $(t, \mathbf{p}_1, \boldsymbol{\eta}_{x_1}), (t, \mathbf{p}_2, \boldsymbol{\eta}_{x_2}) \in D_\delta$. It follows that

$$\begin{aligned} \|\bar{\mathbf{f}}(t, \mathbf{p}_1, \boldsymbol{\eta}_{x_1}) - \bar{\mathbf{f}}(t, \mathbf{p}_2, \boldsymbol{\eta}_{x_2})\| &= \|\mathbf{f}(t, \mathbf{p}_1, \boldsymbol{\eta}_{x_1}, \mathbf{r}(t, \mathbf{p}_1, \boldsymbol{\eta}_{x_1})) - \mathbf{f}(t, \mathbf{p}_2, \boldsymbol{\eta}_{x_2}, \mathbf{r}(t, \mathbf{p}_2, \boldsymbol{\eta}_{x_2}))\|, \\ &\leq k_f(t) \|(\mathbf{p}_1, \boldsymbol{\eta}_{x_1}, \mathbf{r}(t, \mathbf{p}_1, \boldsymbol{\eta}_{x_1})) - (\mathbf{p}_2, \boldsymbol{\eta}_{x_2}, \mathbf{r}(t, \mathbf{p}_2, \boldsymbol{\eta}_{x_2}))\|, \\ &\leq k_f(t)(1 + k_r) \|(\mathbf{p}_1, \boldsymbol{\eta}_{x_1}) - (\mathbf{p}_2, \boldsymbol{\eta}_{x_2})\|, \end{aligned}$$

for any $t \in [t_0, t_f]$ and any $(\mathbf{p}_1, \boldsymbol{\eta}_{x_1}), (\mathbf{p}_2, \boldsymbol{\eta}_{x_2}) \in \pi_{p,x}(D_\delta; t)$.

By replacing \mathbf{f} by $\bar{\mathbf{f}}$ and D by D_δ , it is valid to apply Theorem 2.2 to

$$\dot{\mathbf{u}}(t, \mathbf{p}) = \bar{\mathbf{f}}(t, \mathbf{u}(t, \mathbf{p})),$$

$$\mathbf{u}(t_0, \mathbf{p}) = \mathbf{f}_0(\mathbf{p}),$$

which admits the unique solution $\mathbf{x}(\cdot, \mathbf{p})$ on $[t_0, t_f]$ for each $\mathbf{p} \in N(\mathbf{p}_0)$. Theorem 2.2 yields that, for each $t \in [t_0, t_f]$, $\mathbf{x}_t \equiv \mathbf{x}(t, \cdot)$ is Lipschitz continuous on a neighborhood $\tilde{N}(\mathbf{p}_0) \subset N(\mathbf{p}_0)$ of \mathbf{p}_0 , with Lipschitz constant $k_x \geq 0$. For any $\mathbf{p}_1, \mathbf{p}_2 \in \tilde{N}(\mathbf{p}_0)$,

$$\begin{aligned} \|\mathbf{y}(t, \mathbf{p}_1) - \mathbf{y}(t, \mathbf{p}_2)\| &= \|\mathbf{r}(t, \mathbf{p}_1, \mathbf{x}(t, \mathbf{p}_1)) - \mathbf{r}(t, \mathbf{p}_2, \mathbf{x}(t, \mathbf{p}_2))\|, \\ &\leq k_r \|(\mathbf{p}_1, \mathbf{x}(t, \mathbf{p}_1)) - (\mathbf{p}_2, \mathbf{x}(t, \mathbf{p}_2))\|, \\ &\leq k_r(1 + k_x) \|\mathbf{p}_1 - \mathbf{p}_2\|, \end{aligned}$$

since $\{(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p})) : (t, \mathbf{p}) \in [t_0, t_f] \times \tilde{N}(\mathbf{p}_0)\} \subset D_\delta$. This demonstrates Lipschitz continuity of \mathbf{y}_t on $\tilde{N}(\mathbf{p}_0)$, with a Lipschitz constant that is independent of t . From Theorem 2.2 it also follows that \mathbf{x}_t is L-smooth at \mathbf{p}_0 , which implies that the mapping

\mathbf{y}_t is L-smooth at \mathbf{p}_0 for any $t \in [t_0, t_f]$ since $\mathbf{y}_t(\cdot) \equiv \mathbf{r}_t(\cdot, \mathbf{x}_t(\cdot))$ on $N(\mathbf{p}_0)$ and \mathbf{r}_t is L-smooth at $(\mathbf{p}_0, \mathbf{x}_t(\mathbf{p}_0))$.

Define the following mappings:

$$\begin{aligned}\tilde{\mathbf{r}}_{(0)} &: [t_0, t_f] \times \mathbb{R}^{n_p+n_x} \rightarrow \mathbb{R}^{n_y} : (t, \mathbf{d}) \mapsto [\mathbf{r}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); \mathbf{d}), \\ \tilde{\mathbf{r}}_{(i)} &: [t_0, t_f] \times \mathbb{R}^{n_p+n_x} \rightarrow \mathbb{R}^{n_y} : (t, \mathbf{d}) \mapsto [\tilde{\mathbf{r}}_{(i-1), t}]'(\mathbf{m}_{(i)}, [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(i-1)}(\mathbf{m}_{(i)}); \mathbf{d}), \\ &\forall i \in \{1, \dots, k-1\},\end{aligned}$$

which are well-defined since \mathbf{r} is L-smooth on $D_\delta \supset \{(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\}$. It will be shown by induction that, for each $i \in \{0, 1, \dots, k-1\}$, $\tilde{\mathbf{r}}_{(i)}(\cdot, \mathbf{d})$ is measurable on $[t_0, t_f]$ for each $\mathbf{d} \in \mathbb{R}^{n_p+n_x}$ and $\tilde{\mathbf{r}}_{(i)}(t, \cdot)$ is Lipschitz continuous on $\mathbb{R}^{n_p+n_x}$ for each $t \in [t_0, t_f]$, with a Lipschitz constant that is independent of t .

Consider the base case and choose any $\mathbf{d} \in \mathbb{R}^{n_p+n_x}$. The first part proceeds as in the proof of [10, Theorem 4.1]: by construction,

$$\{(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\} \subset D_\delta,$$

where $\{(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\}$ is compact and D_δ is open. Thus,

$$\{(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) : t \in [t_0, t_f]\} \cap (\mathbb{R}^{1+n_p+n_x} \setminus D_\delta) = \emptyset.$$

Let $\tilde{\mathbf{d}} := (0, \mathbf{d})$. There exists $\varepsilon > 0$ such that for any $t \in [t_0, t_f]$ and any $\tau \in [0, \varepsilon]$,

$$(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) + \tau \tilde{\mathbf{d}} = (t, (\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) + \tau \mathbf{d}) \in D_\delta;$$

this follows from [39, Chap. 2, Sect. 5, Lemma 1]. Since $t \mapsto (\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0))$ is continuous on $[t_0, t_f]$, the composite mapping $t \mapsto \mathbf{r}(t, (\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) + \tau \mathbf{d})$ is continuous on $[t_0, t_f]$ for each $\tau \in [0, \varepsilon]$. The mapping

$$t \mapsto \lim_{\alpha \downarrow 0} \frac{\mathbf{r}(t, (\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) + \alpha \mathbf{d}) - \mathbf{r}(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0))}{\alpha}$$

is the pointwise limit of a sequence of continuous functions and is therefore measurable on $[t_0, t_f]$, from which it follows that $\tilde{\mathbf{r}}_{(0)}(\cdot, \mathbf{d})$ is measurable on $[t_0, t_f]$ for each $\mathbf{d} \in \mathbb{R}^{n_p+n_x}$.

Choose any $t \in [t_0, t_f]$. The function \mathbf{r}_t is Lipschitz continuous on $\pi_{p,x}(D_\delta; t)$ and directionally differentiable at $(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0))$. By (8), k_r acts as a Lipschitz constant for \mathbf{r}_t in a neighborhood of $(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0))$, and as a result

$$\begin{aligned} \|\tilde{\mathbf{r}}_{(0)}(t, \mathbf{d}_1) - \tilde{\mathbf{r}}_{(0)}(t, \mathbf{d}_2)\| &= \|[\mathbf{r}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); \mathbf{d}_1) - [\mathbf{r}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); \mathbf{d}_2)\|, \\ &\leq k_r \|\mathbf{d}_1 - \mathbf{d}_2\|, \quad \forall t \in [t_0, t_f], \quad \forall \mathbf{d}_1, \mathbf{d}_2 \in \mathbb{R}^{n_p+n_x}, \end{aligned}$$

by [9, Theorem 3.1.2]. Hence, $\tilde{\mathbf{r}}_{(0)}(t, \cdot)$ is Lipschitz continuous on $\mathbb{R}^{n_p+n_x}$ for each $t \in [t_0, t_f]$, with Lipschitz constant k_r .

Assume that the claim is true for $i := j \in \{0, 1, \dots, k-2\}$ and choose any vector $\mathbf{d} \in \mathbb{R}^{n_p+n_x}$. From [10, Corollary 4.2], $t \mapsto [\mathbf{x}_r]_{\mathbf{p}_0, \mathbf{M}}^{(i-1)}(\mathbf{m}_{(i)})$ is an absolutely continuous mapping on $[t_0, t_f]$ for each $i \in \{1, \dots, k\}$. Hence, the mapping

$$t \mapsto \begin{bmatrix} \mathbf{m}_{(j+1)} \\ [\mathbf{x}_r]_{\mathbf{p}_0, \mathbf{M}}^{(j)}(\mathbf{m}_{(j+1)}) \end{bmatrix} + \tau \mathbf{d}$$

is absolutely continuous, and therefore measurable, on $[t_0, t_f]$ for any $\tau \geq 0$. By the inductive assumption, $\tilde{\mathbf{r}}_{(j)}(\cdot, \boldsymbol{\eta})$ is a measurable mapping on $[t_0, t_f]$ for any $\boldsymbol{\eta} \in \mathbb{R}^{n_p+n_x}$ and there exists $k_{\tilde{\mathbf{r}}_{(j)}} \geq 0$ such that

$$\|\tilde{\mathbf{r}}_{(j)}(t, \mathbf{d}_1) - \tilde{\mathbf{r}}_{(j)}(t, \mathbf{d}_2)\| \leq k_{\tilde{\mathbf{r}}_{(j)}} \|\mathbf{d}_1 - \mathbf{d}_2\|, \quad \forall t \in [t_0, t_f], \quad \forall \mathbf{d}_1, \mathbf{d}_2 \in \mathbb{R}^{n_p+n_x}.$$

Hence,

$$\|\tilde{\mathbf{r}}_{(j)}(t, \mathbf{d})\| = \|\tilde{\mathbf{r}}_{(j)}(t, \mathbf{d}) - \tilde{\mathbf{r}}_{(j)}(t, \mathbf{0}_{n_p+n_x})\| \leq k_{\tilde{\mathbf{r}}_{(j)}} \|\mathbf{d}\|, \quad \forall (t, \mathbf{d}) \in [t_0, t_f] \times \mathbb{R}^{n_p+n_x}.$$

Consequently, the mapping

$$t \mapsto \frac{\tilde{\mathbf{r}}_{(j)}(t, (\mathbf{m}_{(j+1)}, [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(j)}(\mathbf{m}_{(j+1)})) + \tau \mathbf{d}) - \tilde{\mathbf{r}}_{(j)}(t, \mathbf{m}_{(j+1)}, [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(j)}(\mathbf{m}_{(j+1)}))}{\tau}$$

is Lebesgue integrable, and therefore measurable, on $[t_0, t_f]$ for any $\tau > 0$ by [39, Chap. 1, Sect. 1, Lemma 1]. Then, since the mapping $\tilde{\mathbf{r}}_{(j),t}$ is directionally differentiable at $(\mathbf{m}_{(j+1)}, [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(j)}(\mathbf{m}_{(j+1)}))$, the mapping

$$t \mapsto \lim_{\alpha \downarrow 0} \frac{\tilde{\mathbf{r}}_{(j)}(t, (\mathbf{m}_{(j+1)}, [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(j)}(\mathbf{m}_{(j+1)})) + \alpha \mathbf{d}) - \tilde{\mathbf{r}}_{(j)}(t, \mathbf{m}_{(j+1)}, [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(j)}(\mathbf{m}_{(j+1)}))}{\alpha}$$

is well-defined and is measurable on $[t_0, t_f]$ as the pointwise limit of a sequence of measurable functions. Hence, $\tilde{\mathbf{r}}_{(j+1)}(\cdot, \mathbf{d})$ is measurable on $[t_0, t_f]$ for each $\mathbf{d} \in \mathbb{R}^{n_p + n_x}$.

Again by [9, Theorem 3.1.2], the finite constant $k_{\tilde{\mathbf{r}}_{(j)}}$ acts as a Lipschitz constant for $\tilde{\mathbf{r}}_{(j+1),t} \equiv [\tilde{\mathbf{r}}_{(j),t}]'(\mathbf{m}_{(j+1)}, [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(j)}(\mathbf{m}_{(j+1)}); \cdot)$ on $\mathbb{R}^{n_p + n_x}$:

$$\|\tilde{\mathbf{r}}_{(j+1)}(t, \mathbf{d}_1) - \tilde{\mathbf{r}}_{(j+1)}(t, \mathbf{d}_2)\| \leq k_{\tilde{\mathbf{r}}_{(j)}} \|\mathbf{d}_1 - \mathbf{d}_2\|, \quad \forall t \in [t_0, t_f], \quad \forall \mathbf{d}_1, \mathbf{d}_2 \in \mathbb{R}^{n_p + n_x},$$

implying that $\tilde{\mathbf{r}}_{(j+1)}(t, \cdot)$ is Lipschitz continuous on $\mathbb{R}^{n_p + n_x}$ for each $t \in [t_0, t_f]$, with a Lipschitz constant that is independent of t . The claim is therefore proved by induction.

Define the following mappings:

$$\tilde{\mathbf{x}}_{(i)} : [t_0, t_f] \rightarrow \mathbb{R}^{n_x} : t \mapsto [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(i-1)}(\mathbf{m}_{(i)}), \quad \forall i \in \{1, \dots, k\},$$

$$\tilde{\mathbf{y}}_{(i)} : [t_0, t_f] \rightarrow \mathbb{R}^{n_y} : t \mapsto \tilde{\mathbf{r}}_{(i-1)}(t, \mathbf{m}_{(i)}, [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(i-1)}(\mathbf{m}_{(i)})), \quad \forall i \in \{1, \dots, k\}.$$

Choose any $i := j \in \{1, \dots, k\}$. For each $\mathbf{d} \in \mathbb{R}^{n_p + n_x}$, the mapping $\tilde{\mathbf{r}}_{(j-1)}(\cdot, \mathbf{d})$ is measurable on $[t_0, t_f]$. Moreover, there exists $k_{\tilde{\mathbf{r}}_{(j-1)}} \geq 0$ such that

$$\|\tilde{\mathbf{r}}_{(j-1)}(t, \mathbf{d}_1) - \tilde{\mathbf{r}}_{(j-1)}(t, \mathbf{d}_2)\| \leq k_{\tilde{\mathbf{r}}_{(j-1)}} \|\mathbf{d}_1 - \mathbf{d}_2\|, \quad \forall t \in [t_0, t_f], \quad \forall \mathbf{d}_1, \mathbf{d}_2 \in \mathbb{R}^{n_p + n_x},$$

and

$$\begin{aligned} \|\tilde{\mathbf{r}}_{(j-1)}(t, \mathbf{d})\| &= \|\tilde{\mathbf{r}}_{(j-1)}(t, \mathbf{d}) - \tilde{\mathbf{r}}_{(j-1)}(t, \mathbf{0}_{n_p+n_x})\|, \\ &\leq k_{\tilde{\mathbf{r}}_{(j-1)}} \|\mathbf{d}\|, \quad \forall (t, \mathbf{d}) \in [t_0, t_f] \times \mathbb{R}^{n_p+n_x}. \end{aligned}$$

It was demonstrated earlier that the mapping $\tilde{\mathbf{x}}_{(j)} : t \mapsto [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(j-1)}(\mathbf{m}_{(j)})$ is absolutely continuous on $[t_0, t_f]$. $\tilde{\mathbf{x}}_{(j)}$ is therefore measurable on $[t_0, t_f]$, from which it follows that $\tilde{\mathbf{y}}_{(j)}$ is Lebesgue integrable on $[t_0, t_f]$ by [39, Chap. 1, Sect. 1, Lemma 1].

Define the following matrix-valued functions:

$$\begin{aligned} \tilde{\mathbf{X}} : [t_0, t_f] &\rightarrow \mathbb{R}^{n_x \times k} : t \mapsto [\tilde{\mathbf{x}}_{(1)}(t) \quad \cdots \quad \tilde{\mathbf{x}}_{(k)}(t)], \\ \tilde{\mathbf{Y}} : [t_0, t_f] &\rightarrow \mathbb{R}^{n_y \times k} : t \mapsto [\tilde{\mathbf{y}}_{(1)}(t) \quad \cdots \quad \tilde{\mathbf{y}}_{(k)}(t)]. \end{aligned}$$

For any $k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n_p \times k}$, Theorem 2.2 implies that the LD-derivative mapping $t \mapsto [\mathbf{x}_t]'(\mathbf{p}_0; \mathbf{M})$ is the unique solution on $[t_0, t_f]$ of the following ODE system:

$$\begin{aligned} \dot{\mathbf{U}}(t) &= [\tilde{\mathbf{f}}_{t \setminus Z_t}]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, \mathbf{U}(t))), \\ \mathbf{U}(t_0) &= [\mathbf{f}_0]'(\mathbf{p}_0; \mathbf{M}). \end{aligned} \tag{9}$$

By L-smoothness of \mathbf{q}_t and \mathbf{r}_t at $(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0))$ for each $t \in [t_0, t_f]$, the LD-derivative chain rule (1) yields

$$\begin{aligned} &[\tilde{\mathbf{f}}_{t \setminus Z_t} \circ \mathbf{q}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, \mathbf{A})) \\ &= [\tilde{\mathbf{f}}_{t \setminus Z_t}]'(\mathbf{q}_t(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)); [\mathbf{q}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, \mathbf{A}))), \\ &= [\tilde{\mathbf{f}}_{t \setminus Z_t}]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0), \mathbf{r}(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)); (\mathbf{M}, \mathbf{A}, [\mathbf{r}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, \mathbf{A})))), \end{aligned}$$

for any $(t, \mathbf{A}) \in [t_0, t_f] \times \mathbb{R}^{n_x \times k}$. Since (9) admits the unique solution $\tilde{\mathbf{X}}$ on $[t_0, t_f]$,

$$\dot{\tilde{\mathbf{X}}}(t) = [\mathbf{f}_t \setminus \mathbf{z}_t]' \left(\left[\begin{array}{c} \mathbf{p}_0 \\ \mathbf{x}(t, \mathbf{p}_0) \\ \mathbf{r}(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0)) \end{array} \right]; \left[\begin{array}{c} \mathbf{M} \\ \tilde{\mathbf{X}}(t) \\ [\mathbf{r}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, \tilde{\mathbf{X}}(t))) \end{array} \right] \right), \quad (10)$$

for almost every $t \in [t_0, t_f]$ and

$$\tilde{\mathbf{X}}(t_0) = [\mathbf{f}_0]'(\mathbf{p}_0; \mathbf{M}). \quad (11)$$

For each $t \in [t_0, t_f]$ and each $i \in \{1, \dots, k\}$,

$$\begin{aligned} \tilde{\mathbf{y}}_{(i)}(t) &= \tilde{\mathbf{r}}_{(i-1)}(t, \mathbf{m}_{(i)}, [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(i-1)}(\mathbf{m}_{(i)})), \\ &= [\mathbf{r}_t]_{(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0), (\mathbf{M}, [\mathbf{x}_t]'(\mathbf{p}_0; \mathbf{M}))}^{(i-1)}(\mathbf{m}_{(i)}, [\mathbf{x}_t]_{\mathbf{p}_0, \mathbf{M}}^{(i-1)}(\mathbf{m}_{(i)})), \end{aligned}$$

from which it follows that

$$\begin{aligned} \tilde{\mathbf{Y}}(t) &= [\mathbf{r}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, [\mathbf{x}_t]'(\mathbf{p}_0; \mathbf{M}))), \\ &= [\mathbf{r}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, \tilde{\mathbf{X}}(t))), \quad \forall t \in [t_0, t_f]. \end{aligned} \quad (12)$$

(4) and the definition of the LD-derivative imply that, for each $t \in [t_0, t_f]$,

$$\mathbf{N} := \mathbf{r}'(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{0}_{1 \times k}, \mathbf{M}, [\mathbf{x}_t]'(\mathbf{p}_0; \mathbf{M}))) = [\mathbf{r}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, \tilde{\mathbf{X}}(t)))$$

is the unique solution of

$$\begin{aligned} \mathbf{0}_{n_y \times k} &= \mathbf{g}'(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0), \mathbf{y}(t, \mathbf{p}_0); (\mathbf{0}_{1 \times k}, \mathbf{M}, [\mathbf{x}_t]'(\mathbf{p}_0; \mathbf{M}), \mathbf{N})), \\ &= [\mathbf{g}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0), \mathbf{y}(t, \mathbf{p}_0); (\mathbf{M}, \tilde{\mathbf{X}}(t), \mathbf{N})). \end{aligned}$$

Hence,

$$\mathbf{0}_{n_y \times k} = [\mathbf{g}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0), \mathbf{y}(t, \mathbf{p}_0); (\mathbf{M}, \tilde{\mathbf{X}}(t), \tilde{\mathbf{Y}}(t))), \quad \forall t \in [t_0, t_f]. \quad (13)$$

For each $t \in [t_0, t_f]$, the L-smoothness of \mathbf{y}_t was established earlier; the LD-derivative chain rule yields

$$[\mathbf{y}_t]'(\mathbf{p}_0; \mathbf{M}) = [\mathbf{r}_t]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, [\mathbf{x}_t]'(\mathbf{p}_0; \mathbf{M}))) = \tilde{\mathbf{Y}}(t), \quad \forall t \in [t_0, t_f].$$

Evaluation of (13) at $t = t_0$ yields the fact that $\tilde{\mathbf{Y}}(t_0)$ is the unique solution $\mathbf{Y}_0 \in \mathbb{R}^{n_y \times k}$ of the equation system

$$\mathbf{0}_{n_y \times k} = [\mathbf{g}_{t_0}]'(\mathbf{p}_0, \mathbf{x}_0, \mathbf{y}_0; (\mathbf{M}, [\mathbf{f}_0]'(\mathbf{p}_0; \mathbf{M}), \mathbf{Y}_0)),$$

since $\mathbf{x}(t_0, \mathbf{p}_0) = \mathbf{x}_0$, $\mathbf{y}(t_0, \mathbf{p}_0) = \mathbf{y}_0$, and $\tilde{\mathbf{X}}(t_0) = [\mathbf{f}_0]'(\mathbf{p}_0; \mathbf{M})$. The conclusion of the theorem holds by virtue of (10)-(13) and the observation that $\mathbf{y}(t, \mathbf{p}_0) = \mathbf{r}(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0))$, $\tilde{\mathbf{X}}(t) = [\mathbf{x}_t]'(\mathbf{p}_0, \mathbf{M})$, and $\tilde{\mathbf{Y}}(t) = [\mathbf{y}_t]'(\mathbf{p}_0, \mathbf{M})$ hold for all $t \in [t_0, t_f]$. \square

Remark 4.1 If \mathbf{f} , \mathbf{g} , and \mathbf{f}_0 are C^1 on their respective domains, then $Z_{\mathbf{f}} = \emptyset$ and, as expected, (6) simplifies to

$$\begin{aligned} \dot{\mathbf{X}}(t) &= \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \mathbf{M} + \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{X}(t) + \frac{\partial \mathbf{f}}{\partial \mathbf{y}} \mathbf{Y}(t), \\ \mathbf{0}_{n_y \times k} &= \frac{\partial \mathbf{g}}{\partial \mathbf{p}} \mathbf{M} + \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \mathbf{X}(t) + \frac{\partial \mathbf{g}}{\partial \mathbf{y}} \mathbf{Y}(t), \\ \mathbf{X}(t_0) &= \mathbf{J}\mathbf{f}_0(\mathbf{p}_0)\mathbf{M}, \end{aligned}$$

where the partial derivatives of \mathbf{f} and \mathbf{g} are evaluated at $(t, \mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0), \mathbf{y}(t, \mathbf{p}_0))$, which has been omitted for brevity.

Remark 4.2 Given a regular solution \mathbf{z} of (5) on $[t_0, t_f] \times \{\mathbf{p}_0\}$ through $\{(t_0, \mathbf{p}_0, \mathbf{x}_0, \mathbf{y}_0)\}$ and any nonsingular $\mathbf{M} \in \mathbb{R}^{n_p \times n_p}$, $(\mathbf{X}(t_f), \mathbf{Y}(t_f)) := [\mathbf{z}_{t_f}]'(\mathbf{p}_0; \mathbf{M})$ can be obtained by evaluating the unique solution of the auxiliary nonsmooth DAE system (6) at $t = t_f$.

As an element of the lexicographic subdifferential,

$$\begin{bmatrix} \mathbf{J}_L \mathbf{x}_{t_f}(\mathbf{p}_0; \mathbf{M}) \\ \mathbf{J}_L \mathbf{y}_{t_f}(\mathbf{p}_0; \mathbf{M}) \end{bmatrix} = \mathbf{J}_L \mathbf{z}_{t_f}(\mathbf{p}_0; \mathbf{M})$$

is a computationally relevant object related to the parametric sensitivities of the differential variables \mathbf{x} and algebraic variables \mathbf{y} , respectively, at $t = t_f$. It can be furnished by solving the following linear equation system:

$$\begin{bmatrix} \mathbf{X}(t_f) \\ \mathbf{Y}(t_f) \end{bmatrix} = \begin{bmatrix} \mathbf{J}_L \mathbf{x}_{t_f}(\mathbf{p}_0; \mathbf{M}) \\ \mathbf{J}_L \mathbf{y}_{t_f}(\mathbf{p}_0; \mathbf{M}) \end{bmatrix} \mathbf{M}.$$

Remark 4.3 Mirroring the discussion in Remark 2.1, the right-hand side function $(t, \mathbf{A}) \mapsto [\bar{\mathbf{f}}_{t \setminus z_t}]'(\mathbf{p}_0, \mathbf{x}(t, \mathbf{p}_0); (\mathbf{M}, \mathbf{A}))$ in (9) need not satisfy the Carathéodory conditions, but the k columns of the matrix-valued function $t \mapsto [\mathbf{x}_t]'$ are nonetheless absolutely continuous on $[t_0, t_f]$. However, the k columns of the matrix-valued function $t \mapsto [\mathbf{y}_t]'$ are Lebesgue integrable vector-valued functions mapping $[t_0, t_f]$ to \mathbb{R}^{n_y} , and therefore may exhibit discontinuities with respect to the independent variable (as illustrated in Example 5.2).

5 Examples

In this section, examples are provided to highlight the theory.

Example 5.1 Consider the following nonsmooth parametric DAEs:

$$\begin{aligned} \dot{x}(t, p) &= 0.5 \operatorname{sign}(1-t) \max\{0, p\} y(t, p), \\ 0 &= |x(t, p)| + |y(t, p)| - 1, \\ x(0, p) &= \arctan(p), \end{aligned} \tag{14}$$

where $\text{sign}(\cdot)$ denotes the signum function. Let $p_0 := 0$, $N(p_0) :=]-0.5, 0.5[$, $x_0 := 0$, and $y_0 := 1$. There exists a unique solution $\mathbf{z} \equiv (x, y)$ of (14) on $[0, 2] \times N(p_0)$ through

$$\Omega_0 := \{(t, p, \eta_x, \eta_y) : t = 0, p \in N(p_0), \eta_x = \arctan(p), \eta_y = 1 - |\arctan(p)|\}$$

which is given by

$$\mathbf{z} : (t, p) \mapsto \begin{cases} \begin{bmatrix} (\arctan(p) - 1) \exp(-0.5pt) + 1 \\ (1 - \arctan(p)) \exp(-0.5pt) \end{bmatrix}, & \text{if } (t, p) \in [0, 1[\times]0, 0.5[, \\ \begin{bmatrix} (\beta(p) - 1) \exp(0.5p(t-1)) + 1 \\ (1 - \beta(p)) \exp(0.5p(t-1)) \end{bmatrix}, & \text{if } (t, p) \in [1, 2] \times]0, 0.5[, \\ \begin{bmatrix} \arctan(p) \\ 1 + \arctan(p) \end{bmatrix}, & \text{if } (t, p) \in [0, 2] \times]-0.5, 0], \end{cases}$$

where $\beta :]0, 0.5[\rightarrow]0, 1[: p \mapsto (\arctan(p) - 1) \exp(-0.5p) + 1$. The solution is regular as $\pi_4 \partial g(t, p, x(t, p), y(t, p)) = \{1\}$ for all $(t, p) \in [0, 2] \times N(p_0)$, since $y(t, p) > 0$ for all $(t, p) \in [0, 2] \times N(p_0)$. Note that $\mathbf{z}(t, 0) = (x(t, 0), y(t, 0)) = (0, 1)$ for all $t \in [0, 2]$. For any $\mathbf{d} := (d_1, d_2, d_3) \in \mathbb{R}^3$, $[f_0]'(0; d_1) = d_1$,

$$[f_{i \setminus \{1\}}]'(0, \mathbf{z}(t, 0); \mathbf{d}) = \begin{cases} 0.5 \max\{0, d_1\}, & \text{if } t \in [0, 1[, \\ 0, & \text{if } t = 1, \\ -0.5 \max\{0, d_1\}, & \text{if } t \in]1, 2], \end{cases}$$

$$[g_i]'(0, \mathbf{z}(t, 0); \mathbf{d}) = |d_2| + d_3.$$

By Theorem 4.1, $\mathbf{z}_t \equiv \mathbf{z}(t, \cdot)$ is L-smooth at \mathbf{p}_0 for each $t \in [t_0, t_f]$; for any $m \in \mathbb{R}$, the LD-derivative mapping $\tilde{\mathbf{Z}} \equiv (\tilde{X}, \tilde{Y}) : t \mapsto [\mathbf{z}_t]'(0; m)$ is the unique solution (in the

sense of Theorem 4.1) of the following DAE system:

$$\begin{aligned} \dot{X}(t) &= 0.5 \operatorname{sign}(1-t) \max\{0, m\}, \\ 0 &= |X(t)| + Y(t), \\ X(0) &= m, \end{aligned} \tag{15}$$

on $[0, 2]$ through $\{(0, m, -|m|)\}$. Observe that the initial condition $Y(0)$ in (15) is uniquely determined from $X(0)$ (unlike in (14)), in accordance with Theorem 4.1. For any $m \neq 0$, post-multiplying the unique solution $(\tilde{X}(t), \tilde{Y}(t))$ of (15) by m^{-1} yields:

$$\mathbf{J}_L \mathbf{z}_t(0; m) = \begin{cases} \{(0.5t + 1, -0.5t - 1)\}, & \text{if } t \in [0, 1[, m > 0, \\ \{(-0.5t + 2, 0.5t - 2)\}, & \text{if } t \in [1, 2], m > 0, \\ \{(1, 1)\}, & \text{if } t \in [0, 2], m < 0, \end{cases}$$

so that

$$\partial_L \mathbf{z}_t(0) = \begin{cases} \{(0.5t + 1, -0.5t - 1), (1, 1)\}, & \text{if } t \in [0, 1[, \\ \{(-0.5t + 2, 0.5t - 2), (1, 1)\}, & \text{if } t \in [1, 2]. \end{cases}$$

From the analytic solution, for each $t \in [0, 1[$,

$$\mathbf{J}_L \mathbf{z}_t(p) = \begin{cases} \begin{bmatrix} ((1 + p^2)^{-1} - 0.5t(\arctan(p) - 1)) \exp(-0.5pt) \\ -(1 + p^2)^{-1} - 0.5t(1 - \arctan(p)) \exp(-0.5pt) \end{bmatrix}, & \text{if } p > 0, \\ \begin{bmatrix} (1 + p^2)^{-1} \\ (1 + p^2)^{-1} \end{bmatrix}, & \text{if } p < 0, \end{cases}$$

and, for each $t \in [1, 2]$,

$$\mathbf{J}_{\mathbf{z}_t}(p) = \begin{cases} \begin{bmatrix} (\beta'(p) + 0.5(t-1)(\beta(p) - 1)) \exp(0.5p(t-1)) \\ (-\beta'(p) + 0.5(t-1)(1 - \beta(p))) \exp(0.5p(t-1)) \end{bmatrix}, & \text{if } p > 0, \\ \begin{bmatrix} (1+p^2)^{-1} \\ (1+p^2)^{-1} \end{bmatrix}, & \text{if } p < 0. \end{cases}$$

Observe that, for each $t \in [0, 2]$,

$$\partial_{\mathbf{L}} \mathbf{z}_t(0) \subset \partial_{\mathbf{B}} \mathbf{z}_t(0) = \begin{cases} \{(1 + 0.5t, -1 - 0.5t), (1, 1)\}, & \text{if } t \in [0, 1[, \\ \{(1.5 - 0.5(t-1), -1.5 + 0.5(t-1)), (1, 1)\}, & \text{if } t \in [1, 2], \end{cases}$$

which is as expected since \mathbf{z}_t is PC^1 on $N(p_0)$ for each $t \in [0, 2]$.

Example 5.2 Sensitivities of solutions of (5) with respect to initial data are easily computed by Theorem 4.1. Suppose that \mathbf{z} is a unique regular solution of (5) (with no explicit parametric dependence in the right-hand side functions) on $[t_0, t_f] \times \{\mathbf{c}_0\}$ through $\{(t_0, \mathbf{c}_0, \mathbf{y}_0)\}$ for some $(\mathbf{c}_0, \mathbf{y}_0) \in D_x \times D_y$. Then, with analogous conditions to the hypotheses of Theorem 4.1, the nonsmooth DAE system (6) simplifies to

$$\begin{aligned} \dot{\mathbf{X}}(t) &= [\mathbf{f}_{t \setminus Z_t}]'(\mathbf{x}(t, \mathbf{c}_0), \mathbf{y}(t, \mathbf{c}_0); (\mathbf{X}(t), \mathbf{Y}(t))), \\ \mathbf{0}_{n_y \times k} &= [\mathbf{g}_t]'(\mathbf{x}(t, \mathbf{c}_0), \mathbf{y}(t, \mathbf{c}_0); (\mathbf{X}(t), \mathbf{Y}(t))), \\ \mathbf{X}(t_0) &= \mathbf{M}. \end{aligned} \tag{16}$$

As an illustration, consider the following IVP in DAEs:

$$\begin{aligned} \dot{x}_1(t, \mathbf{c}) &= 1 - y(t, \mathbf{c}), \\ \dot{x}_2(t, \mathbf{c}) &= x_2(t, \mathbf{c}), \\ 0 &= \max\{x_1(t, \mathbf{c}), x_2(t, \mathbf{c})\} + |y(t, \mathbf{c})| - 1, \\ x_1(0, \mathbf{c}) &= c_1, \\ x_2(0, \mathbf{c}) &= c_2. \end{aligned} \tag{17}$$

Let $\mathbf{c}_0 := (0, 0)$, $y_0 := 1$, and $[t_0, t_f] := [0, 1]$. Consider the parameter set

$$C := \{(c_1, c_2) \in \mathbb{R}^2 : 0 \leq c_1 < c_2 \leq 0.3\} \cup \{\mathbf{c}_0\}.$$

The unique solution $\mathbf{z} \equiv (\mathbf{x}, y)$ of (17) on $[0, 1] \times C$ through

$$\Omega_0 := \{(t, \eta_{x_1}, \eta_{x_2}, \eta_y) : t = 0, (\eta_{x_1}, \eta_{x_2}) \in C, \eta_y = 1 - \max\{\eta_{x_1}, \eta_{x_2}\}\}$$

is given by

$$\mathbf{z} : (t, \mathbf{c}) \mapsto \begin{cases} \begin{bmatrix} c_1 + c_2(1 - \exp(-t)) \\ c_2 \exp(-t) \\ 1 - c_2 \exp(-t) \end{bmatrix}, & \text{if } t \in [0, \tau(\mathbf{c})], \\ \begin{bmatrix} (c_1 + c_2(1 - \exp(-\tau(\mathbf{c}))) \exp(t - \tau(\mathbf{c})) \\ c_2 \exp(-t) \\ 1 - (c_1 + c_2(1 - \exp(-\tau(\mathbf{c}))) \exp(t - \tau(\mathbf{c}))) \end{bmatrix}, & \text{if } t \in]\tau(\mathbf{c}), 1], \end{cases}$$

where

$$\tau : C \rightarrow [0, 0.7[: (c_1, c_2) \mapsto \begin{cases} \ln\left(\frac{2c_2}{c_1 + c_2}\right), & \text{if } (c_1, c_2) \in C \setminus \{\mathbf{c}_0\}, \\ 0, & \text{if } (c_1, c_2) = \mathbf{c}_0. \end{cases}$$

The solution mapping \mathbf{z} is regular because $y(t, \mathbf{c}) > 0$ for all $(t, \mathbf{c}) \in [0, 1] \times C$ implies that $\pi_3 \partial \mathbf{g}(t, \mathbf{x}(t, \mathbf{c}), y(t, \mathbf{c})) = \{1\}$ for all $(t, \mathbf{c}) \in [0, 1] \times C$. In fact, there is a unique regular solution of (17) on $[0, 1] \times [-0.3, 0.3]^2$ through a superset of Ω_0 , which can be calculated by inspection and is PC^1 on its domain. However, its complete analytic expression is omitted here to make this example less cumbersome.

The right-hand side functions \mathbf{f} and g in (17) are C^1 and PC^1 on \mathbb{R}^3 , respectively.

Note that $\mathbf{z}(t, \mathbf{0}_2) = (\mathbf{x}(t, \mathbf{0}_2), y(t, \mathbf{0}_2)) = (0, 0, 1)$ for all $t \in [0, 1]$. Let

$$\mathbf{A} := \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \in \mathbb{R}^{3 \times 2}.$$

For any $t \in [0, 1]$ and any $\mathbf{d} := (d_1, d_2, d_3) \in \mathbb{R}^3$,

$$\begin{aligned} [g_t]_{\mathbf{z}(t, \mathbf{0}_2), \mathbf{A}}^{(0)}(\mathbf{d}) &= \lim_{\alpha \downarrow 0} \alpha^{-1} (\max\{\alpha d_1, \alpha d_2\} + |1 + \alpha d_3| - 1), \\ &= \max\{d_1, d_2\} + d_3, \end{aligned}$$

$$\begin{aligned} [g_t]_{\mathbf{z}(t, \mathbf{0}_2), \mathbf{A}}^{(1)}(\mathbf{d}) &= \lim_{\alpha \downarrow 0} \alpha^{-1} (\max\{a_{11} + \alpha d_1, a_{21} + \alpha d_2\} - \max\{a_{11}, a_{21}\} + \alpha d_3), \\ &= \begin{cases} d_1 + d_3, & \text{if } a_{11} > a_{21} \text{ or } a_{11} = a_{21} \text{ and } d_1 \geq d_2, \\ d_2 + d_3, & \text{if } a_{11} < a_{21} \text{ or } a_{11} = a_{21} \text{ and } d_1 < d_2. \end{cases} \end{aligned}$$

Therefore, for any $t \in [0, 1]$,

$$\begin{aligned} [\mathbf{f}_t]'(\mathbf{z}(t, \mathbf{0}_2); \mathbf{A}) &= \mathbf{J}\mathbf{f}_t(\mathbf{z}(t, \mathbf{0}_2))\mathbf{A} = \begin{bmatrix} -a_{31} & -a_{32} \\ -a_{21} & -a_{22} \end{bmatrix}, \\ [g_t]'(\mathbf{z}(t, \mathbf{0}_2); \mathbf{A}) &= \begin{cases} \begin{bmatrix} a_{11} + a_{31} & a_{12} + a_{32} \end{bmatrix}, & \text{if } a_{11} > a_{21} \text{ or } a_{11} = a_{21}, a_{12} \geq a_{22}, \\ \begin{bmatrix} a_{21} + a_{31} & a_{22} + a_{32} \end{bmatrix}, & \text{if } a_{11} < a_{21} \text{ or } a_{11} = a_{21}, a_{12} < a_{22}. \end{cases} \end{aligned}$$

Choose any directions matrix $\mathbf{M} \in \mathbb{R}^{2 \times 2}$ with entries satisfying

$$0 < m_{22} < m_{11} < m_{21} < m_{12} \leq 0.3$$

(which guarantees its nonsingularity). With these right-hand side functions, (16) ad-

mits the unique solution $\tilde{\mathbf{Z}} \equiv (\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) : t \mapsto [\mathbf{z}_t]'(\mathbf{c}_0; \mathbf{M})$ on $[0, 1]$ through

$$\{(0, \mathbf{M}, [-m_{21} \quad -m_{22}])\}$$

given by

$$\tilde{\mathbf{Z}} : t \mapsto \begin{bmatrix} m_{11} + m_{21}(1 - \exp(-t)) & m_{12} + m_{22}(1 - \exp(-t)) \\ m_{21} \exp(-t) & m_{22} \exp(-t) \\ -m_{21} \exp(-t) & -m_{22} \exp(-t) \end{bmatrix},$$

if $t \in [0, \tau(\mathbf{m}_{(1)})]$, and

$$\tilde{\mathbf{Z}} : t \mapsto \begin{bmatrix} \beta(\mathbf{m}_{(1)}) \exp(t - \tau(\mathbf{m}_{(1)})) & \gamma(\mathbf{m}_{(1)}, \mathbf{m}_{(2)}) \exp(t - \tau(\mathbf{m}_{(1)})) \\ m_{21} \exp(-t) & m_{22} \exp(-t) \\ -\beta(\mathbf{m}_{(1)}) \exp(t - \tau(\mathbf{m}_{(1)})) & -\gamma(\mathbf{m}_{(1)}, \mathbf{m}_{(2)}) \exp(t - \tau(\mathbf{m}_{(1)})) \end{bmatrix},$$

if $t \in]\tau(\mathbf{m}_{(1)}), 1]$, where

$$\beta : \mathbf{m}_{(1)} \mapsto m_{11} + m_{21}(1 - \exp(-\tau(\mathbf{m}_{(1)}))),$$

$$\gamma : (\mathbf{m}_{(1)}, \mathbf{m}_{(2)}) \mapsto m_{12} + m_{22}(1 - \exp(-\tau(\mathbf{m}_{(1)}))).$$

The mappings $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}_1$ are absolutely continuous on $[0, 1]$ but $\tilde{\mathbf{Y}}_2$ is not continuous at $\tau(\mathbf{m}_{(1)}) \in]0, 1[$. Post-multiplying $\tilde{\mathbf{Z}}(t_f)$ by \mathbf{M}^{-1} furnishes the following L-derivative:

$$\mathbf{J}_{L\mathbf{z}_{t_f}}(\mathbf{0}_2; \mathbf{M}) = \begin{bmatrix} \exp(1 - \tau(\mathbf{m}_{(1)})) & (1 - \exp(-\tau(\mathbf{m}_{(1)}))) \exp(1 - \tau(\mathbf{m}_{(1)})) \\ 0 & \exp(-1) \\ -\exp(1 - \tau(\mathbf{m}_{(1)})) & -(1 - \exp(-\tau(\mathbf{m}_{(1)}))) \exp(1 - \tau(\mathbf{m}_{(1)})) \end{bmatrix}.$$

From the analytic solution with $0 < c_1 < c_2 \leq 0.3$,

$$\mathbf{J}_{\mathbf{z}_{t_f}}(\mathbf{c}) = \begin{bmatrix} \exp(1 - \tau(\mathbf{c})) & (1 - \exp(-\tau(\mathbf{c})))\exp(1 - \tau(\mathbf{c})) \\ 0 & \exp(-1) \\ -\exp(1 - \tau(\mathbf{c})) & -(1 - \exp(-\tau(\mathbf{c})))\exp(1 - \tau(\mathbf{c})) \end{bmatrix}.$$

Let $\mathbf{c}_{(j)} := (m_{11}/j, m_{21}/j)$ for each $j \in \mathbb{N}$. Then $\tau(\mathbf{c}_{(j)}) = \tau(\mathbf{m}_{(1)})$ for each $j \in \mathbb{N}$ and $\lim_{j \rightarrow \infty} \mathbf{J}_{\mathbf{z}_{t_f}}(\mathbf{c}_{(j)}) = \mathbf{J}_{\mathbf{L}\mathbf{z}_{t_f}}(\mathbf{0}_2; \mathbf{M}) \in \partial_{\mathbf{L}}\mathbf{z}_{t_f}(\mathbf{0}_2) \subset \partial_{\mathbf{B}}\mathbf{z}_{t_f}(\mathbf{0}_2)$, as expected.

6 Conclusions

A theory to compute lexicographic derivatives of solutions of nonsmooth parametric DAEs has been developed. These generalized derivatives are computationally relevant and furnished via the solution of an auxiliary nonsmooth DAE system. The part of this solution mapping that is associated with the algebraic variables exhibits features that are unlike the original nonsmooth parametric DAEs of interest. Namely, it need not be continuous with respect to the independent variable and its initial condition is uniquely determined from the algebraic constraints of the auxiliary nonsmooth DAE system.

Forward sensitivity functions for Carathéodory index-1 semi-explicit DAEs have thus been characterized. Index refers here to a generalized differential index, which is formulated in terms of the projections of Clarke Jacobians being of maximal rank. Existence and regularity of a solution of the nonsmooth parametric DAEs need only be assumed on a finite horizon and at one parameter value for the theory to be applicable. This work is a natural extension of the classical sensitivity results for the analogous smooth case. Numerical solution of large-scale instances of the DAE system (6)

will require automatic methods for evaluation of the LD-derivatives appearing in (6), which is facilitated by a recently developed vector forward mode of automatic differentiation for LD-derivative evaluation [7]. Moreover, developing tractable methods for simulating the auxiliary nonsmooth DAE systems found here is an avenue for future work. Other possible directions for future work include extending the results to “high-index” nonsmooth DAEs and adjoint sensitivity results for nonsmooth DAEs.

Acknowledgments The authors would like to thank the anonymous reviewers for their helpful comments and Kamil A. Khan for his insights.

Funding: This research was financially supported by the Novartis-MIT Center for Continuous Manufacturing and the Natural Sciences and Engineering Research Council of Canada (NSERC).

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Kojima, M., Shindoh, S.: Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. *J. Oper. Res. Soc. Jpn.* **29**, 352–374 (1986)
2. Qi, L., Sun, J.: A nonsmooth version of Newton’s method. *Math. Program.* **58**, 353–367 (1993)
3. Facchinei, F., Fischer, A., Herrich, M.: An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. *Math. Program.* **146**, 1–36 (2014)
4. Kiwiel, K.C.: *Methods of Descent for Nondifferentiable Optimization*. Lecture Notes in Mathematics. Springer, Berlin (1985)

5. Lemaréchal, C., Strodiot, J.J., Bihain, A.: On a bundle algorithm for nonsmooth optimization. In: O.L. Mangasarian, R.R. Meyer, S.M. Robinson (eds.) *Nonlinear Programming 4*. Academic Press, New York (1981)
6. Lukšan, L., Vlček, J.: A bundle-Newton method for nonsmooth unconstrained minimization. *Math. Program.* **83**, 373–391 (1998)
7. Khan, K.A., Barton, P.I.: A vector forward mode of automatic differentiation for generalized derivative evaluation. *Optim. Methods Softw.* **30**(6), 1185–1212 (2015)
8. Nesterov, Y.: Lexicographic differentiation of nonsmooth functions. *Math. Program.* **104**, 669–700 (2005)
9. Scholtes, S.: *Introduction to Piecewise Differentiable Equations*. Springer, New York (2012)
10. Khan, K.A., Barton, P.I.: Generalized derivatives for solutions of parametric ordinary differential equations with non-differentiable right-hand sides. *J. Optim. Theory Appl.* **163**, 355–386 (2014)
11. Khan, K.A., Barton, P.I.: Generalized derivatives for hybrid systems. Submitted (2015)
12. Höffner, K., Khan, K.A., Barton, P.I.: Generalized derivatives of dynamic systems with a linear program embedded. *Automatica* **63**, 198–208 (2016)
13. Khan, K.A., Barton, P.I.: Generalized gradient elements for nonsmooth optimal control problems. In: *IEEE 53rd Annual Conference on Decision and Control (CDC)*, pp. 1887–1892. IEEE (2014)

14. Kunkel, P., Mehrmann, V.: *Differential-Algebraic Equations: Analysis and Numerical Solution*. European Mathematical Society, Zurich (2006)
15. Benyahia, B., Lakerveld, R., Barton, P.I.: A plant-wide dynamic model of a continuous pharmaceutical process. *Ind. Eng. Chem. Res.* **51**(47), 15,393–15,412 (2012)
16. Lakerveld, R., Benyahia, B., Heider, P., Zhang, H., Braatz, R.D., Barton, P.I.: Averaging level control to reduce off-spec material in a continuous pharmaceutical pilot plant. *Processes* **1**, 330–348 (2013)
17. Sahlodin, A.M., Barton, P.I.: Optimal campaign continuous manufacturing. *Ind. Eng. Chem. Res.* **54**(45), 11,344–11,359 (2015)
18. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. SIAM, Philadelphia (1990)
19. Facchinei, F., Pang, J.S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems: Volume II*. Springer, New York (2003)
20. Mordukhovich, B.S.: *Variational Analysis and Generalized Differentiation I: Basic Theory*. Springer, Berlin (2006)
21. Ascher, U.M., Petzold, L.R.: *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM, Philadelphia (1998)
22. Brenan, K.E., Campbell, S.L., Petzold, L.R.: *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. SIAM, Philadelphia (1996)
23. Petzold, L.: Differential/Algebraic equations are not ODE's. *SIAM J. Sci. Stat. Comput.* **3**(3), 367–384 (1982)
24. Rabier, P.J., Rheinboldt, W.C.: Theoretical and numerical analysis of differential-algebraic equations. *Handb. Numer. Anal.* **8**, 183–540 (2002)

25. Cao, Y., Li, S., Petzold, L., Serban, R.: Adjoint sensitivity analysis for differential-algebraic equations: the adjoint DAE system and its numerical solution. *SIAM J. Sci. Comp.* **24**(3), 1076–1089 (2003)
26. Feehery, W.F., Tolsma, J.E., Barton, P.I.: Efficient sensitivity analysis of large-scale differential-algebraic systems. *Appl. Numer. Math.* **25**, 41–54 (1997)
27. Barton, P.I., Lee, C.K.: Modeling, simulation, sensitivity analysis, and optimization of hybrid systems. *ACM Trans. Model. Comput. Simul.* **12**(4), 256–289 (2002)
28. Galán, S., Feehery, W.F., Barton, P.I.: Parametric sensitivity functions for hybrid discrete/continuous systems. *Appl. Numer. Math.* **31**, 17–47 (1999)
29. Ruban, A.I.: Sensitivity coefficients for discontinuous dynamic systems. *J. Comput. Syst. Sci. Int.* **36**(4), 536–542 (1997)
30. Khan, K.A., Saxena, V.P., Barton, P.I.: Sensitivity analysis of limit-cycle oscillating hybrid systems. *SIAM J. Sci. Comput.* **33**(4), 1475–1504 (2011)
31. Wilkins, K.A., Tidor, B., White, J.K., Barton, P.I.: Sensitivity analysis for oscillating dynamical systems. *SIAM J. Sci. Comput.* **31**(4), 2706–2732 (2009)
32. Pang, J.S., Stewart, D.E.: Solution dependence on initial conditions in differential variational inequalities. *Math. Program. B* **116**, 429–460 (2009)
33. Pang, J.S., Stewart, D.E.: Differential variational inequalities. *Math. Program. A* **113**, 345–424 (2008)
34. Hartman, P.: *Ordinary Differential Equations*. SIAM, Philadelphia (2002)
35. Stechliniski, P.G., Barton, P.I.: Well-posedness results for Carathéodory index-1 semi-explicit differential-algebraic equations. Tech. rep., Massachusetts Insti-

tute of Technology, Process Systems Engineering Laboratory, Cambridge, MA
(2016)

36. Sweetser, T.H.: A minimal set-valued strong derivative for vector-valued Lipschitz functions. *J. Optim. Theory Appl.* **23**(4), 549–562 (1977)
37. Imbert, C.: Support functions of the Clarke generalized Jacobian and of its plenary hull. *Nonlinear Anal.* **49**, 1111–1125 (2002)
38. Levy, A.B., Mordukhovich, B.S.: Coderivatives in parametric optimization. *Math. Program. A* **99**, 311–327 (2004)
39. Filippov, A.F.: *Differential Equations with Discontinuous Righthand Sides*. Kluwer Academic Publishers, Dordrecht (1988)