Generalized Derivatives for Solutions of Parametric Ordinary Differential Equations with Non-differentiable Right-Hand Sides

Journal of Optimization Theory and Applications manuscript No.
(will be inserted by the editor)

Generalized Derivatives for Solutions of Parametric Ordinary Differential Equations with Non-differentiable Right-Hand Sides

Kamil A. Khan · Paul I. Barton

Received: date / Accepted: date

Abstract Sensitivity analysis provides useful information for equation-solving, optimization, and post-optimality analysis. However, obtaining useful sensitivity information for systems with nonsmooth dynamic systems embedded is a challenging task. In this article, for any locally Lipschitz continuous mapping between finite-dimensional Euclidean spaces, Nesterov's lexicographic derivatives are shown to be elements of the plenary hull of the (Clarke) generalized Jacobian whenever they exist. It is argued that in applications, and in several established results in nonsmooth analysis, elements of the plenary hull of the generalized Jacobian of a locally Lipschitz continuous function are no less useful than elements of the generalized Jacobian itself. Directional derivatives and lexicographic derivatives of solutions of parametric ordinary differential equation (ODE) systems are expressed as the unique solutions of corresponding ODE systems, under Carathéodory-style assumptions. Hence, the scope of numerical methods for nonsmooth equation solving and local optimization is extended to systems with nonsmooth parametric ODEs embedded.

K.A. Khan · P.I. Barton (Corresponding Author)
Process Systems Engineering Laboratory, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
E-mail: pib@mit.edu

K.A. Khan
E-mail: kamil@mit.edu

Keywords Generalized Jacobians · Sensitivity analysis · Nonsmooth analysis · Ordinary differential equations

Mathematics Subject Classification (2000) 49J52 · 34A12 · 90C31

1 Introduction

For any locally Lipschitz continuous mapping between finite-dimensional Euclidean spaces, *Clarke's generalized Jacobian* [1] is a set-valued mapping that provides useful local sensitivity information. Elements of Clarke's generalized Jacobian are used in *semismooth Newton methods* for equation-solving [2,3], and in *bundle methods* for local optimization [4–6]. Methods have recently been developed to evaluate generalized Jacobian elements for finite compositions of simple smooth and nonsmooth functions [7,8]. However, there is currently no general method for determining generalized Jacobian elements for nonsmooth dynamic systems, which are defined in this article to be parametric Carathéodory ordinary differential equations (ODEs) with right-hand side functions that are not necessarily differentiable with respect to the dependent variables and parameters. These ODEs will be referred to as *nonsmooth parametric ODEs* throughout this article.

Classical results concerning parametric sensitivities of solutions of parametric ODEs require that the ODE right-hand side function has continuous partial derivatives, and imply differentiability of a unique solution with respect to the parameters [9]. These results can be extended to certain hybrid discrete/continuous dynamic systems, in which any discontinuities or kinks in an otherwise differentiable solution are defined as the solutions of equation systems with residual functions that are both continuously differentiable and locally invertible [10]. Nevertheless, Example A.1, in the appendix of this article, shows that a solution of a nonsmooth parametric ODE system is not necessarily differentiable with respect to the parameters. In this case, classical sensitivity results for parametric ODEs do not apply.

Even if the solutions of nonsmooth parametric ODEs are known to be smooth or convex functions of the ODE parameters, there is no general method for evaluating their gradients or subgradients. Such applications arise in global optimization of systems with nonconvex parametric ODE solutions embedded, where convex underestimators of these nonconvex ODE solutions have been described as solutions of corresponding nonsmooth parametric ODEs [11].

Clarke [1, Theorem 7.4.1] presents the primary existing result describing generalized Jacobians of parametric ODE solutions, in which certain supersets of generalized Jacobians of the ODE solutions are constructed. Using properties of these supersets, sufficient conditions for the differentiability of the original ODE solution have been formulated [1, 12].

Pang and Stewart [13, Theorem 11 and Corollary 12] show that when a parametric ODE has a right-hand side function that is *semismooth* in the sense of Qi [2], the generalized Jacobian supersets described by Clarke are in fact *linear Newton approximations* about any domain point. As summarized in Section 7.5.1 of [14], a linear Newton approximation for a locally Lipschitz continuous function about a domain point is a set-valued mapping containing local sensitivity information. Throughout this article, all discussed linear Newton approximations are linear Newton approximations about *every* domain point simultaneously; any reference to a linear Newton approximation of a function at a domain point refers to the value of this linear Newton approximation when evaluated at that domain point. Yunt [15] extends Pang and Stewart's result to adjoint sensitivities, systems described by index-1 differential-algebraic equations, and multi-stage systems with discontinuities in the right-hand side function occurring only at finitely many known values of the independent variable. However, Example A.2, in the appendix of this article, shows that linear Newton approximations are not guaranteed to satisfy certain properties that are satisfied by Clarke's generalized Jacobian. In particular, the linear Newton approximation of a continuously differentiable function at a domain point can include elements other than the derivative of the function at that point. Moreover, the linear Newton approximation of a convex scalar-valued function at a domain point can include elements that are not subgradients of the function at that point. Thirdly, given a convex scalar-valued function on an open set, the fact that the linear Newton approximation of the function at a domain point contains the origin is not a sufficient condition for a global minimum. Clarke's generalized Jacobian for a locally Lipschitz function, on the other hand, includes only the derivative whenever the function is continuously differentiable, and is identical to the convex subdifferential whenever the function is scalar-valued and convex [1]. In the latter case, the fact that the value of Clarke's generalized Jacobian at a domain point contains the origin is sufficient for a global minimum on an open set.

The plenary hull of Clarke's generalized Jacobian has been investigated in [16–18], and is referred to in this article as the *plenary Jacobian*. Though the plenary Jacobian is a superset of the generalized Jacobian, it satisfies several key nonsmooth analysis results in place of the generalized Jacobian. A benefit of the plenary Jacobian is that membership of the plenary Jacobian is easier to verify than membership of Clarke's generalized Jacobian. Moreover, it is argued in this work that the plenary Jacobian is in some sense as good a linear Newton approximation as the generalized Jacobian, and is just as useful in semismooth Newton methods and in bundle methods.

Sensitivities for unique solutions of a smooth parametric ODE system are traditionally expressed as the unique solutions of a corresponding linear ODE system obtained from the original system by application of the chain rule, as summarized in [9, Ch. V, Theorem 3.1]. In this spirit, the goal of this article is to present the first description of a plenary Jacobian element of the unique solution of a nonsmooth parametric ODE system as the unique solution of another ODE system. Nesterov's lexicographic derivatives [19] are used as a tool to construct this plenary Jacobian element.

This article is structured as follows. Section 2 summarizes relevant known mathematical results, and presents the argument that the plenary Jacobian is in some sense as useful as Clarke's generalized Jacobian. Section 3 presents new relations between various generalized derivatives for locally Lipschitz continuous functions, including the key result that any lexicographic derivative is a plenary Jacobian element. Section 4 expresses directional derivatives and lexicographic derivatives for solutions of nonsmooth parametric ODEs as the unique solutions of corresponding ODE systems. Various implications of these results are discussed.

2 Background

Relevant definitions and results from nonsmooth analysis are summarized in this section. These include properties of Nesterov's lexicographic derivatives [19], and of Clarke's generalized Jacobian [1] and its plenary hull [16].

Throughout this article, all vector spaces $\mathbb{R}^p$ are equipped with the Euclidean inner product and norm, and spaces $\mathbb{R}^{n\times p}$ of matrices are equipped with the corresponding induced norm. The *column space* of a matrix $\mathbf{M}\in\mathbb{R}^{n\times p}$ is defined as the set $\mathcal{R}(\mathbf{M}) := \{\mathbf{M}\mathbf{v} : \mathbf{v}\in\mathbb{R}^p\} \subset \mathbb{R}^n$. In the inductive proofs in this article, it will be convenient to refer to an *empty matrix* $\varnothing_{n\times 0}$ of real numbers, with n rows but no columns. In a further abuse of notation, the set $\{\varnothing_{n\times 0}\}$ will be denoted $\mathbb{R}^{n\times 0}$. No operations will be performed on $\varnothing_{n\times 0}$ beyond concatenation, which proceeds as expected:

$$\begin{bmatrix}\mathbf{A} & \varnothing_{n\times 0}\end{bmatrix} = \begin{bmatrix}\varnothing_{n\times 0} & \mathbf{A}\end{bmatrix} := \mathbf{A}, \qquad \forall \mathbf{A}\in\mathbb{R}^{n\times m}, \quad \forall m\in\mathbb{N}\cup\{0\}.$$

Given a collection of vectors $\mathbf{v}_{(1)},\mathbf{v}_{(2)},\dots,\mathbf{v}_{(p)}\in\mathbb{R}^n$, $\begin{bmatrix}\mathbf{v}_{(1)} & \cdots & \mathbf{v}_{(j)}\end{bmatrix}\in\mathbb{R}^{n\times j}$ will denote $\varnothing_{n\times 0}$ when $j=0$.

2.1 Set-Valued Mappings

As summarized by Facchinei and Pang [14], a *set-valued mapping* $F : Y \rightrightarrows Z$ is a function that maps each element of Y to a subset of Z. Suppose that $Y\subset\mathbb{R}^n$ is open and $Z=\mathbb{R}^m$. In this case, F is *upper-semicontinuous* at $\mathbf{y}\in Y$ iff for each $\varepsilon>0$, there exists $\delta>0$ such that whenever $\|\mathbf{z}\|<\delta$,

$$F(\mathbf{y}+\mathbf{z}) \subset F(\mathbf{y}) + \{\mathbf{v}\in\mathbb{R}^m : \|\mathbf{v}\| < \varepsilon\}.$$

If F is upper-semicontinuous at $\mathbf{y}\in Y$, then given any convergent sequences $\{\mathbf{y}^{(i)}\}_{i\in\mathbb{N}}$ in Y and $\{\mathbf{z}^{(i)}\}_{i\in\mathbb{N}}$ in $\mathbb{R}^m$ such that $\lim_{i\to\infty}\mathbf{y}^{(i)}=\mathbf{y}$, $\lim_{i\to\infty}\mathbf{z}^{(i)}=\mathbf{z}$, and $\mathbf{z}^{(i)}\in F(\mathbf{y}^{(i)})$ for each $i\in\mathbb{N}$, it follows that $\mathbf{z}\in F(\mathbf{y})$.

2.2 Directional Derivatives and Clarke's Generalized Jacobian

Given an open set $X \subset \mathbb{R}^n$, a function $\mathbf{f} : X \to \mathbb{R}^m$, some $\mathbf{x} \in X$, and some $\mathbf{d} \in \mathbb{R}^n$, the limit

$$\lim_{\tau \downarrow 0} \frac{\mathbf{f}(\mathbf{x} + \tau\mathbf{d}) - \mathbf{f}(\mathbf{x})}{\tau}$$

is called the *directional derivative* of $\mathbf{f}$ at $\mathbf{x}$ in the direction $\mathbf{d}$ iff it exists, and is denoted $\mathbf{f}'(\mathbf{x};\mathbf{d})$. The function $\mathbf{f}$ is *directionally differentiable* at $\mathbf{x}$ iff $\mathbf{f}'(\mathbf{x};\mathbf{d})$ exists and is finite for each $\mathbf{d} \in \mathbb{R}^n$.

As summarized by Scholtes [20], if $\mathbf{f}$ is directionally differentiable on its domain, then $\mathbf{f}'(\mathbf{x};\cdot)$ is positively homogeneous for each $\mathbf{x} \in X$. If, in addition, $\mathbf{f}$ is locally Lipschitz continuous on its domain, then

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - \mathbf{f}'(\mathbf{x};\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{0}, \qquad \forall \mathbf{x} \in X. \tag{1}$$

Moreover, for any fixed $\mathbf{x} \in \mathbb{R}^n$, the function $\mathbf{f}'(\mathbf{x};\cdot)$ is Lipschitz continuous on $\mathbb{R}^n$.

If $\mathbf{f}$ is (Fréchet) differentiable at some particular $\mathbf{x} \in X$, then the (Fréchet) derivative of $\mathbf{f}$ at $\mathbf{x}$ is denoted $\mathbf{Jf}(\mathbf{x}) \in \mathbb{R}^{m \times n}$. In this case, $\mathbf{f}$ is also directionally differentiable at $\mathbf{x}$, with $\mathbf{f}'(\mathbf{x};\mathbf{d}) = \mathbf{Jf}(\mathbf{x})\,\mathbf{d}$ for each $\mathbf{d} \in \mathbb{R}^n$.

Suppose that $\mathbf{f} : X \to \mathbb{R}^m$ is locally Lipschitz continuous, and let $Z_{\mathbf{f}} \subset X$ denote the set on which $\mathbf{f}$ is not differentiable. By the Rademacher Theorem, $Z_{\mathbf{f}}$ has zero (Lebesgue) measure. The *B-subdifferential* [14] of $\mathbf{f}$ at some particular $\mathbf{x} \in X$ is then:

$$\partial_{\mathrm{B}}\mathbf{f}(\mathbf{x}) := \left\{ \mathbf{H} \in \mathbb{R}^{m \times n} : \forall j \in \mathbb{N}, \exists \mathbf{x}^{(j)} \in X \backslash Z_{\mathbf{f}} \quad \text{s.t. } \lim_{j \to \infty} \mathbf{x}^{(j)} = \mathbf{x},\ \lim_{j \to \infty} \mathbf{Jf}(\mathbf{x}^{(j)}) = \mathbf{H} \right\}.$$

$\partial_{\mathrm{B}}\mathbf{f}(\mathbf{x})$ is necessarily compact and not empty. *Clarke's generalized Jacobian* [1] of $\mathbf{f}$ at $\mathbf{x}$ is then $\partial\mathbf{f}(\mathbf{x}) := \operatorname{conv} \partial_{\mathrm{B}}\mathbf{f}(\mathbf{x})$, which is compact, convex, and not empty. If $\mathbf{f}$ is differentiable at $\mathbf{x}$, then $\mathbf{Jf}(\mathbf{x}) \in \partial\mathbf{f}(\mathbf{x})$. If $\mathbf{f}$ is continuously differentiable at $\mathbf{x}$, then $\partial\mathbf{f}(\mathbf{x}) = \{\mathbf{Jf}(\mathbf{x})\}$. If $\mathbf{f}$ is scalar-valued and convex, then Clarke's generalized Jacobian of $\mathbf{f}$ coincides with the subdifferential from convex analysis. Considered as set-valued mappings, $\partial_{\mathrm{B}}\mathbf{f}$ and $\partial\mathbf{f}$ are both upper-semicontinuous.

2.3 Lexicographic Differentiation

Given an open set $X \subset \mathbb{R}^n$, a function $\mathbf{f}: X \to \mathbb{R}^m$ is *lexicographically smooth* [19] at $\mathbf{x} \in X$ iff it is Lipschitz continuous on a neighborhood of $\mathbf{x}$ and, for any $p \in \mathbb{N}$ and any matrix $\mathbf{M} = \begin{bmatrix}\mathbf{m}^{(1)} & \cdots & \mathbf{m}^{(p)}\end{bmatrix} \in \mathbb{R}^{n\times p}$, the following functions are well-defined:

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

The class of lexicographically smooth functions is closed under composition, and includes all continuously differentiable functions, all convex functions, and all piecewise differentiable functions [21] in the sense of Scholtes [20]. This represents a broad class of nonsmooth functions. The following lemma summarizes some properties and relations involving the functions $\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}$.

Lemma 2.1 *Given an open set $X \subset \mathbb{R}^n$, some $\mathbf{x} \in X$, a function $\mathbf{f}: X \to \mathbb{R}^n$ that is lexicographically smooth at $\mathbf{x}$, some $p \in \mathbb{N}$, and some $\mathbf{M} = \begin{bmatrix}\mathbf{m}^{(1)} & \cdots & \mathbf{m}^{(p)}\end{bmatrix} \in \mathbb{R}^{n\times p}$, the following properties are satisfied:*

1. $\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}(\tau\mathbf{d}) = \tau\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}(\mathbf{d}), \quad \forall \tau \geq 0, \quad \forall \mathbf{d} \in \mathbb{R}^n, \quad \forall j \in \{0,1,\ldots,p\}$,
2. $\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}(\mathbf{d}) = \mathbf{f}^{(j+1)}_{\mathbf{x},\mathbf{M}}(\mathbf{d}) = \cdots = \mathbf{f}^{(p)}_{\mathbf{x},\mathbf{M}}(\mathbf{d})$, $\forall \mathbf{d} \in \operatorname{span}\{\mathbf{m}^{(1)},\ldots,\mathbf{m}^{(j)}\}$, $\forall j \in \{1,\ldots,p\}$,
3. $\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}$ *is linear on* $\operatorname{span}\{\mathbf{m}^{(1)},\ldots,\mathbf{m}^{(j)}\}$ *for each* $j \in \{1,\ldots,p\}$,
4. $\mathbf{f}^{(j-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(j)}) = \mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(j)}) = \cdots = \mathbf{f}^{(p)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(j)}), \quad \forall j \in \{1,\ldots,p\}$,
5. *With* $\tilde{\mathbf{M}} := \begin{bmatrix}\mathbf{m}^{(1)} & \cdots & \mathbf{m}^{(j)} & \mathbf{A}\end{bmatrix}$ *for any particular* $j \in \{0,1,\ldots,p\}$, $q \in \mathbb{N}\cup\{0\}$, *and* $\mathbf{A} \in \mathbb{R}^{n\times q}$, $\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}(\mathbf{d}) = \mathbf{f}^{(j)}_{\mathbf{x},\tilde{\mathbf{M}}}(\mathbf{d})$ *for each* $\mathbf{d} \in \mathbb{R}^n$.

Proof Properties 1, 2, and 3 are demonstrated in [19, Lemma 3]. To obtain the leftmost equation in Property 4, combining the definition of $\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}$ with the positive homogeneity of $\mathbf{f}^{(j-1)}_{\mathbf{x},\mathbf{M}}$ implied by Property 1 yields

$$\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(j)}) = \lim_{\tau\downarrow 0} \frac{(1+\tau)\mathbf{f}^{(j-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(j)}) - \mathbf{f}^{(j-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(j)})}{\tau} = \mathbf{f}^{(j-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(j)}).$$

The remaining equations in Property 4 follow immediately from Property 2. Property 5 follows from the construction of the mappings $\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}$, noting that for each $\mathbf{d}\in\mathbb{R}^n$, $\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}(\mathbf{d})$ is independent of the rightmost $(p-j)$ columns of $\mathbf{M}$. □

Remark 2.1 It follows from Property 5 of Lemma 2.1 that for any $\mathbf{d}\in\mathbb{R}^n$, $\mathbf{f}^{(0)}_{\mathbf{x},\emptyset_{n\times 0}}(\mathbf{d})$ is well-defined, and equals $\mathbf{f}'(\mathbf{x};\mathbf{d})$.

If the columns of $\mathbf{M}\in\mathbb{R}^{n\times p}$ span $\mathbb{R}^n$, then Property 3 of Lemma 2.1 shows that $\mathbf{f}^{(p)}_{\mathbf{x},\mathbf{M}}$ is linear on $\mathbb{R}^n$. Thus, the following *lexicographic subdifferential* of $\mathbf{f}$ at $\mathbf{x}$ is well-defined and not empty:

$$\partial_{\mathrm{L}}\mathbf{f}(\mathbf{x}) := \{\mathbf{J}\mathbf{f}^{(n)}_{\mathbf{x},\mathbf{M}}(\mathbf{0})\in\mathbb{R}^{m\times n} : \mathbf{M}\in\mathbb{R}^{n\times n},\ \det\mathbf{M}\neq 0\}.$$

For any nonsingular $\mathbf{M}\in\mathbb{R}^{n\times n}$, let $\mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{x};\mathbf{M})$ denote the *lexicographic derivative* $\mathbf{J}\mathbf{f}^{(n)}_{\mathbf{x},\mathbf{M}}(\mathbf{0})\in\mathbb{R}^{m\times n}$ appearing in the above expression. In an abuse of notation, for any $\mathbf{M}\in\mathbb{R}^{n\times p}$ with $p\in\mathbb{N}$, let $\tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{f}(\mathbf{x};\mathbf{M})\in\mathbb{R}^{m\times n}$ denote *any* particular matrix for which $\mathbf{f}^{(p)}_{\mathbf{x},\mathbf{M}}(\mathbf{d}) = \tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{f}(\mathbf{x};\mathbf{M})\,\mathbf{d}$ for each $\mathbf{d}\in\mathcal{R}(\mathbf{M})$. Property 3 of Lemma 2.1 shows that at least one such matrix exists. This notation will only be used when the particular choice of matrix satisfying this description is irrelevant. Since $\mathcal{R}(\mathbf{M})$ contains each column of $\mathbf{M} := \begin{bmatrix}\mathbf{m}^{(1)} & \cdots & \mathbf{m}^{(p)}\end{bmatrix}$, it follows that

$$\tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{f}(\mathbf{x};\mathbf{M})\,\mathbf{M} = \begin{bmatrix}\mathbf{f}^{(p)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(1)}) & \cdots & \mathbf{f}^{(p)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(p)})\end{bmatrix}.$$

Thus, $\tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{f}(\mathbf{x};\mathbf{M})\,\mathbf{M}$ is uniquely defined for each $p\in\mathbb{N}$ and each $\mathbf{M}\in\mathbb{R}^{n\times p}$, even though $\tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{f}(\mathbf{x};\mathbf{M})$ may not be. If $\mathbf{M}\in\mathbb{R}^{n\times n}$ is nonsingular, then $\tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{f}(\mathbf{x};\mathbf{M}) = \mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{x};\mathbf{M})$.

According to Nesterov's chain rule for lexicographic derivatives [19, Theorem 5], if $X\subset\mathbb{R}^n$ and $Y\subset\mathbb{R}^m$ are open, and if functions $\mathbf{g}: X\to Y$ and $\mathbf{f}: Y\to\mathbb{R}^q$ are lexicographically smooth, then the composition $\mathbf{f}\circ\mathbf{g}$ is also lexicographically smooth. Moreover, for any nonsingular matrix $\mathbf{M}\in\mathbb{R}^{n\times n}$ and any $\mathbf{x}\in X$,

$$\mathbf{J}_{\mathrm{L}}[\mathbf{f}\circ\mathbf{g}](\mathbf{x};\mathbf{M}) = \tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{f}(\mathbf{g}(\mathbf{x});\mathbf{J}_{\mathrm{L}}\mathbf{g}(\mathbf{x};\mathbf{M})\,\mathbf{M})\;\mathbf{J}_{\mathrm{L}}\mathbf{g}(\mathbf{x};\mathbf{M}). \qquad (2)$$

(Note that the matrix $\mathbf{J}_{\mathrm{L}}\mathbf{g}(\mathbf{x};\mathbf{M})\,\mathbf{M}$ may be rectangular, and that its columns do not necessarily span $\mathbb{R}^m$.) This chain rule offers a means of evaluating a lexicographic derivative $\mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{x};\mathbf{M})$ for a composite function $\mathbf{f}$, without explicitly constructing the intermediate directional derivatives $\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}$. Our recent work [7] provides a computationally tractable method for evaluating generalized Jacobian elements for a broad class of piecewise differentiable functions. It has been shown that these generalized Jacobian elements are also lexicographic derivatives [21].

Unlike the B-subdifferential and Clarke's generalized Jacobian, $\partial_{\mathrm{L}}\mathbf{f}$ is not necessarily an upper-semicontinuous set-valued mapping [19, Example 1]. Nesterov [19, Equation 6.7] shows that when f is scalar-valued, $\partial_{\mathrm{L}} f(\mathbf{x}) \subset \partial f(\mathbf{x})$.

2.4 Plenary Hulls of Generalized Jacobians

The relevant properties of plenary sets and hulls were established by Sweetser [16]. A set $\mathcal{A} \subset \mathbb{R}^{m\times n}$ is *plenary* iff any matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ such that $\mathbf{M}\,\mathbf{d} \in \{\mathbf{H}\,\mathbf{d} : \mathbf{H} \in \mathcal{A}\}$ for all $\mathbf{d} \in \mathbb{R}^n$ is itself an element of $\mathcal{A}$. The intersection of plenary sets is itself plenary. Thus, the *plenary hull* of a set $\mathcal{M} \subset \mathbb{R}^{m\times n}$ is the intersection of all plenary supersets of $\mathcal{M}$, is itself plenary, and is denoted as $\mathrm{plen}\,\mathcal{M}$. It is possible for $\mathrm{plen}\,\mathcal{M}$ to be a strict superset of $\mathcal{M}$, even if $\mathcal{M}$ is both convex and compact.

Now, consider an open set $X \subset \mathbb{R}^n$ and a locally Lipschitz continuous function $\mathbf{f} : X \to \mathbb{R}^m$. The *plenary Jacobian* of $\mathbf{f}$ at some $\mathbf{x} \in X$, $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x}) := \mathrm{plen}\,\partial\mathbf{f}(\mathbf{x})$, has been investigated in [16–18,22]. $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ is convex, compact, and not empty [22], and satisfies:

$$\partial\mathbf{f}(\mathbf{x}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x}) \subset \prod_{i=1}^{m} \partial f_i(\mathbf{x}), \tag{3}$$

where either or both of the above inclusions may be strict. (The rightmost set above denotes the set of matrices $\mathbf{M}$ whose i^{th} row is an element of $\partial f_i(\mathbf{x})$, for every $i \in \{1,\dots,m\}$.) When $\min(m,n) = 1$, however, $\partial\mathbf{f}(\mathbf{x}) = \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$. Since the objective functions in nonlinear programs (NLPs) are scalar-valued, it follows that *bundle methods* for finding local minima for nonsmooth NLPs [4,5] are unaffected if the plenary Jacobian is used in place of Clarke's generalized Jacobian.

Since $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ is compact and $\partial\mathbf{f}(\mathbf{x})$ is both convex and compact, it follows immediately from (3) and [16, Lemma 3.1] that

$$\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x}) = \{\mathbf{M} \in \mathbb{R}^{m\times n} : \forall \mathbf{e} \in \mathbb{R}^n, \exists \mathbf{H} \in \partial\mathbf{f}(\mathbf{x}) \quad \text{s.t. } \mathbf{M}\,\mathbf{e} = \mathbf{H}\,\mathbf{e}\}. \tag{4}$$

The above equation will be used in the next section to determine whether particular matrices are elements of $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$. Combining the above equation with the inclusion $\partial\mathbf{f}(\mathbf{x}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ yields:

$$\{\mathbf{H}\,\mathbf{e} \in \mathbb{R}^m : \mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})\} = \{\mathbf{H}\,\mathbf{e} \in \mathbb{R}^m : \mathbf{H} \in \partial\mathbf{f}(\mathbf{x})\}, \qquad \forall \mathbf{e} \in \mathbb{R}^n. \tag{5}$$

The following proposition shows that if $m = n$, and if certain nonsingularity assumptions apply, then a similar relationship holds between images of inverses of elements of $\partial\mathbf{f}(\mathbf{x})$ and $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$. The condition that $\partial\mathbf{f}(\mathbf{x})$ does not contain any singular matrices is a key assumption in Clarke's inverse function theorem and implicit function theorem for locally Lipschitz continuous functions [1].

Proposition 2.1 *Given an open set $X \subset \mathbb{R}^n$ and a locally Lipschitz continuous function $\mathbf{f} : X \to \mathbb{R}^n$, suppose that for some $\mathbf{x} \in X$, $\partial\mathbf{f}(\mathbf{x})$ does not contain any singular matrices. Then $\{\mathbf{H}^{-1}\mathbf{e} \in \mathbb{R}^n : \mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})\} = \{\mathbf{H}^{-1}\mathbf{e} \in \mathbb{R}^n : \mathbf{H} \in \partial\mathbf{f}(\mathbf{x})\}$ for each $\mathbf{e} \in \mathbb{R}^n$.*

Proof Since $\partial\mathbf{f}(\mathbf{x})$ does not contain any singular matrices, [17, Proposition 3] implies that $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ does not contain any singular matrices either. Since $\partial\mathbf{f}(\mathbf{x}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$, the inclusion $\{\mathbf{H}^{-1}\mathbf{e} \in \mathbb{R}^n : \mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})\} \supset \{\mathbf{H}^{-1}\mathbf{e} \in \mathbb{R}^n : \mathbf{H} \in \partial\mathbf{f}(\mathbf{x})\}$ is trivial for each $\mathbf{e} \in \mathbb{R}^n$. To prove the reverse inclusion, choose any $\mathbf{e} \in \mathbb{R}^n$ and any $\mathbf{A} \in \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$. This implies that $\mathbf{A}$ is nonsingular. By (5),

$$\mathbf{e} = \mathbf{A}(\mathbf{A}^{-1}\mathbf{e}) \in \{\mathbf{H}(\mathbf{A}^{-1}\mathbf{e}) \in \mathbb{R}^n : \mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})\} = \{\mathbf{H}(\mathbf{A}^{-1}\mathbf{e}) \in \mathbb{R}^n : \mathbf{H} \in \partial\mathbf{f}(\mathbf{x})\}.$$

Thus, there exists $\mathbf{B} \in \partial\mathbf{f}(\mathbf{x})$ for which $\mathbf{e} = \mathbf{B}(\mathbf{A}^{-1}\mathbf{e})$. By the hypotheses of the proposition, $\mathbf{B}$ is nonsingular, and so $\mathbf{A}^{-1}\mathbf{e} = \mathbf{B}^{-1}\mathbf{e} \in \{\mathbf{H}^{-1}\,\mathbf{e} \in \mathbb{R}^n : \mathbf{H} \in \partial\mathbf{f}(\mathbf{x})\}$. $\square$

It follows that if the plenary Jacobian is used in place of Clarke's generalized Jacobian in a semismooth Newton method [2], then any sequence of iterates generated by the altered method can necessarily be generated using the original method. Similarly, it follows from (5) that if the plenary Jacobian is used in place of the generalized Jacobian in Clarke's mean value theorem [1, Proposition 2.6.5], then the result is unaffected. Since $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ contains a singular matrix if and only if $\partial\mathbf{f}(\mathbf{x})$ contains a singular matrix [17], Clarke's inverse function theorem [1, Theorem 7.1.1] for locally Lipschitz continuous functions is also unaffected if the generalized Jacobian is replaced with the plenary Jacobian.

As a set-valued mapping on X, $\partial_{\mathrm{P}}\mathbf{f}$ is upper-semicontinuous [17]. Thus, (5), [14, Definition 7.2.2], and [14, Definition 7.5.13] imply that $\partial_{\mathrm{P}}\mathbf{f}$ is a linear Newton approximation of $\mathbf{f}$ at any $\mathbf{x} \in X$. In light of the previous paragraph, $\partial_{\mathrm{P}}\mathbf{f}$ is in some sense as good a linear Newton approximation of $\mathbf{f}$ as $\partial\mathbf{f}$.

3 Relating Generalized Derivatives

Consider an open set $X \subset \mathbb{R}^n$, some $\mathbf{x} \in X$, and a function $\mathbf{f} : X \to \mathbb{R}^m$ that is both locally Lipschitz continuous and directionally differentiable. The main results of this section are the inclusions $\partial_{\mathrm{B}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ and $\partial_{\mathrm{L}}\mathbf{f}(\mathbf{x}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$, with the latter result assuming further that $\mathbf{f}$ is lexicographically smooth at $\mathbf{x}$. It follows immediately that any numerical or analytical method for evaluating an element of $\partial_{\mathrm{B}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0})$ or $\partial_{\mathrm{L}}\mathbf{f}(\mathbf{x})$ is also a method for evaluating an element of $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$.

Lemma 3.1 *Consider a function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ that is positively homogeneous and locally Lipschitz continuous*[1]*. For any $\mathbf{d} \in \mathbb{R}^n$, $\partial_{\mathrm{B}}\mathbf{f}(\mathbf{d}) \subset \partial_{\mathrm{B}}\mathbf{f}(\mathbf{0})$.*

Proof The result is trivial when $\mathbf{d} = \mathbf{0}$, so assume that $\mathbf{d} \neq \mathbf{0}$, and consider any particular $\mathbf{H} \in \partial_{\mathrm{B}}\mathbf{f}(\mathbf{d})$. By definition of the B-subdifferential, there exists a sequence $\{\mathbf{d}^{(i)}\}_{i\in\mathbb{N}}$ in $\mathbb{R}^n$ converging to $\mathbf{d}$, such that $\mathbf{f}$ is differentiable at each $\mathbf{d}^{(i)}$, and such that $\lim_{i\to\infty} \mathbf{Jf}(\mathbf{d}^{(i)}) = \mathbf{H}$. Since $\mathbf{d} \neq \mathbf{0}$ and $\lim_{i\to\infty} \mathbf{d}^{(i)} = \mathbf{d}$, it may be assumed without loss of generality that $\mathbf{d}^{(i)} \neq \mathbf{0}$ for all $i \in \mathbb{N}$. Making use of the positive homogeneity of $\mathbf{f}$, for each $i \in \mathbb{N}$, each $t > 0$, and each nonzero $\mathbf{h} \in \mathbb{R}^n$,

$$\frac{\mathbf{f}(t\mathbf{d}^{(i)} + \mathbf{h}) - \mathbf{f}(t\mathbf{d}^{(i)}) - \mathbf{Jf}(\mathbf{d}^{(i)})\,\mathbf{h}}{\|\mathbf{h}\|} = \frac{\mathbf{f}(\mathbf{d}^{(i)} + \frac{1}{t}\mathbf{h}) - \mathbf{f}(\mathbf{d}^{(i)}) - \mathbf{Jf}(\mathbf{d}^{(i)})\,(\frac{1}{t}\mathbf{h})}{\|\frac{1}{t}\mathbf{h}\|}$$

Noting that $\mathbf{f}$ is differentiable at $\mathbf{d}^{(i)}$ and taking the limit $\mathbf{h} \to \mathbf{0}$ yields:

$$\mathbf{0} = \lim_{\mathbf{h}\to\mathbf{0}} \frac{\mathbf{f}(\mathbf{d}^{(i)} + \frac{1}{t}\mathbf{h}) - \mathbf{f}(\mathbf{d}^{(i)}) - \mathbf{Jf}(\mathbf{d}^{(i)})\,(\frac{1}{t}\mathbf{h})}{\|\frac{1}{t}\mathbf{h}\|} = \lim_{\mathbf{h}\to\mathbf{0}} \frac{\mathbf{f}(t\mathbf{d}^{(i)} + \mathbf{h}) - \mathbf{f}(t\mathbf{d}^{(i)}) - \mathbf{Jf}(\mathbf{d}^{(i)})\,\mathbf{h}}{\|\mathbf{h}\|}.$$

[1] Though irrelevant to this lemma, if $k_{\mathbf{f}}$ is a Lipschitz constant for $\mathbf{f}$ in a neighborhood of $\mathbf{0}$, then $k_{\mathbf{f}}$ is a global Lipschitz constant for $\mathbf{f}$ [20].

Thus, for each $i \in \mathbb{N}$ and each $t > 0$, $\mathbf{f}$ is differentiable at $(t\mathbf{d}^{(i)})$, with a derivative of $\mathbf{Jf}(t\mathbf{d}^{(i)}) = \mathbf{Jf}(\mathbf{d}^{(i)})$. Since $\lim_{i\to\infty} \mathbf{Jf}(\mathbf{d}^{(i)}) = \mathbf{H}$, it follows that

$$\mathbf{H} = \lim_{i\to\infty} \mathbf{Jf}\left(\frac{\mathbf{d}^{(i)}}{2^i\|\mathbf{d}^{(i)}\|}\right).$$

Noting that $\lim_{i\to\infty}\left(\frac{\mathbf{d}^{(i)}}{2^i\|\mathbf{d}^{(i)}\|}\right) = \mathbf{0}$, it follows that $\mathbf{H} \in \partial_{\mathrm{B}}\mathbf{f}(\mathbf{0})$. □

Lemma 3.2 *Consider an open set $X \subset \mathbb{R}^n$, some $\mathbf{x} \in X$, and a function $\mathbf{f} : X \to \mathbb{R}^m$ that is locally Lipschitz continuous and directionally differentiable. If $\mathbf{f}'(\mathbf{x};\cdot)$ is differentiable at some particular $\mathbf{d} \in \mathbb{R}^n$, then $\mathbf{J}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{d}) \in \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$.*

Proof For notational simplicity, define $\mathbf{A} := \mathbf{J}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{d})$. The differentiability of $\mathbf{f}'(\mathbf{x};\cdot)$ at $\mathbf{d}$ implies that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\mathbf{f}'(\mathbf{x};\mathbf{d}+\mathbf{h}) - \mathbf{f}'(\mathbf{x};\mathbf{d}) - \mathbf{A}\mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0}. \tag{6}$$

To prove the lemma, the cases in which $\mathbf{d} = \mathbf{0}$ and $\mathbf{d} \neq \mathbf{0}$ will be considered separately. If $\mathbf{d} = \mathbf{0}$, then applying (6) and the positive homogeneity of $\mathbf{f}'(\mathbf{x};\cdot)$ yields $\mathbf{f}'(\mathbf{x};\mathbf{0}) = \mathbf{0}$, and

$$\mathbf{0} = \lim_{t\downarrow 0} \frac{\mathbf{f}'(\mathbf{x};t\mathbf{h}) - \mathbf{f}'(\mathbf{x};\mathbf{0}) - t\mathbf{A}\mathbf{h}}{t\|\mathbf{h}\|} = \frac{\mathbf{f}'(\mathbf{x};\mathbf{h}) - \mathbf{A}\mathbf{h}}{\|\mathbf{h}\|}, \qquad \forall \mathbf{h} \in \mathbb{R}^n\backslash\{\mathbf{0}\}.$$

Combining these statements, $\mathbf{f}'(\mathbf{x};\mathbf{h}) = \mathbf{A}\mathbf{h}$ for each $\mathbf{h} \in \mathbb{R}^n$. Hence, $\mathbf{f}$ is Gâteaux differentiable at $\mathbf{x}$, with a Gâteaux derivative of $\mathbf{A}$. Since Gâteaux and Fréchet differentiability are equivalent for locally Lipschitz continuous functions on $\mathbb{R}^n$ [1], it follows that $\mathbf{f}$ is Fréchet differentiable at $\mathbf{x}$, with $\mathbf{Jf}(\mathbf{x}) = \mathbf{A}$. Thus, $\mathbf{A} \in \partial\mathbf{f}(\mathbf{x}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$, as required.

Now consider the case in which $\mathbf{d} \neq \mathbf{0}$. Due to (4), it suffices to show that for any particular $\mathbf{e} \in \mathbb{R}^n$, $\mathbf{A}\mathbf{e} = \mathbf{H}\mathbf{e}$ for some $\mathbf{H} \in \partial\mathbf{f}(\mathbf{x})$. This statement is trivial when $\mathbf{e} = \mathbf{0}$, so assume that $\mathbf{e} \neq \mathbf{0}$. It follows from (6) that for any $\varepsilon > 0$, there exists some $\delta_\varepsilon > 0$ such that whenever $|\tau| < \delta_\varepsilon$,

$$\|\mathbf{f}'(\mathbf{x};\mathbf{d}+\tau\mathbf{e}) - \mathbf{f}'(\mathbf{x};\mathbf{d}) - \mathbf{A}(\tau\mathbf{e})\| < \varepsilon\|\tau\mathbf{e}\|.$$

It will be assumed that $\delta_\varepsilon < 1$ without loss of generality, since otherwise, setting $\delta_\varepsilon \leftarrow \min(\delta_\varepsilon, \frac{1}{2})$ does not affect the validity of the above statement. Since $\mathbf{f}'(\mathbf{x};\cdot)$ is positively homogeneous, multiplying both sides of the above inequality by any $\alpha > 0$ and setting $\tau := \frac{1}{2}\delta_\varepsilon$ yields:

$$\|\mathbf{f}'\big(\mathbf{x};\alpha(\mathbf{d}+\tfrac{1}{2}\delta_\varepsilon\mathbf{e})\big)-\mathbf{f}'(\mathbf{x};\alpha\mathbf{d})-\tfrac{1}{2}\alpha\delta_\varepsilon\mathbf{A}\mathbf{e}\| < \tfrac{1}{2}\varepsilon\alpha\delta_\varepsilon\|\mathbf{e}\|, \qquad \forall\alpha > 0. \quad (7)$$

It follows from (1) that for any $\varepsilon > 0$, there exists some $\bar{\delta}_\varepsilon > 0$ such that whenever $\|\mathbf{v}\| \leq \bar{\delta}_\varepsilon$, $\mathbf{f}$ is defined at $(\mathbf{x}+\mathbf{v})$, and

$$\|\mathbf{f}(\mathbf{x}+\mathbf{v})-\mathbf{f}(\mathbf{x})-\mathbf{f}'(\mathbf{x};\mathbf{v})\| < \varepsilon\delta_\varepsilon\|\mathbf{v}\|. \quad (8)$$

It will be assumed that $\lim_{\varepsilon\downarrow 0}\bar{\delta}_\varepsilon = 0$ without loss of generality, since otherwise, setting $\bar{\delta}_\varepsilon \leftarrow \min(\bar{\delta}_\varepsilon,\varepsilon)$ does not affect the validity of the above statement.

Now, choose any fixed $\varepsilon > 0$, and set

$$\alpha_\varepsilon := \frac{\bar{\delta}_\varepsilon}{\|\mathbf{d}\|+\frac{1}{2}\delta_\varepsilon\|\mathbf{e}\|} > 0.$$

The triangle inequality shows that for each $\tau \in [0,\frac{1}{2}\delta_\varepsilon]$,

$$\alpha_\varepsilon\|\mathbf{d}+\tau\mathbf{e}\| \leq \alpha_\varepsilon(\|\mathbf{d}\|+\tau\|\mathbf{e}\|) \leq \alpha_\varepsilon(\|\mathbf{d}\|+\tfrac{1}{2}\delta_\varepsilon\|\mathbf{e}\|) = \bar{\delta}_\varepsilon. \quad (9)$$

Thus, in (8), $\mathbf{v}$ may be set to $(\alpha_\varepsilon(\mathbf{d}+\tau\mathbf{e}))$ for any $\tau \in [0,\frac{1}{2}\delta_\varepsilon]$ to yield:

$$\|\mathbf{f}(\mathbf{x}+\alpha_\varepsilon(\mathbf{d}+\tau\mathbf{e}))-\mathbf{f}(\mathbf{x})-\mathbf{f}'(\mathbf{x};\alpha_\varepsilon(\mathbf{d}+\tau\mathbf{e}))\| < \varepsilon\delta_\varepsilon\alpha_\varepsilon\|\mathbf{d}+\tau\mathbf{e}\| \leq \varepsilon\delta_\varepsilon\bar{\delta}_\varepsilon, \quad (10)$$

Setting τ to 0 and $\frac{1}{2}\delta_\varepsilon$ in (10), respectively, yields:

$$\|\mathbf{f}(\mathbf{x})+\mathbf{f}'(\mathbf{x};\alpha_\varepsilon\mathbf{d})-\mathbf{f}(\mathbf{x}+\alpha_\varepsilon\mathbf{d})\| < \varepsilon\delta_\varepsilon\bar{\delta}_\varepsilon, \quad (11)$$

$$\|\mathbf{f}(\mathbf{x}+\alpha_\varepsilon(\mathbf{d}+\tfrac{1}{2}\delta_\varepsilon\mathbf{e}))-\mathbf{f}(\mathbf{x})-\mathbf{f}'\big(\mathbf{x};\alpha_\varepsilon(\mathbf{d}+\tfrac{1}{2}\delta_\varepsilon\mathbf{e})\big)\| < \varepsilon\delta_\varepsilon\bar{\delta}_\varepsilon. \quad (12)$$

Setting α to α_ε in (7), adding (11) and (12), and applying the triangle inequality yields:

$$\|\mathbf{f}(\mathbf{x}+\alpha_\varepsilon(\mathbf{d}+\tfrac{1}{2}\delta_\varepsilon\mathbf{e}))-\mathbf{f}(\mathbf{x}+\alpha_\varepsilon\mathbf{d})-\tfrac{1}{2}\alpha_\varepsilon\delta_\varepsilon\mathbf{A}\mathbf{e}\| < \varepsilon\delta_\varepsilon(\tfrac{1}{2}\alpha_\varepsilon\|\mathbf{e}\|+2\bar{\delta}_\varepsilon). \quad (13)$$

Now, Clarke's mean value theorem for locally Lipschitz continuous functions [1, Proposition 2.6.5] implies that

$$\begin{aligned}&\mathbf{f}(\mathbf{x}+\alpha_\varepsilon(\mathbf{d}+\tfrac{1}{2}\delta_\varepsilon\mathbf{e}))-\mathbf{f}(\mathbf{x}+\alpha_\varepsilon\mathbf{d})\\ &\qquad\in \operatorname{conv}\{\tfrac{1}{2}\alpha_\varepsilon\delta_\varepsilon\mathbf{H}\mathbf{e} : \exists\tau\in[0,\tfrac{1}{2}\delta_\varepsilon] \text{ s.t. } \mathbf{H}\in\partial\mathbf{f}(\mathbf{x}+\alpha_\varepsilon(\mathbf{d}+\tau\mathbf{e}))\}.\end{aligned}$$

Substituting this result into (13) and applying the Carathéodory Theorem yields the existence of $\lambda_\varepsilon^{(i)} \in [0,1]$, $\tau_\varepsilon^{(i)} \in [0, \frac{1}{2}\delta_\varepsilon]$, and $\mathbf{H}_\varepsilon^{(i)} \in \partial \mathbf{f}(\mathbf{x} + \alpha_\varepsilon(\mathbf{d} + \tau_\varepsilon^{(i)}\mathbf{e}))$ for each $i \in \{1, 2, \ldots, m+1\}$ such that:

$$1 = \sum_{i=1}^{m+1} \lambda_\varepsilon^{(i)}, \quad \text{and} \quad \left\| \frac{1}{2} \sum_{i=1}^{m+1} \lambda_\varepsilon^{(i)} \alpha_\varepsilon \delta_\varepsilon \mathbf{H}_\varepsilon^{(i)} \mathbf{e} - \frac{1}{2} \alpha_\varepsilon \delta_\varepsilon \mathbf{A}\mathbf{e} \right\| < \varepsilon \delta_\varepsilon (\tfrac{1}{2} \alpha_\varepsilon \|\mathbf{e}\| + 2\bar{\delta}_\varepsilon). \tag{14}$$

Dividing both sides of the above inequality by $\frac{1}{2}\alpha_\varepsilon \delta_\varepsilon$, applying the definition of α_ε, and noting that $\delta_\varepsilon < 1$ yields:

$$\left\| \sum_{i=1}^{m+1} \lambda_\varepsilon^{(i)} \mathbf{H}_\varepsilon^{(i)} \mathbf{e} - \mathbf{A}\mathbf{e} \right\| < \varepsilon \left(\|\mathbf{e}\| + \frac{4\bar{\delta}_\varepsilon}{\alpha_\varepsilon} \right) < \varepsilon (3\|\mathbf{e}\| + 4\|\mathbf{d}\|). \tag{15}$$

For each $\varepsilon > 0$ and each $i \in \{1, \ldots, m+1\}$, $\lambda_\varepsilon^{(i)}$ is an element of the compact set $[0,1] \subset \mathbb{R}$, and $\tau_\varepsilon^{(i)}$ is an element of the compact set $[0, \frac{1}{2}\delta_\varepsilon]$. Moreover, if $k_\mathbf{f}$ denotes a Lipschitz constant for $\mathbf{f}$ on $\{\mathbf{y} \in X : \|\mathbf{y} - \mathbf{x}\| \leq \bar{\delta}_1\}$, then, noting that $\lim_{\varepsilon \downarrow 0} \bar{\delta}_\varepsilon = 0$, it follows from (9) and [1, Proposition 2.6.2(d)] that for sufficiently small $\varepsilon > 0$, $\mathbf{H}_\varepsilon^{(i)}$ is an element of the compact set $\{\mathbf{H} \in \mathbb{R}^{m \times n} : \|\mathbf{H}\| \leq k_\mathbf{f}\}$ for each $i \in \{1, \ldots, m+1\}$.

Since any sequence in a compact set has a convergent subsequence, it follows that there exists a sequence $\{\varepsilon_j\}_{j \in \mathbb{N}}$ such that each $\varepsilon_j > 0$, $\lim_{j \to \infty} \varepsilon_j = 0$, and the sequences $\{\lambda_{\varepsilon_j}^{(i)}\}_{j \in \mathbb{N}}$, $\{\tau_{\varepsilon_j}^{(i)}\}_{j \in \mathbb{N}}$, and $\{\mathbf{H}_{\varepsilon_j}^{(i)}\}_{j \in \mathbb{N}}$ converge for each $i \in \{1, \ldots, m+1\}$, permitting the following definitions:

$$\bar{\lambda}^{(i)} := \lim_{j \to \infty} \lambda_{\varepsilon_j}^{(i)}, \qquad \bar{\tau}^{(i)} := \lim_{j \to \infty} \tau_{\varepsilon_j}^{(i)}, \qquad \text{and} \quad \bar{\mathbf{H}}^{(i)} := \lim_{j \to \infty} \mathbf{H}_{\varepsilon_j}^{(i)}.$$

It follows from (14) and (15) that

$$1 = \sum_{i=1}^{m+1} \bar{\lambda}^{(i)}, \qquad \text{and} \qquad \left\| \sum_{i=1}^{m+1} \bar{\lambda}^{(i)} \bar{\mathbf{H}}^{(i)} \mathbf{e} - \mathbf{A}\mathbf{e} \right\| = 0. \tag{16}$$

Since each $\tau_{\varepsilon_j}^{(i)} \in [0, \frac{1}{2}\delta_\varepsilon]$, applying (9) with $\tau := \tau_{\varepsilon_j}^{(i)}$ and $\varepsilon := \varepsilon_j$, taking the limit $j \to \infty$, and noting that $\lim_{\varepsilon \downarrow 0} \bar{\delta}_\varepsilon = 0$,

$$0 \leq \limsup_{j \to \infty} \left\| \alpha_{\varepsilon_j} (\mathbf{d} + \tau_{\varepsilon_j}^{(i)} \mathbf{e}) \right\| \leq \lim_{j \to \infty} \bar{\delta}_{\varepsilon_j} = 0.$$

Thus, for each $i \in \{1, \ldots, m+1\}$, $\lim_{j \to \infty} (\mathbf{x} + \alpha_{\varepsilon_j}(\mathbf{d} + \tau_{\varepsilon_j}^{(i)} \mathbf{e})) = \mathbf{x}$. Moreover, by construction,

$$\mathbf{H}_{\varepsilon_j}^{(i)} \in \partial \mathbf{f}(\mathbf{x} + \alpha_{\varepsilon_j}(\mathbf{d} + \tau_{\varepsilon_j}^{(i)} \mathbf{e})), \qquad \forall i \in \{1, \ldots, m+1\}, \qquad \forall j \in \mathbb{N}.$$

The upper-semicontinuity of Clarke's generalized Jacobian then yields $\bar{\mathbf{H}}^{(i)} \in \partial \mathbf{f}(\mathbf{x})$. Since $\partial \mathbf{f}(\mathbf{x})$ is convex and $\sum_{i=1}^{m+1} \bar{\lambda}^{(i)} = 1$, it follows that $\bar{\mathbf{H}} := \sum_{i=1}^{m+1} \bar{\lambda}^{(i)} \bar{\mathbf{H}}^{(i)} \in \partial \mathbf{f}(\mathbf{x})$. Moreover, (16) shows that $\bar{\mathbf{H}}\mathbf{e} = \mathbf{A}\mathbf{e}$, as required. □

Theorem 3.1 *Given an open set $X \subset \mathbb{R}^n$ and a function $\mathbf{f} : X \to \mathbb{R}^m$ that is locally Lipschitz continuous and directionally differentiable, $\partial_{\mathrm{B}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ for each $\mathbf{x} \in X$.*

Proof Consider any particular $\mathbf{x} \in X$ and any particular $\mathbf{H} \in \partial_{\mathrm{B}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0})$. By definition of the B-subdifferential, there exists a sequence $\{\mathbf{d}^{(i)}\}_{i\in\mathbb{N}}$ in $\mathbb{R}^n$ such that $\mathbf{f}'(\mathbf{x};\cdot)$ is differentiable at each $\mathbf{d}^{(i)}$, and $\lim_{i\to\infty} \mathbf{J}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{d}^{(i)}) = \mathbf{H}$. By Lemma 3.2, $\mathbf{J}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{d}^{(i)}) \in \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ for each $i \in \mathbb{N}$. Since $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ is a closed set [16], it follows that $\mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$. □

Corollary 3.1 *Given an open set $X \subset \mathbb{R}^n$ and a function $\mathbf{f} : X \to \mathbb{R}^m$ that is locally Lipschitz continuous and directionally differentiable,*

$$\partial_{\mathrm{B}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) \subset \partial[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) \subset \partial_{\mathrm{P}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x}), \qquad \forall \mathbf{x} \in X.$$

Proof Consider any particular $\mathbf{x} \in X$. The inclusions

$$\partial_{\mathrm{B}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) \subset \partial[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) \subset \partial_{\mathrm{P}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0})$$

follow immediately from the definitions of Clarke's generalized Jacobian and the plenary hull. Now, Theorem 3.1 yields the inclusion $\partial_{\mathrm{B}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$. Since $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ is convex [16], and since the convex hull of $\partial_{\mathrm{B}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0})$ is the intersection of all of its convex supersets in $\mathbb{R}^{m\times n}$, it follows that

$$\operatorname{conv} \partial_{\mathrm{B}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) = \partial[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x}).$$

Since $\partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ is plenary, and since the plenary hull of $\partial[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0})$ is the intersection of all of its plenary supersets in $\mathbb{R}^{m\times n}$, it follows that $\partial_{\mathrm{P}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$. The required chain of inclusions is therefore established. □

Corollary 3.2 *Given an open set $X \subset \mathbb{R}^n$ and a function $\mathbf{f} : X \to \mathbb{R}^m$ that is lexicographically smooth, $\partial_{\mathrm{L}}\mathbf{f}(\mathbf{x}) \subset \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$ for each $\mathbf{x} \in X$.*

Proof Consider any particular $\mathbf{x} \in X$ and any particular $\mathbf{H} \in \partial_{\mathrm{L}}\mathbf{f}(\mathbf{x})$. By definition of $\partial_{\mathrm{L}}\mathbf{f}(\mathbf{x})$, there exists some nonsingular matrix $\mathbf{M} := \begin{bmatrix}\mathbf{m}^{(1)} & \cdots & \mathbf{m}^{(n)}\end{bmatrix} \in \mathbb{R}^{n\times n}$ such that the following functions are well-defined:

$$\mathbf{f}^{(0)}_{\mathbf{x},\mathbf{M}} : \mathbb{R}^n \to \mathbb{R}^m : \mathbf{h} \mapsto \mathbf{f}'(\mathbf{x};\mathbf{h}),$$
$$\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}} : \mathbb{R}^n \to \mathbb{R}^m : \mathbf{h} \mapsto [\mathbf{f}^{(j-1)}_{\mathbf{x},\mathbf{M}}]'(\mathbf{m}^{(j)};\mathbf{h}), \qquad \forall j \in \{1,\ldots,n\},$$

and such that $\mathbf{f}^{(n)}_{\mathbf{x},\mathbf{M}}$ is linear (and therefore differentiable) on its domain, with a derivative of $\mathbf{J}\mathbf{f}^{(n)}_{\mathbf{x},\mathbf{M}}(\mathbf{0}) = \mathbf{H}$. As an intermediate result, it will be proved by induction on $k = n, (n-1), \ldots, 0$ that for each $k \in \{0,1,\ldots,n\}$, $\mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}^{(k)}_{\mathbf{x},\mathbf{M}}(\mathbf{0})$. For the base case, the differentiability of $\mathbf{f}^{(n)}_{\mathbf{x},\mathbf{M}}$ implies that $\mathbf{H} = \mathbf{J}\mathbf{f}^{(n)}_{\mathbf{x},\mathbf{M}}(\mathbf{0}) \in \partial\mathbf{f}^{(n)}_{\mathbf{x},\mathbf{M}}(\mathbf{0}) \subset \partial_{\mathrm{P}}\mathbf{f}^{(n)}_{\mathbf{x},\mathbf{M}}(\mathbf{0})$.

For the inductive step, suppose that for some $k \in \{1,\ldots,n\}$, $\mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}^{(k)}_{\mathbf{x},\mathbf{M}}(\mathbf{0})$. It follows from the construction of $\mathbf{f}^{(k)}_{\mathbf{x},\mathbf{M}}$ that $\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}$ is directionally differentiable. Moreover, it follows from repeated application of [20, Theorem 3.1.2] that $\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}$ is Lipschitz continuous. Thus, noting that $\mathbf{f}^{(k)}_{\mathbf{x},\mathbf{M}} \equiv [\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}]'(\mathbf{m}^{(k)};\cdot)$, Corollary 3.1 implies that $\partial_{\mathrm{P}}\mathbf{f}^{(k)}_{\mathbf{x},\mathbf{M}}(\mathbf{0}) \subset \partial_{\mathrm{P}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(k)})$. Applying the inductive assumption then yields:

$$\mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(k)}). \tag{17}$$

Now, Lemma 2.1 implies that $\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}$ is positively homogeneous, and so Lemma 3.1 yields

$$\partial_{\mathrm{B}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(k)}) \subset \partial_{\mathrm{B}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{0}) \subset \partial_{\mathrm{P}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{0}).$$

Since $\partial_{\mathrm{P}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{0})$ is convex, it follows that $\partial\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(k)}) \subset \partial_{\mathrm{P}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{0})$. Similarly, since $\partial_{\mathrm{P}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{0})$ is plenary, it follows that $\partial_{\mathrm{P}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{m}^{(k)}) \subset \partial_{\mathrm{P}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{0})$. Thus, (17) implies that $\mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}^{(k-1)}_{\mathbf{x},\mathbf{M}}(\mathbf{0})$, which completes the inductive step.

It follows from this inductive proof that $\mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}^{(0)}_{\mathbf{x},\mathbf{M}}(\mathbf{0}) = \partial_{\mathrm{P}}[\mathbf{f}'(\mathbf{x};\cdot)](\mathbf{0})$. A final application of Corollary 3.1 yields $\mathbf{H} \in \partial_{\mathrm{P}}\mathbf{f}(\mathbf{x})$. □

4 Generalized Derivatives for Solutions of Parametric ODEs

This section extends a result by Pang and Stewart [13] to show that when the right-hand side function of a Carathéodory ODE [23] is directionally differentiable with respect to the dependent variables, then directional derivatives of any ODE solution with respect to its initial condition can be expressed as the solution of another Carathéodory ODE. This result is in turn extended to show that if the right-hand side function of the original ODE is lexicographically smooth with respect to the dependent variables, then lexicographic derivatives of the ODE solution can also be expressed as the unique solution of another ODE. This latter ODE may be decoupled into a sequence of Carathéodory ODEs, but does not necessarily satisfy the Carathéodory assumptions itself.

4.1 Propagating Directional Derivatives

The following theorem extends a result by Pang and Stewart [13, Theorem 7] concerning directional derivatives of ODE solutions to the case in which direct dependence of the ODE right-hand side function on the independent variable is measurable but not necessarily continuous. This theorem and the subsequent corollary show that these directional derivatives uniquely solve a corresponding ODE whose right-hand side function may be discontinuous in the independent variable, even if the right-hand side function of the original ODE was continuous. Hence, allowing for discontinuous dependence of the original right-hand side function on the independent variable is essential when using these results in inductive proofs to describe higher-order directional derivatives and lexicographic derivatives of the ODE solution.

Theorem 4.1 *Given an open, connected set $X \subset \mathbb{R}^n$ and real numbers $t_0 < t_f$, suppose that a function $\mathbf{f} : [t_0, t_f] \times X \to \mathbb{R}^n$ satisfies the following conditions:*

- *the mapping $\mathbf{f}(\cdot, \mathbf{c}) : [t_0, t_f] \to \mathbb{R}^n$ is measurable for each $\mathbf{c} \in X$,*
- *for each $t \in [t_0, t_f]$ except in a zero-measure subset $Z_{\mathbf{f}}$, the mapping $\mathbf{f}(t, \cdot) : X \to \mathbb{R}^n$ is continuous and directionally differentiable,*
- *with $\mathbf{x}(\cdot, \mathbf{c})$ denoting any solution of the parametric ODE system:*

$$\frac{d\mathbf{x}}{dt}(t, \mathbf{c}) = \mathbf{f}(t, \mathbf{x}(t, \mathbf{c})), \qquad \mathbf{x}(t_0, \mathbf{c}) = \mathbf{c}, \tag{18}$$

  there exists a solution $\{\mathbf{x}(t, \mathbf{c}_0) : t \in [t_0, t_f]\} \subset X$ for some $\mathbf{c}_0 \in X$,

– *there exists an open set* $N \subset X$ *such that* $\{\mathbf{x}(t,\mathbf{c}_0) : t \in [t_0,t_f]\} \subset N$*, and such that there exist Lebesgue integrable functions* $k_{\mathbf{f}}, m_{\mathbf{f}} : [t_0,t_f] \to \mathbb{R}_+ \cup \{+\infty\}$ *for which*

$$\|\mathbf{f}(t,\mathbf{c})\| \leq m_{\mathbf{f}}(t), \qquad \forall t \in [t_0,t_f], \quad \forall \mathbf{c} \in N,$$

and

$$\|\mathbf{f}(t,\mathbf{c}_1) - \mathbf{f}(t,\mathbf{c}_2)\| \leq k_{\mathbf{f}}(t)\,\|\mathbf{c}_1 - \mathbf{c}_2\|, \qquad \forall t \in [t_0,t_f], \quad \forall \mathbf{c}_1,\mathbf{c}_2 \in N.$$

Then, for each $t \in [t_0,t_f]$*, the function* $\mathbf{x}_t \equiv \mathbf{x}(t,\cdot)$ *is well-defined and Lipschitz continuous on a neighborhood of* $\mathbf{c}_0$*, with a Lipschitz constant that is independent of* t*. Moreover,* $\mathbf{x}_t$ *is directionally differentiable at* $\mathbf{c}_0$ *for each* $t \in [t_0,t_f]$*, and for each* $\mathbf{d} \in \mathbb{R}^n$*, the mapping* $t \mapsto [\mathbf{x}_t]'(\mathbf{c}_0;\mathbf{d})$ *is the unique solution on* $[t_0,t_f]$ *of the ODE:*

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t}(t) = [\hat{\mathbf{f}}_t]'(\mathbf{x}(t,\mathbf{c}_0);\mathbf{y}(t)), \qquad \mathbf{y}(t_0) = \mathbf{d}, \tag{19}$$

where $\hat{\mathbf{f}}_t : X \to \mathbb{R}^n$ *is defined in terms of* $\mathbf{f}$ *as follows, and is directionally differentiable for each* $t \in [t_0,t_f]$*:*

$$\hat{\mathbf{f}}_t(\mathbf{c}) = \begin{cases} \mathbf{f}(t,\mathbf{c}), & \text{if } t \in [t_0,t_f]\backslash Z_{\mathbf{f}}, \\ \mathbf{0}, & \text{if } t \in Z_{\mathbf{f}}. \end{cases}$$

Proof By [23, Ch. 1, §1, Theorem 2], $\mathbf{x}(\cdot,\mathbf{c}_0)$ is the unique solution of (18) on $[t_0,t_f]$ with $\mathbf{c} = \mathbf{c}_0$. Consequently, by [23, Ch. 1, §1, Theorems 2 and 6], there exists a neighborhood $N_0 \subset N$ of $\mathbf{c}_0$ such that for each $\mathbf{c} \in N_0$, there exists a unique solution $\{\mathbf{x}(t,\mathbf{c}) : t \in [t_0,t_f]\} \subset N$ of (18).

To obtain the Lipschitz continuity of $\mathbf{x}(t,\cdot)$ near $\mathbf{c}_0$, choose any $t \in [t_0,t_f]$ and any $\mathbf{c}_1,\mathbf{c}_2 \in N_0$. Since the ODE solutions $\mathbf{x}(\cdot,\mathbf{c}_1)$ and $\mathbf{x}(\cdot,\mathbf{c}_2)$ exist on $[t_0,t_f]$, it follows that

$$\begin{aligned}\|\mathbf{x}(t,\mathbf{c}_1) - \mathbf{x}(t,\mathbf{c}_2)\| &= \left\|\mathbf{c}_1 + \int_{t_0}^{t} \mathbf{f}(s,\mathbf{x}(s,\mathbf{c}_1))\,\mathrm{d}s - \mathbf{c}_2 - \int_{t_0}^{t} \mathbf{f}(s,\mathbf{x}(s,\mathbf{c}_2))\,\mathrm{d}s\right\|, \\ &\leq \|\mathbf{c}_1 - \mathbf{c}_2\| + \int_{t_0}^{t} k_{\mathbf{f}}(s)\,\|\mathbf{x}(s,\mathbf{c}_1) - \mathbf{x}(s,\mathbf{c}_2)\|\,\mathrm{d}s.\end{aligned}$$

Let $k_{\mathbf{x}} := \exp\left(\int_{t_0}^{t_f} k_{\mathbf{f}}(s)\,\mathrm{d}s\right)$. Applying the version of Gronwall's Inequality described in Section 1 of [24], since the above inequality holds with any $\bar{t} \in [t_0,t]$ in place of t, it follows that

$$\|\mathbf{x}(t,\mathbf{c}_1) - \mathbf{x}(t,\mathbf{c}_2)\| \leq \|\mathbf{c}_1 - \mathbf{c}_2\| \exp\left(\int_{t_0}^{t} k_{\mathbf{f}}(s)\,\mathrm{d}s\right) \leq k_{\mathbf{x}}\|\mathbf{c}_1 - \mathbf{c}_2\|, \quad \forall \mathbf{c}_1,\mathbf{c}_2 \in N_0.$$

This demonstrates the Lipschitz continuity of $\mathbf{x}(t,\cdot)$ near $\mathbf{c}_0$ for each $t \in [t_0, t_f]$, with a Lipschitz constant $k_{\mathbf{x}}$ that is independent of t.

By construction of $\hat{\mathbf{f}}_t$, $\hat{\mathbf{f}}_t$ is directionally differentiable on its domain for each $t \in [t_0, t_f] \backslash Z_{\mathbf{f}}$. $\hat{\mathbf{f}}_t$ is the zero function on $Z_{\mathbf{f}}$, and is therefore also directionally differentiable for each $t \in Z_{\mathbf{f}}$. Hence, $\hat{\mathbf{f}}_t$ is directionally differentiable for each $t \in [t_0, t_f]$. The mapping $\mathbf{g} : [t_0, t_f] \times \mathbb{R}^n \to \mathbb{R}^n : (t, \mathbf{v}) \mapsto [\hat{\mathbf{f}}_t]'(\mathbf{x}(t, \mathbf{c}_0); \mathbf{v})$ is therefore well-defined, and is the right-hand side function of the ODE (19).

Now, choose any particular $\mathbf{v} \in \mathbb{R}^n$. Since $\mathbf{x}(\cdot, \mathbf{c}_0)$ is continuous on the compact set $[t_0, t_f]$, the set $\{\mathbf{x}(t, \mathbf{c}_0) : t \in [t_0, t_f]\} \subset N$ is compact, and does not contain any points in the closed set $(\mathbb{R}^n \backslash N)$. Thus, there exists $\delta > 0$ such that for any $\tau \in [0, \delta]$ and any $t \in [t_0, t_f]$, $(\mathbf{x}(t, \mathbf{c}_0) + \tau \mathbf{v}) \in N$; this is trivial when $N = \mathbb{R}^n$, and follows from [23, Ch. 2, §5, Lemma 1] otherwise. Since $\mathbf{x}(\cdot, \mathbf{c}_0)$ is continuous on $[t_0, t_f]$, [23, Ch. 1, §1, Lemma 1] shows that the mapping $t \mapsto \mathbf{f}(t, \mathbf{x}(t, \mathbf{c}_0) + \tau \mathbf{v})$ is measurable on $[t_0, t_f]$ for each $\tau \in [0, \delta]$. Thus, the mapping $t \mapsto \hat{\mathbf{f}}_t(\mathbf{x}(t, \mathbf{c}_0) + \tau \mathbf{v})$ is also measurable on $[t_0, t_f]$ for each $\tau \in [0, \delta]$.

For each $\tau \in]0, \delta]$, the previous paragraph implies that the following mapping is well-defined and measurable:

$$\boldsymbol{\gamma}_\tau : [t_0, t_f] \to \mathbb{R}^n : t \mapsto \frac{\hat{\mathbf{f}}_t(\mathbf{x}(t, \mathbf{c}_0) + \tau \mathbf{v}) - \hat{\mathbf{f}}_t(\mathbf{x}(t, \mathbf{c}_0))}{\tau}.$$

It follows from the directional differentiability of $\hat{\mathbf{f}}_t$ and the definition of $\mathbf{g}$ that for each $t \in [t_0, t_f]$, $\mathbf{g}(t, \mathbf{v}) = \lim_{\tau \downarrow 0} \boldsymbol{\gamma}_\tau(t)$. Noting that $\mathbf{v} \in \mathbb{R}^n$ was chosen arbitrarily, it follows that for each $\mathbf{v} \in \mathbb{R}^n$, the mapping $\mathbf{g}(\cdot, \mathbf{v})$ is the pointwise limit of a sequence of measurable functions, and is therefore measurable on $[t_0, t_f]$.

Now, define $Z_{k_{\mathbf{f}}} := \{t \in [t_0, t_f] : k_{\mathbf{f}}(t) = +\infty\}$. Since $k_{\mathbf{f}}$ is integrable on $[t_0, t_f]$, $Z_{k_{\mathbf{f}}}$ has zero measure. For each $t \in [t_0, t_f] \backslash (Z_{\mathbf{f}} \cup Z_{k_{\mathbf{f}}})$, the definition of $k_{\mathbf{f}}$ implies that $k_{\mathbf{f}}(t)$ is a finite Lipschitz constant for $\mathbf{f}_t$ near $\mathbf{x}(t, \mathbf{c}_0)$. Thus, [20, Theorem 3.1.2] implies that

$$\|\mathbf{g}(t, \mathbf{v}_1) - \mathbf{g}(t, \mathbf{v}_2)\| \leq k_{\mathbf{f}}(t)\, \|\mathbf{v}_1 - \mathbf{v}_2\|, \qquad \forall t \in [t_0, t_f] \backslash (Z_{\mathbf{f}} \cup Z_{k_{\mathbf{f}}}), \quad \forall \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n.$$

The above relationship still holds if $t \in Z_{\mathbf{f}}$, since $\mathbf{g}(t, \mathbf{v}) = \mathbf{0}$ for each $\mathbf{v} \in \mathbb{R}^n$ in this case. The relationship also holds if $t \in Z_{k_{\mathbf{f}}}$, since $k_{\mathbf{f}}(t) = +\infty$ in this case. Combining these cases,

$$\|\mathbf{g}(t,\mathbf{v}_1)-\mathbf{g}(t,\mathbf{v}_2)\| \leq k_{\mathbf{f}}(t)\,\|\mathbf{v}_1-\mathbf{v}_2\|, \qquad \forall t\in[t_0,t_f], \quad \forall \mathbf{v}_1,\mathbf{v}_2\in\mathbb{R}^n. \tag{20}$$

Choose some $\mathbf{d}\in\mathbb{R}^n$, and let $m_{\mathbf{y}} := \|\mathbf{d}\|\exp\left(\int_{t_0}^{t_f} k_{\mathbf{f}}(s)\,\mathrm{d}s\right)+\|\mathbf{d}\|+1$. Since $\mathbf{g}(t,\mathbf{0})=\mathbf{0}$ for each $t\in[t_0,t_f]$, it follows that whenever $\|\mathbf{v}\|<m_{\mathbf{y}}$,

$$\|\mathbf{g}(t,\mathbf{v})\| = \|\mathbf{g}(t,\mathbf{v})-\mathbf{g}(t,\mathbf{0})\| \leq k_{\mathbf{f}}(t)\,\|\mathbf{v}\| \leq k_{\mathbf{f}}(t)\,m_{\mathbf{y}}. \tag{21}$$

Thus, $\|\mathbf{g}\|$ is bounded above on $[t_0,t_f]\times\{\mathbf{v}\in\mathbb{R}^n : \|\mathbf{v}\|<m_{\mathbf{y}}\}$ by an integrable function of t. By Carathéodory's existence theorem [25, Ch. 2, Theorems 1.1 and 1.3], there exists a solution $\mathbf{y}$ of (19) on $[t_0,\bar{t}]$, where $\bar{t}$ is the least element of $]t_0,t_f]$ for which either $\bar{t}=t_f$, $\|\mathbf{y}(\bar{t})\|\geq m_{\mathbf{y}}$, or both. Now, for each $t\in[t_0,\bar{t}]$, (21) implies that

$$\|\mathbf{y}(t)\| = \left\|\mathbf{d}+\int_{t_0}^{t}\mathbf{g}(s,\mathbf{y}(s))\,\mathrm{d}s\right\| \leq \|\mathbf{d}\|+\int_{t_0}^{t}k_{\mathbf{f}}(s)\,\|\mathbf{y}(s)\|\,\mathrm{d}s.$$

Thus, Gronwall's Inequality [24] implies that

$$\|\mathbf{y}(\bar{t})\| \leq \|\mathbf{d}\|\exp\left(\int_{t_0}^{\bar{t}}k_{\mathbf{f}}(s)\,\mathrm{d}s\right) \leq \|\mathbf{d}\|\exp\left(\int_{t_0}^{t_f}k_{\mathbf{f}}(s)\,\mathrm{d}s\right) < m_{\mathbf{y}}.$$

Comparing this inequality with the definition of $\bar{t}$, it follows that $\bar{t}=t_f$, and so there exists a solution $\mathbf{y}$ of (19) on $[t_0,t_f]$. Moreover, (20), (21), and [23, Ch. 1, §1, Theorem 2] show that this solution is unique.

The remainder of this proof proceeds similarly to the proof of [13, Theorem 7]. For sufficiently small $\bar{\tau}>0$, $(\mathbf{c}_0+\tau\mathbf{d})\in N_0$. Thus, for each choice of $t\in[t_0,t_f]$ and $\tau\in]0,\bar{\tau}]$, let

$$\mathbf{e}_{\mathbf{x}}(t,\tau) := \frac{\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d})-\mathbf{x}(t,\mathbf{c}_0)}{\tau}-\mathbf{y}(t),$$

and

$$\mathbf{e}_{\mathbf{f}}(t,\tau) := \frac{\mathbf{f}(t,\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d}))-\mathbf{f}(t,\mathbf{x}(t,\mathbf{c}_0))}{\tau}-\mathbf{g}\left(t,\frac{\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d})-\mathbf{x}(t,\mathbf{c}_0)}{\tau}\right).$$

It follows from the established bounds that for each $t\in[t_0,t_f]$ and each $\tau\in]0,\bar{\tau}]$,

$$\begin{aligned}\|\mathbf{e}_{\mathbf{f}}(t,\tau)\| &\leq \left\|\frac{\mathbf{f}(t,\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d}))-\mathbf{f}(t,\mathbf{x}(t,\mathbf{c}_0))}{\tau}\right\|+\left\|\mathbf{g}\left(t,\frac{\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d})-\mathbf{x}(t,\mathbf{c}_0)}{\tau}\right)\right\|,\\ &\leq \frac{k_{\mathbf{f}}(t)}{\tau}\|\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d})-\mathbf{x}(t,\mathbf{c}_0)\|+\frac{k_{\mathbf{f}}(t)}{\tau}\|\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d})-\mathbf{x}(t,\mathbf{c}_0)\|,\\ &\leq 2k_{\mathbf{x}}k_{\mathbf{f}}(t)\|\mathbf{d}\|.\end{aligned} \tag{22}$$

Now, (20) and the definitions of $\mathbf{e_x}$ and $\mathbf{e_f}$ imply that for each $t \in [t_0, t_f]$ and $\tau \in]0, \bar{\tau}]$,

$$\begin{aligned}
\|\mathbf{e_x}(t,\tau)\| &= \left\| \int_{t_0}^{t} \left(\frac{\mathbf{f}(s,\mathbf{x}(s,\mathbf{c}_0+\tau\mathbf{d})) - \mathbf{f}(s,\mathbf{x}(s,\mathbf{c}_0))}{\tau} - \mathbf{g}(s,\mathbf{y}(s)) \right) \mathrm{d}s \right\|, \\
&= \left\| \int_{t_0}^{t} \left(\mathbf{e_f}(s,\tau) + \mathbf{g}\left(s, \frac{\mathbf{x}(s,\mathbf{c}_0+\tau\mathbf{d}) - \mathbf{x}(s,\mathbf{c}_0)}{\tau}\right) - \mathbf{g}(s,\mathbf{y}(s)) \right) \mathrm{d}s \right\|, \\
&\leq \int_{t_0}^{t} \left(\|\mathbf{e_f}(s,\tau)\| + k_{\mathbf{f}}(s)\|\mathbf{e_x}(s,\tau)\| \right) \mathrm{d}s.
\end{aligned}$$

Since $\|\mathbf{e_x}(\cdot,\tau)\|$ is continuous, it is bounded on the compact set $[t_0, t_f]$. Hence, the mapping $t \mapsto k_{\mathbf{f}}(t)\|\mathbf{e_x}(t,\tau)\|$ is integrable on $[t_0, t_f]$. This permits application of a variation [24, Theorem 2] of Gronwall's Inequality, which yields the following for any $t \in [t_0, t_f]$ and $\tau \in]0, \bar{\tau}]$:

$$0 \leq \|\mathbf{e_x}(t,\tau)\| \leq \int_{t_0}^{t} \|\mathbf{e_f}(s,\tau)\| \exp\left(\int_{s}^{t} k_{\mathbf{f}}(r)\,\mathrm{d}r \right) \mathrm{d}s \leq k_{\mathbf{x}} \int_{t_0}^{t} \|\mathbf{e_f}(s,\tau)\|\,\mathrm{d}s. \tag{23}$$

Substituting (22) into (23) for each $t \in [t_0, t_f]$ and $\tau \in]0, \bar{\tau}]$ yields $\|\mathbf{e_x}(t,\tau)\| \leq m_{\mathbf{e_x}}$, where

$$m_{\mathbf{e_x}} := 2\|\mathbf{d}\|(k_{\mathbf{x}})^2 \int_{t_0}^{t_f} k_{\mathbf{f}}(s)\,\mathrm{d}s.$$

Now, for each $t \in [t_0, t_f] \backslash (Z_{\mathbf{f}} \cup Z_{k_{\mathbf{f}}})$, the assumptions of the theorem imply that $\mathbf{f}(t,\cdot)$ is directionally differentiable and Lipschitz continuous on X, with a Lipschitz constant of $k_{\mathbf{f}}(t)$. Hence, (1) implies that for each $t \in [t_0, t_f] \backslash (Z_{\mathbf{f}} \cup Z_{k_{\mathbf{f}}})$, for each $\varepsilon > 0$, there exists some $\delta_{t,\varepsilon} > 0$ such that if $\|\mathbf{h}\| < \delta_{t,\varepsilon}$,

$$\|\mathbf{f}(t,\mathbf{x}(t,\mathbf{c}_0)+\mathbf{h}) - \mathbf{f}(t,\mathbf{x}(t,\mathbf{c}_0)) - \mathbf{g}(t,\mathbf{h})\| \leq \varepsilon\|\mathbf{h}\|.$$

Moreover, the established Lipschitz continuity of $\mathbf{x}(t,\cdot)$ on N_0 for each $t \in [t_0, t_f]$ implies that for any $\varepsilon > 0$, any $t \in [t_0, t_f] \backslash (Z_{\mathbf{f}} \cup Z_{k_{\mathbf{f}}})$, and any $\tau \in]0, \min(\bar{\tau}, \frac{\delta_{t,\varepsilon}}{k_{\mathbf{x}}\|\mathbf{d}\|+1})[$,

$$\|\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d}) - \mathbf{x}(t,\mathbf{c}_0)\| \leq k_{\mathbf{x}}\tau\|\mathbf{d}\| < \delta_{t,\varepsilon}.$$

Thus, if $t \in [t_0, t_f] \backslash (Z_{\mathbf{f}} \cup Z_{k_{\mathbf{f}}})$ and $0 < \tau < \min(\bar{\tau}, \frac{\delta_{t,\varepsilon}}{k_{\mathbf{x}}\|\mathbf{d}\|+1})$,

$$\begin{aligned}
&\|\mathbf{f}(t,\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d})) - \mathbf{f}(t,\mathbf{x}(t,\mathbf{c}_0)) - \mathbf{g}(t,\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d}) - \mathbf{x}(t,\mathbf{c}_0))\| \\
&\quad \leq \varepsilon\|\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d}) - \mathbf{x}(t,\mathbf{c}_0)\|.
\end{aligned}$$

Noting that $\mathbf{g}(t,\cdot)$ is positively homogeneous and that $\tau > 0$, dividing both sides of the above inequality by τ yields the following, for each $t \in [t_0,t_f]\backslash(Z_{\mathbf{f}} \cup Z_{k_{\mathbf{f}}})$, each $\varepsilon > 0$, and each $\tau \in]0, \min(\bar{\tau}, \frac{\delta_{t,\varepsilon}}{k_{\mathbf{x}}\|\mathbf{d}\|+1})[$:

$$\|\mathbf{e_f}(t,\tau)\| \leq \varepsilon \left\| \frac{\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d}) - \mathbf{x}(t,\mathbf{c}_0)}{\tau} \right\| = \varepsilon\|\mathbf{e_x}(t,\tau)+\mathbf{y}(t)\| < \varepsilon(m_{\mathbf{e_x}}+m_{\mathbf{y}}).$$

Thus, $\lim_{\tau\downarrow 0}\|\mathbf{e_f}(t,\tau)\| = 0$ for almost all $t \in [t_0,t_f]$. Using this limit and the bound (22), applying the dominated convergence theorem to (23) yields $\lim_{\tau\downarrow 0}\|\mathbf{e_x}(t,\tau)\| = 0$ for each $t \in [t_0,t_f]$. Hence,

$$\lim_{\tau\downarrow 0} \frac{\mathbf{x}(t,\mathbf{c}_0+\tau\mathbf{d}) - \mathbf{x}(t,\mathbf{c}_0)}{\tau} = \mathbf{y}(t), \qquad \forall t \in [t_0,t_f].$$

Noting that $\mathbf{d} \in \mathbb{R}^n$ was chosen arbitrarily, it follows that for each $\mathbf{d} \in \mathbb{R}^n$, the directional derivative $[\mathbf{x}_t]'(\mathbf{c}_0;\mathbf{d})$ exists and is finite for each $t \in [t_0,t_f]$, and so $\mathbf{x}_t$ is directionally differentiable at $\mathbf{c}_0$ for each $t \in [t_0,t_f]$. Moreover, the above equation shows that $t \mapsto [\mathbf{x}_t]'(\mathbf{c}_0;\mathbf{d})$ is the unique solution $\mathbf{y}$ of (19) on $[t_0,t_f]$. □

Corollary 4.1 *Under the assumptions of Theorem 4.1, and using the same notation as in the theorem, the mapping* $\mathbf{g} : [t_0,t_f] \times \mathbb{R}^n \to \mathbb{R}^n : (t,\mathbf{v}) \mapsto [\hat{\mathbf{f}}_t]'(\mathbf{x}(t,\mathbf{c}_0);\mathbf{v})$ *satisfies the following conditions:*

- *the mapping* $\mathbf{g}(\cdot,\mathbf{v}) : [t_0,t_f] \to \mathbb{R}^n$ *is measurable for each* $\mathbf{v} \in \mathbb{R}^n$*,*
- *for each* $t \in [t_0,t_f]$ *except in a zero-measure set* $Z_{\mathbf{g}}$*, the mapping* $\mathbf{g}(t,\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ *is defined and continuous,*
- *for each* $\mathbf{d} \in \mathbb{R}^n$*, there exists a solution* $\{\mathbf{z}(t,\mathbf{d}) : t \in [t_0,t_f]\} \subset \mathbb{R}^n$ *of the parametric ODE system:*
$$\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t}(t,\mathbf{d}) = \mathbf{g}(t,\mathbf{z}(t,\mathbf{d})), \qquad \mathbf{z}(t_0,\mathbf{d}) = \mathbf{d}.$$
- *for each* $\mathbf{d} \in \mathbb{R}^n$*, there exists an open set* $N_{\mathbf{g}}(\mathbf{d}) \subset \mathbb{R}^n$ *such that*
$$\{\mathbf{z}(t,\mathbf{d}) : t \in [t_0,t_f]\} \subset N_{\mathbf{g}}(\mathbf{d}),$$
and such that there exist Lebesgue integrable functions $k_{\mathbf{g}}, m_{\mathbf{g}} : [t_0,t_f] \to \mathbb{R}_+ \cup \{+\infty\}$ *for which*
$$\|\mathbf{g}(t,\mathbf{v})\| \leq m_{\mathbf{g}}(t), \qquad \forall t \in [t_0,t_f], \quad \forall \mathbf{v} \in N_{\mathbf{g}}(\mathbf{d}),$$

and

$$\|\mathbf{g}(t,\mathbf{v}_1)-\mathbf{g}(t,\mathbf{v}_2)\| \leq k_{\mathbf{g}}(t)\,\|\mathbf{v}_1-\mathbf{v}_2\|, \qquad \forall t \in [t_0,t_f], \quad \forall \mathbf{v}_1,\mathbf{v}_2 \in N_{\mathbf{g}}(\mathbf{d}).$$

If, in addition, the mapping $\mathbf{f}(t,\cdot):X\to\mathbb{R}^n$ is lexicographically smooth for each $t\in[t_0,t_f]\backslash Z_{\mathbf{f}}$, then the mapping $\mathbf{g}(t,\cdot):\mathbb{R}^n\to\mathbb{R}^n$ is lexicographically smooth for each $t\in[t_0,t_f]$. In this case, the set $Z_{\mathbf{g}}$ described above may be set to $\emptyset$.

Proof The measurability of $\mathbf{g}(\cdot,\mathbf{v})$ and the existence and Lipschitz continuity of $\mathbf{g}(t,\cdot)$ except on some zero-measure set $Z_{\mathbf{g}} \subset (Z_{\mathbf{f}} \cup Z_{k_{\mathbf{f}}})$ were established in the proof of Theorem 4.1. For any $\mathbf{d}\in\mathbb{R}^n$, setting $\mathbf{z}(\cdot,\mathbf{d})$ to be the unique solution $\mathbf{y}$ of (19) establishes the existence of the trajectory $\{\mathbf{z}(t,\mathbf{d}):t\in[t_0,t_f]\}$. The existence of a set $N_{\mathbf{g}}(\mathbf{d})$ and functions $k_{\mathbf{g}}$ and $m_{\mathbf{g}}$ satisfying the claimed properties follows from the proof of Theorem 4.1 as well, with the identifications $N_{\mathbf{g}}(\mathbf{d}):=\{\mathbf{v}\in\mathbb{R}^n:\|\mathbf{v}\|<m_{\mathbf{y}}\}$, $k_{\mathbf{g}}\equiv k_{\mathbf{f}}$ on $[t_0,t_f]$, and $m_{\mathbf{g}}:t\mapsto k_{\mathbf{f}}(t)\,m_{\mathbf{y}}$.

Now, suppose that the mapping $\mathbf{f}_t\equiv\mathbf{f}(t,\cdot):X\to\mathbb{R}^n$ is lexicographically smooth for each $t\in[t_0,t_f]\backslash Z_{\mathbf{f}}$. Choose any fixed $t\in[t_0,t_f]\backslash Z_{\mathbf{f}}$. The construction of $\mathbf{g}$ implies that $\mathbf{g}(t,\cdot)\equiv[\mathbf{f}_t]'(\mathbf{x}(t,\mathbf{c}_0);\cdot)$. Since $\mathbf{f}_t$ is lexicographically smooth on X, it follows that $\mathbf{g}(t,\cdot)$ is lexicographically smooth on $\mathbb{R}^n$. Now, choose any fixed $t\in Z_{\mathbf{f}}$. By construction of $\mathbf{g}$, $\mathbf{g}(t,\cdot)$ is the zero function, which is trivially lexicographically smooth. Combining these cases, $\mathbf{g}(t,\cdot)$ is lexicographically smooth on $\mathbb{R}^n$ for each $t\in[t_0,t_f]$. Since this demonstrates *a posteriori* that $\mathbf{g}(t,\cdot)$ is continuous on $\mathbb{R}^n$ for each $t\in[t_0,t_f]$, the set $Z_{\mathbf{g}}$ described in the statement of the corollary may be set to $\emptyset$. □

4.2 Propagating Lexicographic Derivatives

The following corollary extends the results of the previous subsection to describe the higher-order directional derivatives of the solution of a nonsmooth parametric ODE. The subsequent theorem uses this result to express the lexicographic derivatives of the unique solution of an ODE with a lexicographically smooth right-hand side function as the unique solution of another ODE. Some implications of this result are discussed.

Corollary 4.2 *Given an open, connected set $X\subset\mathbb{R}^n$ and real numbers $t_0<t_f$, suppose that a function $\mathbf{f}:[t_0,t_f]\times X\to\mathbb{R}^n$ satisfies the conditions of Theorem 4.1, and*

suppose in addition that $\mathbf{f}(t,\cdot)$ *is lexicographically smooth on* X *for each* $t \in [t_0,t_f]\backslash Z_{\mathbf{f}}$*. Then, for each* $t \in [t_0,t_f]$*, with the function* $\mathbf{x}_t \equiv \mathbf{x}(t,\cdot)$ *defined as in the statement of Theorem 4.1,* $\mathbf{x}_t$ *is lexicographically smooth at* $\mathbf{c}_0$*. Moreover, for each* $p \in \mathbb{N}$*, each* $\mathbf{M} := \left[\mathbf{m}^{(1)} \cdots \mathbf{m}^{(p)}\right] \in \mathbb{R}^{n\times p}$*, each* $j \in \{0,1,\ldots,p\}$*, and each* $\mathbf{d} \in \mathbb{R}^n$*, the mapping* $t \mapsto [\mathbf{x}_t]^{(j)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{d})$ *is the unique solution on* $[t_0,t_f]$ *of the ODE:*

$$\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t}(t) = \mathbf{h}_{(j)}(t,\mathbf{z}(t)), \qquad \mathbf{z}(t_0) = \mathbf{d}, \tag{24}$$

where the functions $\mathbf{h}_{(j)} : [t_0,t_f] \times \mathbb{R}^n \to \mathbb{R}^n$ *are defined inductively as follows:*

$$\begin{aligned}
&\mathbf{h}_{(0)} : (t,\mathbf{v}) \mapsto [\hat{\mathbf{f}}_t]'(\mathbf{x}(t,\mathbf{c}_0);\mathbf{v}),\\
&\mathbf{h}_{(j)} : (t,\mathbf{v}) \mapsto [\mathbf{h}_{(j-1),t}]'\left([\mathbf{x}_t]^{(j-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(j)});\mathbf{v}\right), \qquad \forall j \in \{1,\ldots,p\},
\end{aligned}$$

where $\hat{\mathbf{f}}_t : X \to \mathbb{R}^n$ *is defined for each* $t \in [t_0,t_f]$ *as in the statement of Theorem 4.1, and where* $\mathbf{h}_{(j),t} \equiv \mathbf{h}_{(j)}(t,\cdot)$*. Lastly, for each* $t \in [t_0,t_f]$ *and each* $j \in \{0,1,\ldots,p\}$*, let*

$$\mathbf{Y}(t,j,\mathbf{c}_0,\mathbf{M}) := \left[[\mathbf{x}_t]^{(0)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(1)}) \;\; [\mathbf{x}_t]^{(1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(2)}) \;\; \cdots \;\; [\mathbf{x}_t]^{(j-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(j)})\right] \in \mathbb{R}^{n\times j}.$$

(Thus, $\mathbf{Y}(t,0,\mathbf{c}_0,\mathbf{M}) = \varnothing_{n\times 0}$*.) The functions* $\mathbf{h}_{(j)}$ *satisfy:*

$$\mathbf{h}_{(j)}(t,\mathbf{v}) = [\hat{\mathbf{f}}_t]^{(j)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,j,\mathbf{c}_0,\mathbf{M})}(\mathbf{v}), \qquad \forall (t,\mathbf{v}) \in [t_0,t_f]\times\mathbb{R}^n, \quad \forall j \in \{0,1,\ldots,p\}.$$

Proof Corollary 4.1 shows that $\hat{\mathbf{f}}_t$ is lexicographically smooth at $\mathbf{x}(t,\mathbf{c}_0)$ for each $t \in [t_0,t_f]$. Now, choose any fixed $p \in \mathbb{N}$ and $\mathbf{M} \in \mathbb{R}^{n\times p}$. It will be shown by induction on $j \in \{0,1,\ldots,p\}$ that for every such j and every $\mathbf{d} \in \mathbb{R}^n$, the mapping $t \mapsto [\mathbf{x}_t]^{(j)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{d})$ is the unique solution on $[t_0,t_f]$ of the ODE (24), that $\mathbf{h}_{(j)}(t,\cdot)$ is lexicographically smooth for each $t \in [t_0,t_f]$, and that $\mathbf{h}_{(j)}$ satisfies the assumptions of Theorem 4.1 in place of $\mathbf{f}$, with $Z_{\mathbf{f}} = \emptyset$.

The case in which $j = 0$ follows immediately from Theorem 4.1 and Corollary 4.1. For the inductive step, suppose that for some $k \in \{1,\ldots,p\}$ and every $\mathbf{d} \in \mathbb{R}^n$, the mapping $t \mapsto [\mathbf{x}_t]^{(k-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{d})$ is the unique solution on $[t_0,t_f]$ of (24), and that $\mathbf{h}_{(k-1)}$ satisfies the assumptions of Theorem 4.1 in place of $\mathbf{f}$. The existence of k^{th}-order directional derivatives of $\mathbf{x}_t$ is not assumed *a priori*. Applying Theorem 4.1 with $\mathbf{h}_{(k-1)}$ in place of $\mathbf{f}$, with $\mathbf{m}^{(k)}$ in place of $\mathbf{c}_0$, with the mapping $(t,\mathbf{d}) \mapsto [\mathbf{x}_t]^{(k-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{d})$ in place of $\mathbf{x}$, and with $Z_{\mathbf{f}} = \emptyset$, for each $\mathbf{d} \in \mathbb{R}^n$, the mapping $t \mapsto [(\mathbf{x}_t)^{(k-1)}_{\mathbf{c}_0,\mathbf{M}}]'(\mathbf{m}^{(k)};\mathbf{d})$ is the unique solution on $[t_0,t_f]$ of the ODE:

$$\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t}(t) = [\mathbf{h}_{(k-1),t}]'\Big([\mathbf{x}_t]^{(k-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(k)});\mathbf{z}(t)\Big), \qquad \mathbf{z}(t_0) = \mathbf{d}.$$

Applying the definition of $\mathbf{h}_{(k)}$, it follows immediately that $t \mapsto [(\mathbf{x}_t)^{(k-1)}_{\mathbf{c}_0,\mathbf{M}}]'(\mathbf{m}^{(k)};\mathbf{d})$ is the unique solution on $[t_0,t_f]$ of (24) with $j := k$. Moreover, Theorem 4.1 shows that $[\mathbf{x}_t]^{(k-1)}_{\mathbf{c}_0,\mathbf{M}}$ is directionally differentiable at $\mathbf{m}^{(k)}$ for each $t \in [t_0,t_f]$, implying that $[\mathbf{x}_t]^{(k)}_{\mathbf{c}_0,\mathbf{M}} \equiv [(\mathbf{x}_t)^{(k-1)}_{\mathbf{c}_0,\mathbf{M}}]'(\mathbf{m}^{(k)};\cdot)$. Combining these remarks, for each $\mathbf{d} \in \mathbb{R}^n$, the mapping $t \mapsto [\mathbf{x}_t]^{(k)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{d})$ uniquely solves the ODE (24) with $j := k$. To complete the inductive step, Corollary 4.1 shows that $\mathbf{h}_{(k)}(t,\cdot)$ is lexicographically smooth for each $t \in [t_0,t_f]$, and that $\mathbf{h}_{(k)}$ satisfies the assumptions of Theorem 4.1 in place of $\mathbf{f}$, with $Z_{\mathbf{f}} = \emptyset$.

Since p and $\mathbf{M}$ were arbitrary in the above inductive argument, this argument shows that $\mathbf{x}_t$ is lexicographically smooth at $\mathbf{c}_0$ for each $t \in [t_0,t_f]$, as required.

Next, a simpler inductive proof shows that $\mathbf{h}_{(j)}(t,\cdot) \equiv [\hat{\mathbf{f}}_t]^{(j)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,j,\mathbf{c}_0,\mathbf{M})}$ for each $t \in [t_0,t_f]$ and each $j \in \{0,1,\ldots,p\}$, as follows. For the base case, the definition of $\mathbf{h}_{(0)}$ implies that for each $t \in [t_0,t_f]$,

$$\mathbf{h}_{(0)}(t,\mathbf{v}) = [\hat{\mathbf{f}}_t]'(\mathbf{x}(t,\mathbf{c}_0);\mathbf{v}) = [\hat{\mathbf{f}}_t]^{(0)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,0,\mathbf{c}_0,\mathbf{M})}(\mathbf{v}), \qquad \forall \mathbf{v} \in \mathbb{R}^n,$$

as required. For the inductive step, suppose that for some $k \in \{1,\ldots,p\}$,

$$\mathbf{h}_{(k-1),t} \equiv \mathbf{h}_{(k-1)}(t,\cdot) \equiv [\hat{\mathbf{f}}_t]^{(k-1)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,k-1,\mathbf{c}_0,\mathbf{M})}, \qquad \forall t \in [t_0,t_f].$$

The constructive definition of $\mathbf{h}_{(k)}$, the inductive assumption, and the definitions of $\mathbf{Y}(t,k-1,\mathbf{c}_0,\mathbf{M})$ and $\mathbf{Y}(t,k,\mathbf{c}_0,\mathbf{M})$ imply that, for each $t \in [t_0,t_f]$,

$$\mathbf{h}_{(k)}(t,\cdot) \equiv [(\hat{\mathbf{f}}_t)^{(k-1)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,k-1,\mathbf{c}_0,\mathbf{M})}]'\Big([\mathbf{x}_t]^{(k-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(k)});\cdot\Big) \equiv [\hat{\mathbf{f}}_t]^{(k)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,k,\mathbf{c}_0,\mathbf{M})}.$$

This completes the inductive step. □

Using the notation of Corollary 4.2, if $\mathbf{e}^{(1)},\ldots,\mathbf{e}^{(n)}$ denote the coordinate vectors in $\mathbb{R}^n$, then for any nonsingular $\mathbf{M} \in \mathbb{R}^{n\times n}$ and any $t \in [t_0,t_f]$,

$$\mathbf{J}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M}) = \Big[[\mathbf{x}_t]^{(n)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{e}^{(1)}) \;\; \cdots \;\; [\mathbf{x}_t]^{(n)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{e}^{(n)})\Big].$$

Thus, Corollary 4.2 provides a method for evaluating lexicographic derivatives of $\mathbf{x}(t,\cdot)$. Without further assumptions, though, this method is computationally expensive in the worst case, as it involves construction and evaluation of the ODE right-hand side function $(t,\mathbf{v}) \mapsto \mathbf{h}_{(j)}(t,\mathbf{v}) = [\hat{\mathbf{f}}_t]^{(j)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,j,\mathbf{c}_0,\mathbf{M})}(\mathbf{v})$ for each $j \in \{0,1,\ldots,n\}$. If the forward mode of automatic differentiation is used to construct these mappings using the identity

$$[\hat{\mathbf{f}}_t]^{(j)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,j,\mathbf{c}_0,\mathbf{M})} \equiv [(\hat{\mathbf{f}}_t)^{(j-1)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,j-1,\mathbf{c}_0,\mathbf{M})}]'(\mathbf{m}^{(j)};\cdot),$$

then the overall cost of this construction scales worst-case exponentially with j, relative to the cost of evaluating $\mathbf{f}$. To avoid this computational burden, the following theorem expresses $\mathbf{J}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M})$ in terms of the unique solution of an ODE, without requiring explicit construction of the intermediate directional derivatives $[\hat{\mathbf{f}}_t]^{(j)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,j,\mathbf{c}_0,\mathbf{M})}$.

Theorem 4.2 *Given an open, connected set $X \subset \mathbb{R}^n$ and real numbers $t_0 < t_f$, suppose that a function $\mathbf{f} : [t_0,t_f] \times X \to \mathbb{R}^n$ satisfies the following conditions:*

- *the mapping $\mathbf{f}(\cdot,\mathbf{c}) : [t_0,t_f] \to \mathbb{R}^n$ is measurable for each $\mathbf{c} \in X$,*
- *for each $t \in [t_0,t_f]$ except in a zero-measure subset $Z_{\mathbf{f}}$, the mapping $\mathbf{f}(t,\cdot) : X \to \mathbb{R}^n$ is lexicographically smooth,*
- *with $\mathbf{x}(\cdot,\mathbf{c})$ denoting any solution of the parametric ODE system:*

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}(t,\mathbf{c}) = \mathbf{f}(t,\mathbf{x}(t,\mathbf{c})), \qquad \mathbf{x}(t_0,\mathbf{c}) = \mathbf{c},$$

  there exists a solution $\{\mathbf{x}(t,\mathbf{c}_0) : t \in [t_0,t_f]\} \subset X$ for some $\mathbf{c}_0 \in X$,
- *there exists an open set $N \subset X$ such that $\{\mathbf{x}(t,\mathbf{c}_0) : t \in [t_0,t_f]\} \subset N$, and such that there exist Lebesgue integrable functions $k_{\mathbf{f}}, m_{\mathbf{f}} : [t_0,t_f] \to \mathbb{R}_+ \cup \{+\infty\}$ for which*

$$\|\mathbf{f}(t,\mathbf{c})\| \le m_{\mathbf{f}}(t), \qquad \forall t \in [t_0,t_f], \quad \forall \mathbf{c} \in N,$$

  and

$$\|\mathbf{f}(t,\mathbf{c}_1) - \mathbf{f}(t,\mathbf{c}_2)\| \le k_{\mathbf{f}}(t)\,\|\mathbf{c}_1 - \mathbf{c}_2\|, \qquad \forall t \in [t_0,t_f], \quad \forall \mathbf{c}_1,\mathbf{c}_2 \in N.$$

Then, for each $t \in [t_0,t_f]$, the function $\mathbf{x}_t \equiv \mathbf{x}(t,\cdot)$ is well-defined and Lipschitz continuous on a neighborhood of $\mathbf{c}_0$, with a Lipschitz constant that is independent of t.

Moreover, $\mathbf{x}_t$ *is lexicographically smooth at* $\mathbf{c}_0$*; for any* $p \in \mathbb{N}$ *and any* $\mathbf{M} \in \mathbb{R}^{n\times p}$*, the mapping* $t \mapsto \tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M})\,\mathbf{M}$ *is the unique solution on* $[t_0,t_f]$ *of the following ODE:*

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}(t) = \tilde{\mathbf{J}}_{\mathrm{L}}\hat{\mathbf{f}}_t(\mathbf{x}(t,\mathbf{c}_0);\mathbf{A}(t))\,\mathbf{A}(t), \qquad \mathbf{A}(t_0) = \mathbf{M}, \tag{25}$$

where $\hat{\mathbf{f}}_t : X \to \mathbb{R}^n$ *is defined in terms of* $\mathbf{f}$ *as follows, and is lexicographically smooth by construction for each* $t \in [t_0,t_f]$*:*

$$\hat{\mathbf{f}}_t(\mathbf{c}) = \begin{cases} \mathbf{f}(t,\mathbf{c}), & \text{if } t \in [t_0,t_f]\backslash Z_{\mathbf{f}}, \\ \mathbf{0}, & \text{if } t \in Z_{\mathbf{f}}. \end{cases}$$

Proof For each $t \in [t_0,t_f]$, the lexicographic smoothness of $\mathbf{x}_t$ at $\mathbf{c}_0$ was established in Corollary 4.2, implying that $\tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M})\,\mathbf{M}$ is uniquely defined for each $\mathbf{M} \in \mathbb{R}^{n\times p}$. Moreover, it was established in the proof of Theorem 4.1 that $\mathbf{x}_t$ is Lipschitz continuous on a neighborhood of $\mathbf{c}_0$ for each $t \in [t_0,t_f]$, with a Lipschitz constant that is independent of t.

Now, consider any fixed $p \in \mathbb{N}$ and $\mathbf{M} := \left[\mathbf{m}^{(1)} \cdots \mathbf{m}^{(p)}\right] \in \mathbb{R}^{n\times p}$. As an intermediate result, it will be shown by induction that for each $j \in \{1,\ldots,p\}$, the coupled ODE system:

$$\frac{\mathrm{d}\mathbf{z}_{(i)}}{\mathrm{d}t}(t) = [\hat{\mathbf{f}}_t]^{(i-1)}_{\mathbf{x}(t,\mathbf{c}_0),[\mathbf{z}_{(1)}(t)\ \mathbf{z}_{(2)}(t)\ \cdots\ \mathbf{z}_{(i-1)}(t)]}(\mathbf{z}_{(i)}(t)), \quad \mathbf{z}_{(i)}(t_0) = \mathbf{m}^{(i)}, \quad \forall i \in \{1,\ldots,j\} \tag{26}$$

has a unique solution on $[t_0,t_f]$, in which $\mathbf{z}_{(i)}(t) = [\mathbf{x}_t]^{(i-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(i)})$ for each $t \in [t_0,t_f]$ and each $i \in \{1,\ldots,j\}$. (Note that the right-hand sides of the coupled ODEs above are all well-defined, since Corollary 4.1 established the lexicographic smoothness of $\hat{\mathbf{f}}_t$ at $\mathbf{x}(t,\mathbf{c}_0)$ for each $t \in [t_0,t_f]$).

The case in which $j = 1$ follows immediately from Corollary 4.2. For the inductive step, suppose that for some $k \in \{2,3,\ldots,p\}$, the coupled ODE system (26) has a unique solution on $[t_0,t_f]$ when $j := k-1$, in which $\mathbf{z}_{(i)}(t) = [\mathbf{x}_t]^{(i-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(i)})$ for each $t \in [t_0,t_f]$ and each $i \in \{1,\ldots,k-1\}$. Now, consider the case in which $j := k$. In this case, the ODEs in (26) with $i \in \{1,\ldots,k-1\}$ are unchanged from the case in which $j = k-1$. Thus, by the inductive assumption, the ODEs in (26) with $i \in \{1,\ldots,k-1\}$ have unique solutions on $[t_0,t_f]$ in which $\mathbf{z}_{(i)}(t) = [\mathbf{x}_t]^{(i-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(i)})$ for each $t \in [t_0,t_f]$.

As a result, the ODE in (26) with $i = k$ becomes:

$$\frac{\mathrm{d}\mathbf{z}_{(k)}}{\mathrm{d}t}(t) = [\hat{\mathbf{f}}_t]^{(k-1)}_{\mathbf{x}(t,\mathbf{c}_0),\mathbf{Y}(t,k-1,\mathbf{c}_0,\mathbf{M})}(\mathbf{z}_{(k)}(t)), \qquad \mathbf{z}_{(k)}(t_0) = \mathbf{m}^{(k)}, \tag{27}$$

with $\mathbf{Y}(t,k-1,\mathbf{c}_0,\mathbf{M})$ defined as in the statement of Corollary 4.2. This corollary shows that (27) is uniquely solved on $[t_0,t_f]$ by the mapping $\mathbf{z}_{(k)} : t \mapsto [\mathbf{x}_t]^{(k-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(k)})$. Combining this statement with the inductive assumption completes the inductive step.

Using this inductive result, the coupled ODE system:

$$\frac{\mathrm{d}\mathbf{z}_{(i)}}{\mathrm{d}t}(t) = [\hat{\mathbf{f}}_t]^{(i-1)}_{\mathbf{x}(t,\mathbf{c}_0),[\mathbf{z}_{(1)}(t)\ \mathbf{z}_{(2)}(t)\ \cdots\ \mathbf{z}_{(i-1)}(t)]}(\mathbf{z}_{(i)}(t)), \quad \mathbf{z}_{(i)}(t_0) = \mathbf{m}^{(i)}, \quad \forall i \in \{1,\dots,p\} \tag{28}$$

has a unique solution on $[t_0,t_f]$, in which $\mathbf{z}_{(i)}(t) = [\mathbf{x}_t]^{(i-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(i)})$ for each $i \in \{1,\dots,p\}$. Using Properties 4 and 5 in Lemma 2.1, it follows that for each $i \in \{1,\dots,p\}$, each $t \in [t_0,t_f]$, and each choice of $\mathbf{v}_{(1)},\dots,\mathbf{v}_{(p)} \in \mathbb{R}^n$,

$$[\hat{\mathbf{f}}_t]^{(i-1)}_{\mathbf{x}(t,\mathbf{c}_0),[\mathbf{v}_{(1)}\ \cdots\ \mathbf{v}_{(i-1)}]}(\mathbf{v}_{(i)}) = \tilde{\mathbf{J}}_{\mathrm{L}}\hat{\mathbf{f}}_t(\mathbf{x}(t,\mathbf{c}_0);\begin{bmatrix}\mathbf{v}_{(1)} & \cdots & \mathbf{v}_{(p)}\end{bmatrix})\,\mathbf{v}_{(i)}.$$

Since $\mathbf{v}_{(i)}$ is a column of $\begin{bmatrix}\mathbf{v}_{(1)} & \cdots & \mathbf{v}_{(p)}\end{bmatrix}$, there is no ambiguity in the final term in the above equations. Thus, the following coupled ODE system is equivalent to (28):

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{z}_{(i)}}{\mathrm{d}t}(t) = \tilde{\mathbf{J}}_{\mathrm{L}}\hat{\mathbf{f}}_t(\mathbf{x}(t,\mathbf{c}_0);\begin{bmatrix}\mathbf{z}_{(1)}(t) & \cdots & \mathbf{z}_{(p)}(t)\end{bmatrix})\,\mathbf{z}_{(i)}(t), \\ \mathbf{z}_{(i)}(t_0) = \mathbf{m}^{(i)}, \end{cases} \qquad \forall i \in \{1,\dots,p\}, \tag{29}$$

and therefore has the same unique solution on $[t_0,t_f]$ as (28). Moreover, Property 4 in Lemma 2.1 and the definition of $\tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M})$ imply that

$$[\mathbf{x}_t]^{(i-1)}_{\mathbf{c}_0,\mathbf{M}}(\mathbf{m}^{(i)}) = \tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M})\,\mathbf{m}^{(i)}, \qquad \forall t \in [t_0,t_f], \qquad \forall i \in \{1,\dots,p\}.$$

Since $\mathbf{m}^{(i)}$ is a column of $\mathbf{M}$, the quantity $\tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M})\,\mathbf{m}^{(i)}$ in the above expression is uniquely defined for each choice of t and i. Thus, the unique solution of (29) on $[t_0,t_f]$ satisfies $\mathbf{z}_{(i)}(t) = \tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M})\,\mathbf{m}^{(i)}$ for each $i \in \{1,\dots,p\}$ and each $t \in [t_0,t_f]$. The coupled ODEs (29) may be written as the columns of a single ODE with the matrix-valued dependent variable $\mathbf{A} := \begin{bmatrix}\mathbf{z}_{(1)} & \cdots & \mathbf{z}_{(p)}\end{bmatrix}$ to yield the ODE (25), which therefore has the unique solution:

$$t \mapsto \begin{bmatrix}\tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M})\,\mathbf{m}^{(1)} & \cdots & \tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M})\,\mathbf{m}^{(p)}\end{bmatrix} = \tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{M})\,\mathbf{M}$$

on $[t_0,t_f]$. □

Corollary 3.2 and Theorem 4.2 together show that plenary Jacobian elements can be obtained for solutions of a nonsmooth parametric ODE, provided that lexicographic derivatives can be evaluated for the ODE right-hand side function, and provided that the unique solution of the ODE (25) can be determined or approximated numerically. This implies the following corollaries, which make use of *a priori* knowledge concerning the differentiability or convexity of the solution to a nonsmooth parametric ODE. These results do not require differentiability or convexity assumptions on the ODE right-hand side function.

Corollary 4.3 *Suppose that the hypotheses of Theorem 4.2 hold, and let* $\mathbf{x}_t \equiv \mathbf{x}(t,\cdot)$. *If* $\mathbf{x}_{t_f}$ *is known to be differentiable at* $\mathbf{c}_0$*, then the ODE:*

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}(t) = \tilde{\mathbf{J}}_{\mathrm{L}}\hat{\mathbf{f}}_t(\mathbf{x}(t,\mathbf{c}_0);\mathbf{A}(t))\,\mathbf{A}(t), \qquad \mathbf{A}(t_0) = \mathbf{I} \tag{30}$$

has a unique solution $\mathbf{A}$ *on* $[t_0,t_f]$*, which satisfies* $\mathbf{A}(t_f) = \mathbf{J}\mathbf{x}_{t_f}(\mathbf{c}_0)$.

Proof By Theorem 4.2, the mapping $\mathbf{A} : t \mapsto \mathbf{J}_{\mathrm{L}}\mathbf{x}_t(\mathbf{c}_0;\mathbf{I})$ is the unique solution on $[t_0,t_f]$ of (30). Since $\mathbf{x}_{t_f}$ is differentiable at $\mathbf{c}_0$, it follows from [19] that

$$\mathbf{J}_{\mathrm{L}}\mathbf{x}_{t_f}(\mathbf{c}_0;\mathbf{I}) \in \partial_{\mathrm{L}}\mathbf{x}_{t_f}(\mathbf{c}_0) = \{\mathbf{J}\mathbf{x}_{t_f}(\mathbf{c}_0)\}.$$

Thus, $\mathbf{A}(t_f) = \mathbf{J}_{\mathrm{L}}\mathbf{x}_{t_f}(\mathbf{c}_0;\mathbf{I}) = \mathbf{J}\mathbf{x}_{t_f}(\mathbf{c}_0)$. □

Now, for any function $\mathbf{g} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ that is piecewise differentiable in the sense of Scholtes [20], $\partial_{\mathrm{L}}\mathbf{g}(\mathbf{x}) \subset \partial\mathbf{g}(\mathbf{x})$ for each $\mathbf{x} \in X$ [21]. It follows that if the ODE right-hand side function $(t,\mathbf{c}) \mapsto \mathbf{f}(t,\mathbf{c})$ is piecewise differentiable with respect to $\mathbf{c}$ for almost all $t \in [t_0,t_f]$, then the solution to (25) is also an element of the linear Newton approximation to $\mathbf{x}(t,\cdot)$ at $\mathbf{c}_0$ described in [13, Corollary 12], right-multiplied by $\mathbf{M}$.

While the ODE (25) has a unique solution, the following example shows that its right-hand side function, $(t,\mathbf{A}) \mapsto \tilde{\mathbf{J}}_{\mathrm{L}}\hat{\mathbf{f}}_t(\mathbf{x}(t,\mathbf{c}_0);\mathbf{A})\,\mathbf{A}$, is not necessarily continuous with respect to $\mathbf{A}$ at almost every fixed $t \in [t_0,t_f]$. Thus, (25) is not necessarily a Carathéodory ODE. As the proof of Theorem 4.2 suggests, however, the columns of (25) can be decoupled to yield a sequence of Carathéodory ODEs, each with a unique solution.

Example 4.1 Consider the following parametric ODE system with two differential variables:

$$\frac{\mathrm{d}x_1}{\mathrm{d}t}(t,\mathbf{p}) = \frac{\mathrm{d}x_2}{\mathrm{d}t}(t,\mathbf{p}) = \max(x_1(t,\mathbf{p}), x_2(t,\mathbf{p})), \qquad \mathbf{x}(0,\mathbf{p}) = \mathbf{p}.$$

This ODE system satisfies the Carathéodory existence and uniqueness conditions when $\mathbf{x}(t,\mathbf{p})$ is restricted to any bounded neighborhood of $\mathbf{p}$; when $\mathbf{p} = (0,0)$, the unique solution is $\mathbf{x}(t,\mathbf{0}) := (x_1(t,\mathbf{0}), x_2(t,\mathbf{0})) = \mathbf{0}$ for each $t \in \mathbb{R}$. Now, with

$$\mathbf{A} := \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbb{R}^{2\times 2}, \qquad \text{and} \quad \mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2 : \mathbf{c} \mapsto (\max(c_1,c_2), \max(c_1,c_2)),$$

it follows that $\mathbf{f}$ is the composition of continuously differentiable functions and the function $\mathbf{c} \mapsto \max(c_1,c_2)$, and is therefore lexicographically smooth. Since $\mathbf{f}$ is not an explicit function of t, it follows that $\mathbf{f}$ itself plays the role of $\hat{\mathbf{f}}_t$ in Theorems 4.1 and 4.2. By inspection, for any $\mathbf{d} \in \mathbb{R}^2$ and any $t \in \mathbb{R}$,

$$\mathbf{f}^{(0)}_{\mathbf{x}(t,\mathbf{0}),\mathbf{A}}(\mathbf{d}) = \begin{cases} (d_1,d_1), & \text{if } d_1 \geq d_2, \\ (d_2,d_2), & \text{if } d_1 < d_2; \end{cases}$$

$$\begin{aligned} \mathbf{f}^{(1)}_{\mathbf{x}(t,\mathbf{0}),\mathbf{A}}(\mathbf{d}) &= [\mathbf{f}^{(0)}_{\mathbf{x}(t,\mathbf{0}),\mathbf{A}}]'((a_{11},a_{21});\mathbf{d}), \\ &= \begin{cases} (d_1,d_1), & \text{if } a_{11} > a_{21}, \text{ or if } a_{11} = a_{21} \text{ and } d_1 \geq d_2, \\ (d_2,d_2), & \text{if } a_{11} < a_{21}, \text{ or if } a_{11} = a_{21} \text{ and } d_1 < d_2. \end{cases} \end{aligned}$$

Using Lemma 2.1, it follows that:

$$\begin{aligned} \tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{f}(\mathbf{x}(t,\mathbf{0});\mathbf{A})\,\mathbf{A} &= \left[\mathbf{f}^{(2)}_{\mathbf{x}(t,\mathbf{0}),\mathbf{A}}(a_{11},a_{21}) \ \mathbf{f}^{(2)}_{\mathbf{x}(t,\mathbf{0}),\mathbf{A}}(a_{12},a_{22})\right], \\ &= \left[\mathbf{f}^{(0)}_{\mathbf{x}(t,\mathbf{0}),\mathbf{A}}(a_{11},a_{21}) \ \mathbf{f}^{(1)}_{\mathbf{x}(t,\mathbf{0}),\mathbf{A}}(a_{12},a_{22})\right], \\ &= \begin{cases} \begin{bmatrix} a_{11} & a_{12} \\ a_{11} & a_{12} \end{bmatrix}, & \text{if } a_{11} > a_{21}, \text{ or if } a_{11} = a_{21} \text{ and } a_{12} \geq a_{22}, \\ \begin{bmatrix} a_{21} & a_{22} \\ a_{21} & a_{22} \end{bmatrix}, & \text{if } a_{11} < a_{21}, \text{ or if } a_{11} = a_{21} \text{ and } a_{12} < a_{22}. \end{cases} \end{aligned}$$

It follows that for any $t \in \mathbb{R}$, the mapping $\mathbf{A} \mapsto \tilde{\mathbf{J}}_{\mathrm{L}}\mathbf{f}(\mathbf{x}(t,\mathbf{0});\mathbf{A})\,\mathbf{A}$ is discontinuous at any $\mathbf{A} \in \mathbb{R}^{2\times 2}$ for which both $a_{11} = a_{21}$ and $a_{12} \neq a_{22}$.

The following example presents a straightforward application of Theorem 4.2, in which the relevant ODE systems can all be solved analytically.

Example 4.2 Consider the function:

$$\mathbf{f}:\mathbb{R}^2\to\mathbb{R}^2:\mathbf{y}\mapsto\begin{bmatrix}(1-y_2)|y_1|\\ 1\end{bmatrix},$$

and the following nonsmooth parametric ODE system with two differential variables, in which $\mathbf{c}:=(c_1,c_2)\in\mathbb{R}^2$ denotes a parameter:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}(t,\mathbf{c})=\mathbf{f}(\mathbf{x}(t,\mathbf{c})),\qquad \mathbf{x}(0,\mathbf{c})=\mathbf{c}.$$

It is readily verified that this ODE system is uniquely solved by the mapping:

$$\mathbf{x}:\mathbb{R}\times\mathbb{R}^2\to\mathbb{R}^2:(t,\mathbf{c})\mapsto\begin{cases}\begin{bmatrix}c_1\exp\left(-\frac{1}{2}t^2+(1-c_2)t\right)\\ c_2+t\end{bmatrix}, & \text{if } c_1\geq 0,\\ \begin{bmatrix}c_1\exp\left(\frac{1}{2}t^2+(c_2-1)t\right)\\ c_2+t\end{bmatrix}, & \text{if } c_1<0.\end{cases}\tag{31}$$

Thus, $\mathbf{x}(t,\mathbf{0})=(0,t)$ for each $t\in\mathbb{R}$.

B-subdifferentials of the parametric ODE solution can be evaluated analytically in this case, as follows. For each fixed $t\in\mathbb{R}$, the mapping $\mathbf{x}_t\equiv\mathbf{x}(t,\cdot)$ is evidently differentiable at all domain points $\mathbf{c}$ for which $c_1\neq 0$. The definition of the B-subdifferential can thus be used to show that, when $\mathbf{c}=\mathbf{0}$,

$$\partial_{\mathrm{B}}\mathbf{x}_t(\mathbf{0})=\left\{\begin{bmatrix}\exp\left(-\frac{1}{2}t^2+t\right) & 0\\ 0 & 1\end{bmatrix},\begin{bmatrix}\exp\left(\frac{1}{2}t^2-t\right) & 0\\ 0 & 1\end{bmatrix}\right\}.$$

Thus, (4) can be used to show that

$$\partial_{\mathrm{P}}\mathbf{x}_t(\mathbf{0})=\partial\mathbf{x}_t(\mathbf{0})=\mathrm{conv}\left\{\begin{bmatrix}\exp\left(-\frac{1}{2}t^2+t\right) & 0\\ 0 & 1\end{bmatrix},\begin{bmatrix}\exp\left(\frac{1}{2}t^2-t\right) & 0\\ 0 & 1\end{bmatrix}\right\}.$$

Elements of the linear Newton approximation of $\mathbf{x}_t$ described in [13, Corollary 12] can be evaluated as follows. The function $\mathbf{f}$ is evidently differentiable at all domain points $\mathbf{y}$ for which $y_1 \neq 0$. Thus, for each $t \in \mathbb{R}$, Clarke's generalized Jacobian of $\mathbf{f}$ is evaluated at $\mathbf{x}(t,\mathbf{0}) = (0,t)$ to be:

$$\partial \mathbf{f}(\mathbf{x}(t,\mathbf{0})) = \left\{ \lambda\,(1-t) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} : \lambda \in [-1,1] \right\}, \qquad \forall t \in \mathbb{R}.$$

Now, define the mapping:

$$h : \mathbb{R} \times [-1,1] \to [-1,1] : (t,\mu) \mapsto \begin{cases} \mu, & \text{if } t \leq 1, \\ -\mu, & \text{if } t > 1. \end{cases}$$

The above results show that the linear Newton approximation of $\mathbf{x}_t$ at $\mathbf{0}$ described in [13, Corollary 12] includes the solutions of the following ODE for all $\mu \in [-1,1]$:

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}(t,\mu) = h(t,\mu)\,(1-t) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{A}(t,\mu), \qquad \mathbf{A}(0,\mu) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This ODE is readily solved to yield:

$$\mathbf{A}(t,\mu) = \begin{cases} \begin{bmatrix} \exp(\mu(-\frac{1}{2}t^2 + t)) & 0 \\ 0 & 1 \end{bmatrix}, & \text{if } t \leq 1, \\ \begin{bmatrix} \exp(\mu(\frac{1}{2}t^2 - t + 1)) & 0 \\ 0 & 1 \end{bmatrix}, & \text{if } t > 1. \end{cases}$$

Thus, for each $t > 1$, the linear Newton approximation $\Gamma \mathbf{x}_t(\mathbf{0})$ of $\mathbf{x}_t$ at $\mathbf{0}$ described in [13, Corollary 12] is such that

$$\operatorname{conv}\left\{ \begin{bmatrix} \exp(\frac{1}{2}t^2 - t + 1) & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} \exp(-\frac{1}{2}t^2 + t - 1) & 0 \\ 0 & 1 \end{bmatrix} \right\} \subset \Gamma \mathbf{x}_t(\mathbf{0}).$$

Lexicographic derivatives of the parametric ODE solution can be evaluated using Theorem 4.2 as follows. Following a similar approach to Example 4.1, the following is obtained for each $\mathbf{A} \in \mathbb{R}^{2\times 2}$. Here, a_{ij} denotes the (i,j)–element of $\mathbf{A}$.

$$\tilde{\mathbf{J}}_{\mathrm{L}} \mathbf{f}(\mathbf{x}(t,\mathbf{0});\mathbf{A})\,\mathbf{A} = \begin{cases} \begin{bmatrix} (1-t)a_{11} & (1-t)a_{12} \\ 0 & 0 \end{bmatrix}, & \text{if } a_{11} > 0, \text{ or if } a_{11} = 0 \text{ and } a_{12} \geq 0, \\ \begin{bmatrix} (t-1)a_{11} & (t-1)a_{12} \\ 0 & 0 \end{bmatrix}, & \text{if } a_{11} < 0, \text{ or if } a_{11} = 0 \text{ and } a_{12} < 0. \end{cases}$$

Thus, for any nonsingular $\mathbf{M} \in \mathbb{R}^{2\times 2}$, Theorem 4.2 shows that the mapping $t \mapsto \mathbf{J}_{\mathrm{L}}\mathbf{x}_t(\mathbf{0};\mathbf{M})\,\mathbf{M}$ is the unique solution of the ODE:

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}(t) = \begin{cases} (1-t)\begin{bmatrix} a_{11}(t) & a_{12}(t) \\ 0 & 0 \end{bmatrix}, & \text{if } a_{11}(t) > 0, \text{ or if } a_{11}(t) = 0 \text{ and } a_{12}(t) \geq 0, \\ (t-1)\begin{bmatrix} a_{11}(t) & a_{12}(t) \\ 0 & 0 \end{bmatrix}, & \text{if } a_{11}(t) < 0, \text{ or if } a_{11}(t) = 0 \text{ and } a_{12}(t) < 0, \end{cases}$$

$$\mathbf{A}(0) = \mathbf{M}.$$

This ODE can be solved by inspection; post-multiplying the result by $\mathbf{M}^{-1}$ yields:

$$\mathbf{J}_{\mathrm{L}}\mathbf{x}_t(\mathbf{0};\mathbf{M}) = \begin{cases} \begin{bmatrix} \exp(-\frac{1}{2}t^2+t) & 0 \\ 0 & 1 \end{bmatrix}, & \text{if } m_{11} > 0, \text{ or if } m_{11} = 0 \text{ and } m_{12} \geq 0, \\ \begin{bmatrix} \exp(\frac{1}{2}t^2-t) & 0 \\ 0 & 1 \end{bmatrix}, & \text{if } m_{11} < 0, \text{ or if } m_{11} = 0 \text{ and } m_{12} < 0, \end{cases}$$

and so

$$\partial_{\mathrm{L}}\mathbf{x}_t(\mathbf{0}) = \left\{ \begin{bmatrix} \exp(-\frac{1}{2}t^2+t) & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} \exp(\frac{1}{2}t^2-t) & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

This result is readily confirmed by lexicographic differentiation of (31) with respect to $\mathbf{c}$ at $\mathbf{c} = \mathbf{0}$.

Collecting the above results, and noting that, for each $t > 1$,

$$\exp(-\tfrac{1}{2}t^2+t-1) < \min\{\exp(-\tfrac{1}{2}t^2+t), \exp(\tfrac{1}{2}t^2-t)\},$$

$$\text{and} \qquad \exp(\tfrac{1}{2}t^2-t+1) > \max\{\exp(-\tfrac{1}{2}t^2+t), \exp(\tfrac{1}{2}t^2-t)\},$$

it follows that, for this example,

$$\partial_{\mathrm{L}}\mathbf{x}_t(\mathbf{0}) = \partial_{\mathrm{B}}\mathbf{x}_t(\mathbf{0}) \subset \partial\mathbf{x}_t(\mathbf{0}) = \partial_{\mathrm{P}}\mathbf{x}_t(\mathbf{0}) \subset \varGamma\mathbf{x}_t(\mathbf{0}), \qquad \forall t > 1.$$

The rightmost inclusion above is strict. In particular, when $t = 2$, the evaluated generalized derivatives satisfy:

$$\partial_{\mathrm{L}}\mathbf{x}_2(\mathbf{0}) = \partial_{\mathrm{P}}\mathbf{x}_2(\mathbf{0}) = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\} \subset \left\{ \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} : \lambda \in [\tfrac{1}{\mathrm{e}}, \mathrm{e}] \right\} \subset \varGamma\mathbf{x}_2(\mathbf{0}).$$

Although $\mathbf{x}_2$ is *strictly differentiable* at $\mathbf{0}$ in the sense of [1], $\varGamma\mathbf{x}_2(\mathbf{0})$ evidently contains elements other than $\mathbf{J}\mathbf{x}_2(\mathbf{0})$.

The result of Theorem 4.2 is easily extended to cover ODEs whose initial conditions are nontrivial functions of parameters $\mathbf{p} \in \mathbb{R}^{n_p}$:

$$\frac{d\mathbf{x}}{dt}(t,\mathbf{p}) = \mathbf{f}(t,\mathbf{x}(t,\mathbf{p})), \qquad \mathbf{x}(t_0,\mathbf{p}) = \mathbf{f}_0(\mathbf{p}). \tag{32}$$

provided that $\mathbf{f}$ satisfies the hypotheses of Theorem 4.2 (with $\mathbf{f}_0(\mathbf{p}_0)$ in place of $\mathbf{c}_0$ for some $\mathbf{p}_0 \in \mathbb{R}^{n_p}$), and provided that $\mathbf{f}_0 : \mathbb{R}^{n_p} \to \mathbb{R}^n$ is lexicographically smooth at $\mathbf{p}_0$. Introducing the auxiliary parametrized ODE:

$$\frac{d\mathbf{z}}{dt}(t,\mathbf{c}) = \mathbf{f}(t,\mathbf{z}(t,\mathbf{c})), \qquad \mathbf{z}(t_0,\mathbf{c}) = \mathbf{c},$$

and defining $\mathbf{x}_t \equiv \mathbf{x}(t,\cdot)$ and $\mathbf{z}_t \equiv \mathbf{z}(t,\cdot)$, it follows that $\mathbf{x}_t \equiv \mathbf{z}_t \circ \mathbf{f}_0$ for each t. Now, for any nonsingular $\mathbf{M} \in \mathbb{R}^{n\times n}$, let $\mathbf{B} := \mathbf{J}_\mathrm{L}\mathbf{f}_0(\mathbf{p}_0;\mathbf{M})\,\mathbf{M}$. Applying the chain rule (2) and post-multiplying the result by $\mathbf{M}$ yields:

$$\mathbf{J}_\mathrm{L}\mathbf{x}_t(\mathbf{p}_0;\mathbf{M})\,\mathbf{M} = \tilde{\mathbf{J}}_\mathrm{L}\mathbf{z}_t(\mathbf{f}_0(\mathbf{p}_0);\mathbf{B})\,\mathbf{B}. \tag{33}$$

Thus, $\mathbf{J}_\mathrm{L}\mathbf{x}_t(\mathbf{p}_0;\mathbf{M})$ can be evaluated by the following procedure:

Step 1: Evaluate $\mathbf{B}$.
Step 2: Use Theorem 4.2 to evaluate $\tilde{\mathbf{J}}_\mathrm{L}\mathbf{z}_t(\mathbf{f}_0(\mathbf{p}_0);\mathbf{B})\,\mathbf{B}$.
Step 3: Evaluate $\mathbf{J}_\mathrm{L}\mathbf{x}_t(\mathbf{p}_0;\mathbf{M})$ by solving the linear equation system (33).

Theorem 4.1 may be extended to cover (32) in a similar fashion.

This result may be extended in turn to parametric ODEs whose right-hand side functions depend explicitly on parameters $\mathbf{p} \in \mathbb{R}^{n_p}$:

$$\frac{d\mathbf{x}}{dt}(t,\mathbf{p}) = \mathbf{f}(t,\mathbf{p},\mathbf{x}(t,\mathbf{p})), \qquad \mathbf{x}(t_0,\mathbf{p}) = \mathbf{f}_0(\mathbf{p}). \tag{34}$$

Considering $\mathbf{p}$ as a constant dependent variable instead, the following ODE is constructed in terms of the augmented dependent variable $\mathbf{z} \equiv (\mathbf{p},\mathbf{x})$, and is equivalent to (34):

$$\frac{d\mathbf{z}}{dt}(t,\mathbf{p}) = \mathbf{h}(t,\mathbf{z}(t,\mathbf{p})), \qquad \mathbf{z}(t_0,\mathbf{p}) = \mathbf{h}_0(\mathbf{p}),$$

where

$$\mathbf{h} : (t,(\mathbf{q},\mathbf{c})) \mapsto \begin{bmatrix} \mathbf{0} \\ \mathbf{f}(t,\mathbf{q},\mathbf{c}) \end{bmatrix}, \qquad \text{and} \qquad \mathbf{h}_0 : \mathbf{q} \mapsto \begin{bmatrix} \mathbf{q} \\ \mathbf{f}_0(\mathbf{q}) \end{bmatrix}.$$

Provided that $\mathbf{h}$ satisfies conditions analogous to the hypotheses of Theorem 4.2, the above ODE in $\mathbf{z}$ may be treated in the same manner as (32). In the special case in which x_t is scalar-valued and convex on some neighborhood of $\mathbf{p}_0$, the discussion in Section 6.2 of [19] implies that $\mathbf{J}_\mathrm{L}x_t(\mathbf{p}_0;\mathbf{M})$ is a subgradient of x_t at $\mathbf{p}_0$. Hence, Theorem 4.2 describes certain subgradients of any convex solution of a nonsmooth parametric ODE system as the unique solutions of corresponding ODEs.

5 Conclusions

Corollary 3.2 shows that lexicographic derivatives of any lexicographically smooth function are also plenary Jacobian elements, which have been argued to be as useful as elements of Clarke's generalized Jacobian in methods for equation-solving and optimization. Theorems 4.1 and 4.2 describe directional derivatives and lexicographic derivatives for the unique solution of a parametric ODE system as the unique solutions of other ODEs. If the original ODE solution is known to be a scalar-valued convex function of the ODE parameters, then a subgradient is described, without requiring smoothness or convexity of the ODE right-hand side function. Similarly, if a differentiable function is the unique solution of a parametric ODE with a nonsmooth right-hand side, then its derivatives can be expressed as the solutions of corresponding ODE systems. To our knowledge, this work provides the first description of generalized derivatives of solutions of nonsmooth parametric ODEs that exhibit these properties.

Whether lexicographic derivatives for an arbitrary lexicographically smooth function are necessarily elements of Clarke's generalized Jacobian remains an open theoretical problem. An extension of the adjoint sensitivities of ODE solutions described in [12] to general nonsmooth parametric ODEs would also be useful, yet the definition of the plenary Jacobian suggests that it may not be a suitable generalized derivative for adjoint analyses.

Future work will involve developing numerical methods that use Theorem 4.2 to evaluate plenary Jacobian elements for nonsmooth dynamic systems arising in practical applications. Though the ODE (25) in Theorem 4.2 is well-posed and has a unique solution, existing numerical methods for ODE solution typically require some regularity of any discontinuities in the right-hand side function. Nevertheless, we expect that under certain regularity assumptions on the ODE right-hand side function, the ODE (25) will become tractable to solve numerically, without sacrificing too much generality from Theorem 4.2. Extensions of the theory in this work to sensitivity analysis of index-1, semi-explicit differential-algebraic equations are also being investigated.

Appendix

The following example shows that the unique solution of a nonsmooth parametric ODE is not necessarily differentiable with respect to the ODE parameters.

Example A.1 Consider the following parametric ODE, with $c \in \mathbb{R}$ denoting a scalar parameter:

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t,c) = |x(t,c)|, \qquad x(0,c) = c.$$

By inspection, this ODE is uniquely solved by the mapping:

$$x : (t,c) \mapsto \begin{cases} c\,\mathrm{e}^{t}, & \text{if } c \geq 0, \\ c\,\mathrm{e}^{-t}, & \text{if } c < 0. \end{cases}$$

Hence, for any fixed $t \neq 0$, the mapping $x(t,\cdot)$ is continuous but not differentiable at 0.

The following example illustrates the properties of linear Newton approximations described in Section 1.

Example A.2 Consider the mappings $f : \mathbb{R} \to \mathbb{R} : x \mapsto x$, $g : \mathbb{R} \to \mathbb{R} : x \mapsto \max(x,0)$, and $h : \mathbb{R} \to \mathbb{R} : x \mapsto \min(x,0)$. Using [1, Theorem 2.5.1], the Clarke generalized Jacobians of g and h are evaluated as:

$$\partial g(x) = \begin{cases} \{0\}, & \text{if } x < 0, \\ [0,1], & \text{if } x = 0, \\ \{1\}, & \text{if } x > 0, \end{cases} \qquad \partial h(x) = \begin{cases} \{1\}, & \text{if } x < 0, \\ [0,1], & \text{if } x = 0, \\ \{0\}, & \text{if } x > 0. \end{cases}$$

Now, g and h are each piecewise linear, and are therefore semismooth [14]. Since $f \equiv g + h$ on $\mathbb{R}$, it follows from [14, Corollary 7.5.18] that the following set-valued mapping is a linear Newton approximation for f:

$$\Gamma f : x \mapsto \partial g(x) + \partial h(x) = \begin{cases} \{1\}, & \text{if } x < 0, \\ [0,2], & \text{if } x = 0, \\ \{1\}, & \text{if } x > 0. \end{cases}$$

By inspection, f is convex and continuously differentiable on its domain, and has a derivative of $\mathbf{J}f(x) = 1$ for each $x \in \mathbb{R}$. In addition, f does not have any local minima on $\mathbb{R}$. However, although $\mathbf{J}f(0) \neq 0$, $0 \in \Gamma f(0)$.

Acknowledgements This paper is based on work funded by the National Science Foundation under grant CBET-0933095.

References

1. Clarke, F.H.: Optimization and Nonsmooth Analysis. SIAM, Philadelphia, PA (1990)
2. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Math. Program. **58**, 353–367 (1993)
3. Kojima, M., Shindo, S.: Extension of Newton and quasi-Newton methods to systems of PC^1 equations. J. Oper. Res. Soc. Jpn **29**, 352–375 (1986)
4. Kiwiel, K.C.: Methods of Descent for Nondifferentiable Optimization. Lecture Notes in Mathematics. Springer-Verlag, 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, NY (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.: Evaluating an element of the Clarke generalized Jacobian of a composite piecewise differentiable function. ACM T. Math. Software **39**, 23:1–23:28 (2013)
8. Griewank, A.: On stable piecewise linearization and generalized algorithmic differentiation. Optim. Method. Softw. **28**, 1139–1178 (2013)
9. Hartman, P.: Ordinary Differential Equations, second edn. SIAM, Philadelphia, PA (2002)
10. Galán, S., Feehery, W.F., Barton, P.I.: Parametric sensitivity functions for hybrid discrete/continuous systems. Appl. Numer. Math. **31**, 17–47 (1999)
11. Scott, J.K., Barton, P.I.: Improved relaxations for the parametric solutions of ODEs using differential inequalities. J. Global Optim. **57**, 143–176 (2013)

12. Yunt, M., Khan, K.A., Barton, P.I.: Parametric sensitivity analysis of dynamic systems using nonsmooth analysis: I. Theory. Submitted (2013)
13. Pang, J.S., Stewart, D.E.: Solution dependence on initial conditions in differential variational inequalities. Math. Program. B **116**, 429–460 (2009)
14. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer-Verlag New York, Inc., New York, NY (2003)
15. Yunt, M.: Nonsmooth dynamic optimization of systems with varying structure. Ph.D. thesis, Massachusetts Institute of Technology (2011)
16. Sweetser, T.: A minimal set-valued strong derivative for vector-valued Lipschitz functions. J. Optimiz. Theory App. **23**, 549–562 (1977)
17. Xu, H.: Set-valued approximations and Newton's methods. Math. Program. **84**, 401–420 (1999)
18. Hiriart-Urruty, J.B.: Characterizations of the plenary hull of the generalized Jacobian matrix. Math. Program. Stud. **17**, 1–12 (1982)
19. Nesterov, Y.: Lexicographic differentiation of nonsmooth functions. Math. Program. B **104**, 669–700 (2005)
20. Scholtes, S.: Introduction to Piecewise Differentiable Equations. SpringerBriefs in Optimization. Springer, New York, NY (2012)
21. Khan, K.A., Barton, P.I.: On a vector forward AD mode for generalized Jacobian element evaluation. In preparation (2013)
22. Imbert, C.: Support functions of the Clarke generalized Jacobian and of its plenary hull. Nonlinear Anal.-Theor. **49**, 1111–1125 (2002)
23. Filippov, A.F.: Differential Equations with Discontinuous Righthand Sides. Kluwer Academic Publishers, Dordrecht (1988)
24. Willett, D., Wong, J.S.W.: On the discrete analogues of some generalizations of Gronwall's inequality. Monatsh. Math. **69**, 362–367 (1965)
25. Coddington, E.A., Levinson, N.: Theory of Ordinary Differential Equations. McGraw Hill Co., Inc., New York, NY (1955)