Path differentiability of ODE flows

Swann Marx, Edouard Pauwels

Path differentiability of ODE flows

Swann Marx[1] and Edouard Pauwels[2]

Abstract

We consider flows of ordinary differential equations (ODEs) driven by path differentiable vector fields. Path differentiable functions constitute a proper subclass of Lipschitz functions which admit conservative gradients, a notion of generalized derivative compatible with basic calculus rules. Our main result states that such flows inherit the path differentiability property of the driving vector field. We show indeed that forward propagation of derivatives given by the sensitivity differential inclusions provide a conservative Jacobian for the flow. This allows to propose a nonsmooth version of the adjoint method, which can be applied to integral costs under an ODE constraint. This result constitutes a theoretical ground to the application of small step first order methods to solve a broad class of nonsmooth optimization problems with parametrized ODE constraints. This is illustrated with the convergence of small step first order methods based on the proposed nonsmooth adjoint.

1 Introduction

1.1 General context

We consider the ordinary differential equation (ODE for short), for some $T > 0$

$$\begin{aligned} &\dot{X}(t) = F(X(t)), \qquad \forall t \in [0,T] \\ &X(0) = x, \end{aligned} \tag{1}$$

where $F\colon \mathbb{R}^p \to \mathbb{R}^p$ is a Lipschitz function and $x \in \mathbb{R}^p$. We denote by $\phi\colon \mathbb{R}^p \times [0,T] \to \mathbb{R}^p$ the corresponding flow which associates to $(x,t) \in \mathbb{R}^p \times [0,T]$ the value $X(t)$ where $X\colon \mathbb{R} \to \mathbb{R}^p$ is the solution to (1) with $X(0) = x$. The flow ϕ typically inherits the regularity of F. For example if F is C^1, then ϕ is also C^1 (see *e.g.* [25, Section 17.6]).

In our setting, the flow ϕ is Lipschitz (see *e.g.*, [25, Section 17.4]). A notion of generalized derivative adapted to Lipschitz function is due to Clarke. The Clarke Jacobian takes values in subsets of $\mathbb{R}^{p\times p}$. We denote by $J_F\colon \mathbb{R}^p \rightrightarrows \mathbb{R}^{p\times p}$ the Clarke Jacobian of F [18, Section 2.6]. For $x \in \mathbb{R}^p$, it is defined as follows

$$J_F^c(x) = \operatorname{conv}\left\{v \in \mathbb{R}^{p\times p},\, \{x_k\}_{k\in\mathbb{N}} \subset R,\, x_k \to x,\, \operatorname{Jac}_F(x_k) \to v, k \to \infty\right\},$$

[1]CNRS UMR 6004, LS2N, École Centrale de Nantes, F-44000 Nantes, France.
[2]IRIT, Université de Toulouse, CNRS, 118 route de Narbonne, F-31400 Toulouse, France.

where R is any full measure set where F is differentiable and Jac_F is the usual Jacobian of F. Our main question of interest is to obtain generalized derivatives of ϕ from the knowledge of Clarke Jacobian of F. This question requires to take a more detailed look at the regularity of F.

1.2 Path differentiability of the flow

The class of Lipschitz functions F is too large for our purpose. Indeed, for generic Lipschitz F the Clarke Jacobian, $\partial^c F$ carries no information about the function itself [39, 13, 14]. In particular, it is proved in [13] that generic 1-Lipschitz functions have the same constant subgradient.

Therefore, to obtain meaningful calculus rules, we need to restrict F to be in a well behaved subclass. We choose the class of path differentiable functions, which was identified by several authors to be well behaved from a nonsmooth analysis perspective [38, 12, 7]. Let us emphasize that, although this is a negligible subclass of Lischitz functions, it is ubiquitous in potential applications as all semi-algebraic functions (more generally definable functions) are path differentiable [7]. This encompasses virtually any function Lipschitz F which can be written using an elementary logical formula involving elementary real operations including powers, exponential, logarithms, quotients, including large classes of numerical programs [8].

Following [7], F is called path differentiable, if it satisfies a chain rule along absolutely continuous curves: for any absolutely continuous $\gamma\colon [0,1] \to \mathbb{R}^p$, we have for almost all $t \in [0,1]$,

$$\frac{d}{dt} F(\gamma(t)) = D\dot{\gamma}(t), \qquad \forall D \in J_F^c(\gamma(t)) \subset \mathbb{R}^{p\times p}.$$

In a first step, we will be interested in the following question regarding regularity of ϕ

Does path differentiability of F imply path differentiability of ϕ?

We provide a positive answer to this question. The result is stated in Corollary 1. The proof is based on a differential inclusion which generalizes the variational equation for smooth ODEs (see for example [25, Section 17.6]) to the Lipschitz vector field F. This variational inclusion is described in [18, Section 7.4]. For any $x \in \mathbb{R}^p$, the latter is defined by the differential inclusion

$$\begin{aligned} &\dot{V}(t) \in J_F^c(\phi(x,t))V(t), \text{ for almost all } t \in [0,T] \\ &V(0) = I \in \mathbb{R}^{p\times p}. \end{aligned} \tag{2}$$

where V is to be found among absolutely continuous functions from $[0,T]$ to $\mathbb{R}^{p\times p}$. Equation (2) can be seen as a formal differentiation of equation (1). As proved in [18, Theorem 7.4.1], the Clarke Jacobian of the flow J_ϕ^c is to be found among the solutions of (2). More precisely, denoting by ψ the function $x \mapsto \phi(x,T)$, we have

$$J_\psi^c(x) \subset U(x) := \{V(T),\, V \text{ solution of (2) }\}$$

for all $x \in \mathbb{R}^p$. However, as shown in [4, Example 3.8], this inclusion can be strict even for a relatively simple F in $\mathbb{R}^2$ (see Section 2.1). Following [7], path differentiability of the flow is characterized by existence of a conservative Jacobian for ψ. More precisely, we show that the set valued map U, despite not being necessarily equal to the Clarke Jacobian of ψ, still satisfies the chain rule: for any absolutely continuous $\gamma\colon [0,1] \to \mathbb{R}^p$, we have for almost all $t \in [0,1]$,

$$\frac{d}{dt}\psi(\gamma(t)) = D\dot{\gamma}(t), \qquad \forall D \in U(\gamma(t)) \subset \mathbb{R}^{p\times p}.$$

This is the result given in Theorem 1. Thanks to this chain rule property, U is a conservative Jacobian of ψ, characterized as the set of solutions to a sensitivity differential inclusion. The existence of such a conservative Jacobian implies that ψ inherits the path-differentiable regularity of F.

The mapping U being a conservative Jacobian of ψ has several consequences for the flow. For example, as given in [7, Corollary 5], we have for Lebesgue almost all $x \in \mathbb{R}^p$

$$U(x) = \{\operatorname{Jac}\psi(x)\},$$

which means that the differential inclusion (2) provides a unique matrix that is the (classical) Jacobian of ψ. This allows to draw a connection with more classical notions of generalized derivatives. For example, using [23, Theorem 6.5], the mapping U (restricted to the set where it is a singleton) can be interpreted as a weak derivative of the flow ψ in the sense of Sobolev spaces.

1.3 Optimizing integral costs under ODE constraints

First motivations to address path differentiability of the flow relates to optimization problems of the form

$$\begin{aligned} \min_{\theta\in\mathbb{R}^m} \quad L(\theta) := \quad & \int_{t=0}^{t=T} \ell(Z(t))dt + \ell_T(Z(T)), \\ \text{where} \quad & Z(0) = \bar{z} \\ & \dot{Z}(t) = H(Z(t),\theta), \quad \forall t \in [0,T] \end{aligned} \tag{3}$$

where $\ell : \mathbb{R}^p \to \mathbb{R}$, $\ell_T : \mathbb{R}^p \to \mathbb{R}$ and $H\colon \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^p$ are Lipschitz, path differentiable functions and $\bar{z} \in \mathbb{R}^p$ is fixed. Note moreover that the flow depends on a given parameter $\theta \in \mathbb{R}^m$. The decision variable in problem (3) is a parameter vector θ. Such an optimization problem appears in many applications such as machine learning [16], data assimilation [28] or geophysics [32].

We will consider first order methods of gradient type to tackle problem (3) algorithmically. These methods generate sequences by recursively following negative gradient directions. The function L is Lipschitz and possibly nonsmooth so we need to use a generalized notion of gradient. The integral part of the loss L consists in a composition of the flow and a Lipschitz integral cost. However, Clarke Jacobian of the flow may be strictly contained in solutions of the variational inclusion (2), see [4, Example 3.8]. Fortunately, conservative

gradients can be used in place of usual gradients in a nonsmooth optimization context, provided that the objective function is path differentiable [7, 9]. Therefore, the main questions we need to adress are the following:

Is the loss L path differentiable? How to obtain a conservative gradient for L?

We leverage our main result on path differentiability of the flow, and the compatibility of conservative Jacobian with calculus rules to show that L is indeed path differentiable. More precisely, we show that a formal differentiation of L (application of integral differentiation rules which hold in the smooth case) using solutions of the variational inclusion (2) provides a conservative gradient for L, this is described in Corollary 5.

Numerical computation of a solution of the variational inclusion (2), for example using Euler discretization [21], requires to solve a differential inclusion of size $p \times p$. In the context of smooth ODEs, it is known that the size of the system to be solved can be reduced to p by using the adjoint method (see e.g., [15]) at the cost of solving an ODE backward in time. We derive a nonsmooth counterpart of the adjoint system using the conservative Jacobian framework and show that solutions to the adjoint system are elements of the conservative gradient for the loss L given in Corollary 5. This is described in Corollary 6.

Application of known results in nonsmooth optimization [9] show that using the prosposed conservative gradient in place of a gradient in a small step first order method context induces a minimizing behavior and generates sequences attracted by sets defined by an optimality condition. In other word, the output given by the proposed adjoint methods may be used as a first order optimization oracle to implement gradient type methods for the problem (3). This result is formally described in Corollary 7.

1.4 Related work

Combination of adjoint differentiation and small step methods of gradient type is at the heart of numerical methods for training neural ordinary differential equations models [16, 22]. Our results provide a theoretical ground for these approaches for which dedicated numerical librairies exist and are broadly used, such as `torchdiffeq` in `python`. These constitute one of the motivations for our investigation.

The use of Clarke's generalized derivatives in a dynamical systems context has been at the heart of nonsmooth analysis developments [18], in variational analysis [17] and stability analysis [19, 1]. More recent contributions include existence and Lipschitz regularity of nonsmooth differential algebraic equations [37] and generalizations in Wasserstein space [10].

The variational inclusion dates back to the work of Clarke [18]. Providing meaning to this equation has been an active topic of research. Let us mention the work of [31] which prove semismoothness of the flow induced by semismooth gradient fields. In this case the variational inclusion becomes an equation and allows to obtain directional derivatives. This result was extended by [27] to handle possibly discontinuous time dependency and lexicographic derivatives [30]. Deducing lexicographic derivatives from variational equation was extended to differential algebraic equations in [36]. All these works are centered

around notions of directional derivatives and forward derivative propagation. We are not aware of further interpretations of the variational inclusion (2) beyond directional derivatives and forward propagation. In an optimization context, directional derivatives are not sufficient as one needs to find candidate descent directions. This constitutes another motivation for the proposed developments.

1.5 Organization

The paper is organized as follows. Section 2 provides notations, definitions and details about the example of Clarke Jacobian forward propagation failure in [4, Example 3.8]. Section 3 contains preliminary results with their proofs. Section 4 is devoted to the first main result, the flow of (1) inherits path differentiability of F. Section 5 shows that integral costs in optimization with ODE constraints are path differentiable as soon as the the loss function is path differentiable. It is also proved that the adjoint method can be applied in this context to estimate elements of the corresponding conservative gradient. Section 6 is devoted to an extension of these results, from initial conditions dependency to the more general parametric case described in (3). The latter includes also convergence guaranties for the small step gradient like method. Some concluding remarks are collected in Section 7 together with further research lines. Finally, Appendix A gathers technical results used throughout the paper.

2 Notation and definitions

Notation. Set $\mathbb{R}_+ = [0, \infty)$. Given $p \in \mathbb{N}$, we will denote $\|\cdot\|$ the norm and $\langle \cdot, \cdot \rangle$ the scalar product in $\mathbb{R}^p$. We will denote by $\|\cdot\|_{op}$ the operator norm for matrices, i.e. if $A \in \mathbb{R}^{p\times p}$, then $\|A\|_{op} := \sup_{\|v\|\leq 1} \|Av\|$. The Frobenius norm is defined and denoted by $\|A\|_F := \sqrt{\mathrm{Tr}(A^\top A)}$, where $A \in \mathbb{R}^{p\times p}$, Tr is the trace operator and $A^\top$ is the transpose of A. The supremum norm is denoted by $\|\cdot\|_\infty$.

We recall that, due to Rademacher theorem [23, Theorem 3.1], any locally Lipschitz function is almost everywhere differentiable. *Absolutely continuous* curves $\gamma : \mathbb{R} \to \mathbb{R}^p$ are functions admitting a Lebesgue integrable derivative (defined for almost all $t \in \mathbb{R}$), such that for any $t \geq 0$:

$$\gamma(t) - \gamma(0) = \int_0^t \dot{\gamma}(s) ds.$$

Given three metric spaces H, S and Y, a *Carathédory function* $f\colon (x,t) \in H \times S \mapsto f(x,t) \in Y$ is a function such that $x \mapsto f(x,t)$ is Borel measurable for each $t \in S$ and such that $t \mapsto f(x,t)$ is continuous for each $x \in H$. We say that a function $f : \mathbb{R}^p \to \mathbb{R}$ is *lower semi-continuous*, if for every every sequence $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^p$ such that $\lim_{k\to\infty} x_k = \bar{x}$, one has $f(\bar{x}) \leq \liminf_{k\to+\infty} f(x_k)$.

A set valued map $D : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ is a function from $\mathbb{R}^p$ to a subset of $\mathbb{R}^q$. We say that D has a *closed graph* if, for any convergent sequences $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^p$ and $(v_k)_{k\in\mathbb{N}} \subset \mathbb{R}^q$, with $v_k \in D(x_k)$, one has $\lim_{k\to\infty} v_k \in D(\lim_{k\to+\infty} x_k)$.

Path differentiablity and conservative Jacobians. The notion of path differentiable functions has been introduced in [7], this class of regularity allows to apply basic differential calculus rules such as the chain rule. As explained in [7], the notion of *conservativity*, defined just below, is crucial to define path differentiable functions.

Definition 1 (Conservative Jacobian) *Let $D : \mathbb{R}^p \rightrightarrows \mathbb{R}^{n\times p}$ be a locally bounded, graph closed, nonempty valued map and $f : \mathbb{R}^p \to \mathbb{R}^n$ be a locally Lipschitz continuous function. Then, D is said to be a conservative Jacobian of f if and only if, for any absolutely continuous curve $\gamma : [0, 1] \to \mathbb{R}^p$, the function $t \mapsto f(\gamma(t))$ satisfies, for almost all $t \in [0, 1]$*

$$\frac{d}{dt} f(\gamma(t)) = V\dot{\gamma}(t),\ \forall V \in D(\gamma(t)).$$

Equivalently, D is a conservative Jacobian of f if and only if, for any measurable selection $V(t) \in D(\gamma(t))$ for all $t \in [0, 1]$,:

$$f(\gamma(1)) - f(\gamma(0)) = \int_0^1 V(t)\dot{\gamma}(t)dt.$$

When $n = 1$, we say that D is a conservative gradient.

Conservatives gradients are defined in the same way for real valued functions (see [7]). Throughout the paper, we require conservative gradients and Jacobians to be convex. This is not too restrictive due to the following remark.

Remark 1 *It follows from the definition that if D is conservative, then its pointwise convex hull $x \rightrightarrows \mathrm{conv}\{D(x)\}$ is also conservative [7].*

Conservativity leads to the notion of path differentiability:

Definition 2 (Path differentiable function) *We say that $f : \mathbb{R}^p \to \mathbb{R}^n$ is path differentiable if there exists a set valued map D such that D is a conservative Jacobian for f.*

Remark 2 *If J_f is a conservative Jacobian for f, then we have $J_f^c(x) \subset \mathrm{conv}\{J_f(x)\}$ for all x [7], in particular J_f^c is conservative. Hence J_f^c being conservative is a characterization of path differentiability of f as stated in the intoduction.*

Dynamical systems. Consider $F\colon \mathbb{R}^p \to \mathbb{R}^p$ the Lipschitz function given in (1) that is assumed path differentiable. We denote by $J_F\colon \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ a bounded convex valued conservative Jacobian for the vector field F which appears in (2). Throughout the paper, we denote by $K > 0$ a bound on the operator norm of J_F, that is,

$$\sup_{x\in\mathbb{R}^p, J\in J_F(x)} \|J\|_{\mathrm{op}} \leq K. \tag{4}$$

We introduce the following map

$$\begin{aligned} U\colon \mathbb{R}^p \times [0,T] &\to \mathbb{R}^{p\times p} \\ (x,t) &\rightrightarrows V(t) \qquad V \text{ solution to (2)}, \end{aligned} \tag{5}$$

which is a candidate for being a conservative Jacobian for the flow ϕ of (1). The main result of this paper is to show that $(x,t) \rightrightarrows (U(x,t), F(\phi(x,t))$ is a conservative Jacobian for the flow ϕ, where we have used matrix concatenation.

Since F is assumed to be Lipschitz, the set of solutions to (2) is composed by Lipschitz functions, as stated in the following lemma.

Lemma 1 *For any $x \in \mathbb{R}^p$, $T > 0$, the set of solutions to (2) is non empty and only contains L-Lipschitz functions with $L = K\sqrt{p}\exp(KT)$.*

Proof : By hypotheses on J_F, the solution set of (2) is nonempty and defined on maximal intervals invoking [3, Theorem 4, p. 101].

It remains to show that solutions to (2) are bounded. Indeed, once one has a bound on V, one deduces a bound on $\dot{V}$ through the following inequality, which holds for a.e. $t \in [0,T]$ and is sufficient to ensure Lipschicity,

$$\|\dot{V}(t)\|_F \leq K\|V(t)\|_F. \tag{6}$$

Using (2), (4) and (6), for a.e. $t \in [0,T]$, one has:

$$\frac{d}{dt}\|V(t)\|_F^2 = 2\mathrm{Tr}(V(t)^\top \dot{V}(t)) \leq 2K\|V(t)\|_F^2.$$

Thanks to Lemma 6, and using the fact that $V(0) = I$ (see (2)), one deduces that, for all $t \in [0,T]$:

$$\|V(t)\|_F^2 \leq \|V(0)\|_F^2 \exp(2KT) = p\exp(2KT)$$

This latter equation together with (6) shows that the solutions V to (2) are L-Lipschitz with $L = K\sqrt{p}\exp(KT)$. This concludes the proof of the Lemma. □

Remark 3 (On the Lipschitz assumption) *The vector field F has been supposed to be Lipschitz with a uniform bound on a conservative Jacobian in (4), which is a stronger assumption than the (classical) local Lipschitz assumption. Under local Lipschicity, solutions only exist in a time interval which could be bounded with endpoint depending on initial condition. The global Lispchicity assumption allows to avoid such discussions, but there exist possible extensions which would allow to relax it:*

1. *If we suppose the trajectories (and the initial condition) to belong to some compact set, then local and global Lipschicity will be essentially equivalent for our purpose. For example if F maximal monotone [6, Chapter 7], one can show that the trajectories of (1) are bounded, independently of the initial condition.*

2. *A different possibility is to assume that all solutions to* (1) *are well defined on* $[0,T]$ *for any initial condition. One can see our global Lipschicity assumption as a sufficient condition.*

Remark 4 *It is worth mentioning that the unknown of* (2) *is a matrix, and not a vector as it is commonly defined in textbooks such as [3, 24]. It is always possible to identify a* $p \times p$ *matrix with a vector of dimension* $p \times p$*, and the meaning of matrix differential inclusion follows by using this identification.*

2.1 Failure of formal differentiation with Clarke Jacobian

Following [4, Example 3.8], consider an instance of (1) as follows

$$\begin{pmatrix} \dot{X}_1 \\ \dot{X}_2 \end{pmatrix} = \begin{pmatrix} (1 - X_2)|X_1| \\ 1 \end{pmatrix},$$

it can be proved that for any initilization $X_1(0)$ and $X_2(0) = 0$, we have $X_1(2) = X_1(0)$ and $X_2(2) = X_2(0)+2$, therefore, the flow is differentiable at $T = 2$ and its Jacobian is the identity matrix. Furthermore, if $X_1(0) = 0$, then we actually have $X_1(t) = 0$ for all $t \in \mathbb{R}$. However the variational inclusion (2) for the particular initialization $X_1(0) = X_2(0) = 0$ reads

$$\dot{M} \in \begin{pmatrix} [-|1-t|, |1-t|] & 0 \\ 0 & 0 \end{pmatrix} M$$

with $M(0) = I$ where we chose the conservative derivative of absolute value to be the usual derivative everywhere except at 0 where it is the segment $[-1, 1]$. All entries of M remain constant in time, except for the first one which we denote by m. The two extreme solutions for m are given by $\dot{m} = |1 - t|m$ and $\dot{m} = -|1 - t|m$ which leads to $m(2) \in [1/e, e]$. Therefore the variational inclusion fails to provide the correct subgradient for $X_1(2)$ with respect to the initial condition $X_1(0)$ (this should be 1). This example highlights the fact that it is not possible to prove that the sensitivity analysis differential inclusion (2) provides subgradients in general. In this example, the discrepancy occurs at the origin only, and, as described in the forthcoming results, the solutions to (2) actually provide a conservative Jacobian for the flow.

3 Preliminary results

Fix any $T > 0$, we define the following mapping:

$$\begin{aligned} \mathcal{U} \colon \mathbb{R}^p &\rightrightarrows \mathcal{C}([0,T], \mathbb{R}^{p\times p}) \\ x &\rightrightarrows \{ t \mapsto V(t),\, V \text{ solution to (2) } \}. \end{aligned}$$

We call the mapping $\mathcal{U}$ the solution mapping of the differential inclusion (2) whose values are Lipschitz functions from $[0,T]$ to $\mathbb{R}^{p\times p}$ (see Lemma 1). Note that, for all $t \in [0,T]$,

$U(x,t) = \{V(t),\ V \in \mathcal{U}(x)\}$, where U is defined in (5). We introduce a Castaing representation for functions with values in Lipschitz subsets of $\mathcal{C}([0,T])$ which will allow to specify some technical measurability issues for $\mathcal{U}$.

Proposition 1 (Castaing representation of solution mappings) *Given $T > 0$, $L > 0$, denote by $\mathcal{L}$ the space of L-Lipschitz functions from $[0,T]$ to $\mathbb{R}^q$, endowed with the supremum norm. Consider a set-valued map $\mathcal{V} : \mathbb{R}^m \rightrightarrows \mathcal{L}$ with closed graph and non empty values. Then $\mathcal{V}$ admits a* Castaing representation*, that is, a sequence of Borel measurable functions from $\mathbb{R}^m$ to $\mathcal{L}$, $(M_n)_{n\in\mathbb{N}}$, such that $\mathcal{V} = \overline{\{M_1(x), M_2(x), \ldots\}}$, for each $x \in \mathbb{R}^m$, where the closure and Borel measurability are induced by L^∞ norm over continuous functions. Furthermore, for all $i \in \mathbb{N}$, M_i can be seen as a function $\mathbb{R}^m \times [0,T] \to \mathbb{R}^q$ and we have that $(x,t) \mapsto M_i(x,t)$ is a Carathéodory function, L Lipschitz in t for fixed x and Borel measurable in x for fixed t.*

Proof : Recall that σ-compact sets are sets defined as the union of countably many compact subspaces. Since the domain of the $\mathcal{V}$ is obviously σ-compact and takes values in the space $\mathcal{L}$ which is, by Lemma 10, also σ-compact, then on can invoke [2, Theorem 18.20] to deduce that $\mathcal{V}$ is Borel measurable. Finally, using [2, Corollary 18.14], there exists a Castaing representation for $\mathcal{V}$. Moreover, by [2, Theorem 4.55], this representation is actually a sequence of Carathéodory functions, which is our desired result. □

We are now in position to state a technical representation result for the set of solutions of (2). Fix any $T > 0$ and any absolutely continuous function $\gamma : [0,1] \to \mathbb{R}^p$. We define the following mapping:

$$\begin{aligned} \mathcal{U}_\gamma \colon [0,1] &\rightrightarrows \mathcal{C}([0,T], \mathbb{R}^{p\times p}) \\ r &\rightrightarrows V \in \mathcal{U}(\gamma(r)), \end{aligned}$$

which corresponds to the set of solutions to (2) with the initial condition given by $x = \gamma(r)$ in (1).

Lemma 2 *The map $\mathcal{U}_\gamma$ is locally bounded, has a closed graph, is Borel measurable and admits a countable collection of dense Carathéodory selections $M \colon [0,1] \times [0,T] \mapsto \mathbb{R}^{p\times p}$ which are absolutely continuous in time and Borel measurable in r. For each such M, there is a Lebesgue measurable selection $S(r,t) \in J_F(\phi(\gamma(r),t)$ for all $(r,t) \in [0,1] \times \mathbb{R}$, such that, for all $r \in [0,1]$ and for almost all $t \in [0,T]$,*

$$\frac{\partial}{\partial t} M(r,t) = S(r,t) M(r,t).$$

Proof : Since J_F is bounded, ϕ is Lipschitz and γ is absolutely continuous, one can deduce that $\mathcal{U}_\gamma$ is locally bounded. Using the fact that γ is absolutely continuous (and therefore has a closed graph) and invoking [24, Corollary 1 and Theorem 3, §7, Chapter 2], one can deduce that $\mathcal{U}_\gamma$ has a closed graph.

Using Proposition 1, $\mathcal{U}_\gamma$ admits a Castaing representation as the closure of a countable dense set of Carathéodory selections. It remains to show that the functions composing this representation satisfy the claimed differential equation and to construct the proposed S.

Fix M an element of this Castaing representation. With a slight abuse of notation, for the rest of this proof, we will see M as a function of (r,t) by identifying M with $(r,t) \mapsto M(r,t)$. As a Carathéodory function, M is jointly Borel measurable in r and t as stated in [2, Lemma 4.51]. Denote by $\frac{\partial}{\partial t}M$ the partial derivative of M with respect to t when it exists. Since M is Lipschitz (hence absolutely continuous) with respect to t, for any $r \in [0,1]$, $\frac{\partial}{\partial t}M(r,t)$ is defined for almost all $t \in [0,T]$.

Consider the set $E \subset [0,1] \times [0,T]$ the set where $\frac{\partial}{\partial t}M(r,t)$ exists. By Lemma 3, E has full Lebesgue measure and $(r,t) \mapsto \frac{\partial}{\partial t}M(r,t)$ is Lebesgue measurable. Furthermore, for all $r \in [0,1]$, $\{t \in [0,T], (r,t) \in E\}$ has full measure by Lipschicity of M in the variable t for fixed r. The set E is a measure space with the induced subspace measure.

Consider the function

$$\begin{aligned} f \colon \mathbb{R}^{p\times p} \times E &\to \mathbb{R}_+ \\ (S,r,t) &\mapsto \left\| \frac{\partial}{\partial t}M(r,t) - SM(r,t) \right\|^2, \end{aligned}$$

which is jointly Lebesgue measurable in (r,t) for a fixed $S \in \mathbb{R}^{p\times p}$ since the sum of the Lebesgue measurable functions $\frac{\partial}{\partial t}M$ and $-SM(r,t)$ is Lebesgue measurable, and because the composition of this sum with the norm function (which is continuous) is also Lebesgue measurable. This function is also continuous in S for fixed $(r,t) \in [0,1] \times [0,T]$, implying then that f is a Carathédory function. Consider $K > 0$ the global upper bound on $\|J_F\|_{op}$ as in (4). By [2, Corollary 18.8], the set valued map

$$\begin{aligned} \mathcal{S}_1 \colon E &\rightrightarrows \mathbb{R}^{p\times p} \\ (r,t) &\rightrightarrows \left\{ S \in \mathbb{R}^{p\times p},\ \|S\| \leq K,\ f(S,r,t) = 0 \right\} \end{aligned}$$

is measurable since S belongs to a compact set. We extend $\mathcal{S}_1$ to $[0,1] \times [0,T]$ by setting $\mathcal{S}_1(r,t) = \emptyset$ if $(r,t) \notin E$. Measurability of $\mathcal{S}_1$ is preserved applying [2, Definition 18.1]. Now consider the set valued function

$$\begin{aligned} \mathcal{S}_2 \colon [0,1] \times [0,T] &\rightrightarrows \mathbb{R}^p \\ (r,t) &\rightrightarrows J_F(\phi(\gamma(r),t)). \end{aligned}$$

Since the graph of J_F is closed, and using moreover the continuity of the functions ϕ and γ, the function $S \mapsto \operatorname{dist}(S, J_F(\phi(\gamma(r),t)))$ is lower semicontinuous, hence Borel measurable by Lemma 4. It implies that it is a Carathéodory function, proving that $\mathcal{S}_2$ is Borel measurable [2, Theorem 18.5]. Now consider the intersection set valued map:

$$\begin{aligned} \mathcal{S} \colon [0,1] \times [0,T] &\rightrightarrows \mathbb{R}^p \\ (r,t) &\mapsto \mathcal{S}_1(r,t) \cap \mathcal{S}_2(r,t). \end{aligned}$$

It is measurable [2, Lemma 18.4, Item 3] and compact valued. Consider the set $\tilde{E} = \{(r,t) \in [0,1] \times [0,T],\ \mathcal{S}(x,t) \neq \emptyset\}$, which is measurable (see discussion after Definition

18.1 in [2]). We have that $\tilde{E} \subset E$ because $\mathcal{S}$ is empty valued outside of E and $\tilde{E} = \{(r,t) \in E, \frac{d}{dt}M(r,t) \in J_F(\phi(\gamma(r),t))M(r,t)\}$ by construction.

Since for any $r \in [0,1]$, $M(r) \in \mathcal{U}_\gamma(r)$, it holds for almost all $t \in [0,T]$ that $\frac{d}{dt}M(r,t) \in J_F(\phi(\gamma(r),t))M(r,t)$. In other words, for all $r \in [0,1]$, $\{t \in [0,T], (r,t) \in \tilde{E}\}$ has full measure. This is by definition of $\mathcal{U}_\gamma$. Therefore, by Fubini's Theorem [34, Theorem 16 Section 20.2], $\tilde{E}$ has full measure.

Set for all $(r,t) \in [0,1] \times [0,T]$, $\tilde{\mathcal{S}}(r,t) = \mathcal{S}(r,t)$ if $\mathcal{S}(r,t) \neq \emptyset$ (that is $(r,t) \in \tilde{E}$), and $J_F(\phi(\gamma(r)),t)$ otherwise, it satisfies $\tilde{\mathcal{S}}(r,t) \subset J_F(\phi(\gamma(r),t))$ for all $(r,t) \in [0,1] \times [0,T]$ and has nonempty values. The mapping $\tilde{\mathcal{S}}$ is measurable and has non empty closed values [2, Theorem 18.13]. Therefore it admits a measurable selection $S\colon [0,1] \times [0,T] \mapsto \mathbb{R}^{p \times p}$, which is the desired function. This achieves the proof. □

Remark 5 *Given M and S as in Lemma 2, we have by [24, Theorem 2, §1, Chapter1] that, for all $r \in [0,1]$, $t \mapsto M(r,t)$ is the unique absolutely continuous solution to*

$$\frac{\partial}{\partial t}M(r,t) = S(r,t)M(r,t).$$

Remark 6 *Note that, using the same arguments, the solution mapping $\mathcal{U}$ defined at the beginning of the section is also locally bounded and has a closed graph. Indeed, since J_F is bounded and ϕ is Lipschitz, it is clear that $\mathcal{U}$ is locally bounded. Then using [24, Corally 1 and Theorem 3, §7,Chapter 2], one deduces that $\mathcal{U}$ has a closed graph.*

4 Path differentiability of the flow

This section is devoted to the proof of our main result, conservativity of the mapping defined in (2) for the flow of (1). We first prove that, for any $T \geq 0$, the mapping U evaluated at $t = T$ is conservative for the flow evaluated at $t = T$.

Theorem 1 *For all $T \geq 0$, the mapping $x \rightrightarrows U(x,T)$ is conservative for $x \mapsto \phi(x,T)$.*

Proof : If $T = 0$, then the statement is obvious. Then, we restrict our analysis to the case where $T > 0$.

Consider an absolutely continuous path $\gamma\colon [0,1] \to \mathbb{R}^p$. Let $M\colon [0,1] \times [0,T] \to \mathbb{R}^{p \times p}$ be a Carathéodory function as in Proposition 1 such that, for all $(r,t) \in [0,1] \times [0,T]$, $M(r,t) \in U(\gamma(r),t)$. Consider the Lebesgue measurable selection $S(r,t) \in J_F(\phi(\gamma(r),t))$ for all $(r,t) \in [0,1] \times [0,T]$ as given by Lemma 2, such that, for all $r \in [0,1]$ and almost all $t \in [0,T]$

$$\frac{\partial}{\partial t}M(r,t) = S(r,t)M(r,t). \tag{7}$$

In addition, we have, for all $r \in [0,1]$ and all $t \in [0,T]$,

$$\phi(\gamma(r),t) - \gamma(r) = \int_{s=0}^{s=t} F(\phi(\gamma(r),s))ds, \tag{8}$$

since $\phi(\gamma(r),0) = \gamma(r)$.

Since ϕ is Lipschitz, for each $s \in [0,T]$, $r \mapsto \phi(\gamma(r),s)$ is an absolutely continuous loop. Therefore, it is differentiable at almost all $r \in [0,1]$. Applying Lemma 3 shows that the function

$$g\colon (r,s) \mapsto \frac{d}{dr}\phi(\gamma(r),s),$$

is well defined for all $s \in [0,t]$ and almost all $r \in [0,1]$, and Lebesgue measurable in (r,s). Therefore, for all $t \in [0,T]$, for almost all $r \in [0,1]$, it follows from (8) that

$$g(r,t) - \dot{\gamma}(r) = \frac{d}{dr}\int_{s=0}^{s=t} F(\phi(\gamma(r),s))ds.$$

The integrand is jointly integrable in (r,s), and absolutely continuous in r for each s. It follows by Lemma 5 that, for all $t \geq 0$ and for almost all $r \in [0,1]$

$$g(r,t) - \dot{\gamma}(r) = \int_{s=0}^{s=t} \frac{\partial}{\partial r}F(\phi(\gamma(r),s))ds.$$

Since F is path differentiable, we have, for all $s \in [0,t]$, for almost all $r \in [0,1]$

$$\begin{aligned}\frac{\partial}{\partial r}F(\phi(\gamma(r),s)) &= J \times g(r,s) \qquad \forall J \in J_F(\phi(\gamma(r),s)) \\ &= S(r,s)g(r,s),\end{aligned}$$

where S is the Lebesgue measurable selection defined in (7). Therefore, by integration, we have, for all $t \in [0,T]$, for almost all $r \in [0,1]$

$$g(r,t) - \dot{\gamma}(r) = \int_{s=0}^{s=t} S(r,s)g(r,s)ds, \tag{9}$$

Now, we rewrite (7) by integration, for all $(r,t) \in [0,1] \times [0,T]$, using $M(r,0) = I$

$$M(r,t) - I = \int_{s=0}^{s=t} S(r,s)M(r,s)ds.$$

Multiplying both sides of the latter equation by $\dot{\gamma}(r)$, that is defined for almost all $r \in [0,1]$, one has, for all $t \geq 0$, for almost all $r \in [0,1]$

$$M(r,t)\dot{\gamma}(r) - \dot{\gamma}(r) = \int_{s=0}^{s=t} S(r,s)M(r,t)\dot{\gamma}(r)ds. \tag{10}$$

Combining both (9) and (10), we have, for all $t \geq 0$, for almost all $r \in [0,1]$

$$\begin{aligned}\|M(r,t)\dot{\gamma}(r) - g(r,t)\| &= \left\| \int_{s=0}^{s=t} S(r,s)(M(r,s)\dot{\gamma}(r) - g(r,s))ds \right\| \\ &\leq \int_{s=0}^{s=t} \|S(r,s)(M(r,s)\dot{\gamma}(r) - g(r,s))\| \, ds \\ &\leq K \int_{s=0}^{s=t} \|(M(r,s)\dot{\gamma}(r) - g(r,s))\| \, ds\end{aligned}$$

where K is a bound on J_F given in (4). Integrating with respect to r and using Fubini's theorem, we have, for all $t \in [0,T]$

$$\int_{r=0}^{r=1} \|M(r,t)\dot{\gamma}(r) - g(r,t)\| dr \leq K \int_{s=0}^{s=t} \int_{r=0}^{r=1} \|(M(r,s)\dot{\gamma}(r) - g(r,s))\| \, drds$$

By Lemma 7, one obtains that, for all $t \geq 0$

$$\int_{r=0}^{r=1} \|M(r,t)\dot{\gamma}(r) - g(r,t)\| dr = 0.$$

Therefore, we have, for all $t \geq 0$ and all $r \in [0,1]$

$$\phi(\gamma(r),t) - \phi(\gamma(0),t) = \int_{u=0}^{u=r} g(u,t)du = \int_{u=0}^{u=r} M(u,t)\dot{\gamma}(u)du.$$

Since M was an arbitrary Carathéodory function in a countable dense subset of such selections, one can apply Lemma 8. This shows that $x \rightrightarrows U(x,t)$ is conservative for $x \mapsto \phi(x,t)$. This concludes the proof. □

From the latter result, one can deduce that the flow ϕ is path differentiable for all $t \geq 0$. It is stated in the following corollary.

Corollary 1 *The mapping $(x,t) \rightrightarrows (U(x,t), F(\phi(x,t)))$ is conservative for ϕ and in particular, ϕ is path differentiable.*

Proof : Consider the following dynamical system on $\mathbb{R}^{p+1}$

$$\begin{aligned}\dot{Y}(s) &= \alpha(s)F(Y(s)) \\ \dot{\alpha}(s) &= 0\end{aligned} \tag{11}$$

Consider $\tilde{F}\colon \mathbb{R}^{p+1} \to \mathbb{R}^{p+1}$ the vector field associated to the ODE in (11) with the state (Y,α). It is given by

$$\tilde{F}(Y(s),\alpha) = \begin{pmatrix} \alpha F(Y(s)) \\ 0 \end{pmatrix}$$

We can compute a conservative Jacobian for $\tilde{F}$ using the product rule of differential calculus and component-wise aggregation, both valid for conservative Jacobians [7, Lemmas 3 and 5]. We obtain a conservative Jacobian for $\tilde{F}$ as follows:

$$(x,\alpha) \rightrightarrows \begin{pmatrix} \alpha J_F(x) & F(x) \\ 0 & 0 \end{pmatrix} \tag{12}$$

Denote by $\tilde{\phi}\colon \mathbb{R}^{p+1} \to \mathbb{R}^{p+1}$ the flow associated to (11), and recall that ϕ is the flow of the system (1). We have, by a simple rescaling of time, for any $x \in \mathbb{R}^p$, $\alpha \in \mathbb{R}$ and any $s \in [0,1]$

$$\phi(x, \alpha s) = \tilde{\phi}(x, \alpha, s). \tag{13}$$

Setting $\alpha = t$ and $s = 1$, by Theorem 1, the mapping $(x,t) \mapsto \tilde{\phi}(x,t,1) = \phi(x,t)$ is path differentiable jointly in (x,t). Let us compute a conservative Jacobian from Theorem 1. The differential inclusion in (2) can be expressed blockwise. Fix $x_0 \in \mathbb{R}^p$ and $\alpha_0 \in \mathbb{R}$ initial conditions for (11) and denote by $Y\colon [0,1] \to \mathbb{R}^p$ the solution to (11), note that $\alpha(t) = \alpha_0$ for all $t \in [0,1]$. Then, one has, for all $t \in [0,1]$, $Y(s) = X(\alpha_0 s)$ where X is the solution to (1) starting at x_0. Moreover, one has

$$\begin{pmatrix} \dot{V}_1(s) & \dot{V}_2(s) \\ \dot{V}_3(s) & \dot{V}_4(s) \end{pmatrix} \in \begin{pmatrix} \alpha_0 J_F(Y(s))V_1(s) + F(Y(s))V_3(s) & \alpha_0 J_F(Y(s))V_2(s) + F(Y(s))V_4(s) \\ 0 & 0 \end{pmatrix}, \tag{14}$$

where $V_1 \in \mathbb{R}^{p\times p}$ and $V_1(0)$ is the identity, $V_2 \in \mathbb{R}^{p\times 1}$ and $V_2(0) = 0$, $V_3 \in \mathbb{R}^{1\times p}$ and $V_3(0) = 0$, $V_4 \in \mathbb{R}$ and $V_4(0) = 1$. It follows that $V_3 = 0$ and $V_4 = 1$ for all t and

$$\begin{aligned} \dot{V}_1(s) &\in \alpha_0 J_F(Y(s))V_1(s) = \alpha_0 J_F(X(\alpha_0 s))V_1(s) \\ \dot{V}_2(s) &\in \alpha_0 J_F(Y(s))V_2(s) + F(Y(s)). \end{aligned} \tag{15}$$

The two dynamics are independant. Furthermore, solutions of the first line are also solutions of (2) modulo a simple time rescaling by a factor α_0. This is more explicitly written $V_1(s) \in U(x_0, \alpha s)$ for all $s \in [0,1]$, where U is given in (5). Conversely, any $V \in U(x_0, \alpha s)$ is related to a solution of the first line of (15). Let us show that $V_2\colon t \mapsto sF(X(\alpha_0 s))$ is the unique solution to the second line. By path differentiability of F, the function $s \mapsto F(X(\alpha_0 s))$ is differentiable for almost all t, such that

$$\begin{aligned} \frac{d}{ds}F(X(\alpha_0 s)) &= J(X(\alpha_0 s))\frac{d}{ds}X(\alpha_0 s)) && \forall J \in J_F(X(\alpha_0 s)) \\ &= \alpha_0 J(X(\alpha_0 s))F(X(\alpha_0 s)) && \forall J \in J_F(X(\alpha_0 s)) \end{aligned}$$

The function $s \mapsto sF(X(\alpha_0 s))$ is absolutely continuous and multiplication by s is a differentiable operation. Then, for almost all $s \in [0,1]$, substituting Y for X

$$\frac{d}{ds}[sF(Y(s))] = \alpha_0 J(Y(s))\,[sF(Y(s))] + F(Y(s)) \qquad \forall J \in J_F(Y(s)) \tag{16}$$

Now, given a measurable selection in $s \mapsto S(s) \in J_F(Y(s))$, the function $(s, V_2) \mapsto S(s)V_2$ is Lipschitz in its second argument, so that the corresponding solution V_2 in (15) is unique [24, Theorem 2, §1, Chapter 1]. Moreover, by (16), since $0F(Y(0)) = 0$, we have

$V_2(s) = sF(Y(s))$ for all $s \in [0,1]$. This shows that any solution to (14) is given by $V_3 = 0$, $V_4 = 1$, and for all $s \in [0,1]$,

$$V_1(s) \in U(x_0, \alpha_0 s)$$
$$V_2(s) = sF(X(\alpha_0 s)).$$

Thanks to Theorem 1, we have

$$(x, \alpha) \rightrightarrows (U(x,\alpha), F(\phi(x,\alpha))),$$

is conservative for the mapping $(x,\alpha) \mapsto \tilde{\phi}(x,\alpha,1)$. Using the fact that $\phi(x,\alpha) = \tilde{\phi}(x,\alpha,1)$ for all $x \in \mathbb{R}^p$, $\alpha \in \mathbb{R}$, this proves the desired result using $\alpha = t$. □

5 Consequences: backward and forward derivatives

In this section, we focus on an optimization of integral costs under ODE constraint and prove that, as soon as the ODE vector field and the integrand are path differentiable, then the integral cost is itself path differentiable. One should see these results as consecutive results of Sections 3 and 4. We provide further results about forward and backward derivatives propagation with a nonsmooth adjoint system.

5.1 Differentiation of a terminal cost

The following result is a direct consequence of Theorem 1 and the fact that product of conservative Jacobian is a also conservative Jacobian, as stated in [7, Lemma 5].

Corollary 2 *Let $\delta_T \colon \mathbb{R}^p \to \mathbb{R}$ be locally Lipschitz and path differentiable. Let $D_{\delta_T} \colon \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ be conservative gradient for δ_T. Then the following set*

$$D_T \colon x \rightrightarrows \left\{ V^\top u,\, V \in U(x,T),\, u \in D_{\delta_T}(\phi(x,T)) \right\} \tag{17}$$

is a conservative gradient for $x \mapsto \delta_T(\phi(x,T))$.

5.2 Forward propagation of derivatives of integral costs

In this subsection, we show how our framework allows to compute forward derivatives of integral costs. Such results already exist in a nonsmooth context with other classes of functions such as the functions admitting lexicographic derivatives (see e.g., [4]). Note that this result will be instrumental in deriving a backward derivative propagation in the form of an adjoint system.

Theorem 2 *Let $\delta \colon \mathbb{R}^p \to \mathbb{R}$ be locally Lipschitz and path differentiable. Let $D_\delta \colon \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ be a conservative Jacobian for δ, with convex values. For $T > 0$, set*

$$\Delta(x) = \int_{t=0}^{t=T} \delta(\phi(x,t))dt. \tag{18}$$

Then the following set valued field is a conservative gradient for Δ,

$$D_\Delta \colon x \rightrightarrows \left\{ \int_{t=0}^{t=T} V(t)^\top w(t) dt,\, V \in \mathcal{U}(x),\ w \in \mathcal{W}(x) \right\} \tag{19}$$

where $\mathcal{W}(x)$ is the set of measurable selections $w(t) \in D_\delta(\phi(x,t))$ for all $t \in [0,T]$ and $x \in \mathbb{R}^p$. In particular V could be any solution of (2).

Proof : We prove first that D_Δ has a closed graph, nonempty values and is locally bounded. Both $\mathcal{U}(x)$ and $\mathcal{W}(x)$ are nonempty valued and locally bounded thanks to the local boundedness of D_δ, ϕ and $\mathcal{U}$ proved in Lemma 1. Therefore D_Δ is locally bounded and nonempty valued. Second, we sketch the proof of graph closedness. Consider a sequence $(x_k)_{k\in\mathbb{N}}$ converging to $\bar{x}$, and $(d_k)_{k\in\mathbb{N}}$ converging to $\bar{d}$, such that, for each $k \in \mathbb{N}$, the sequence $(d_k)_{k\in\mathbb{N}}$ is defined by

$$d_k = \left\{ \int_{t=0}^{t=T} V_k(t)^\top w_k(t) dt,\, V_k \in \mathcal{U}(x_k),\, w_k \in \mathcal{W}(x_k) \right\}.$$

The sequence $(V_k)_{k\in\mathbb{N}}$ is bounded and Lipschitz (as proven in Lemma 1) uniformly in k, therefore we can use the Arzelá-Ascoli's Theorem [6, Theorem 4.25]: up to a subsequence, V_k converges to a given $\bar{V}$ uniformly on $[0,T]$. As detailed in Remark 6, it holds that $\bar{V} \in \mathcal{U}(\bar{x})$. The sequence $(w_k)_{k\in\mathbb{N}}$ is bounded in $L^2([0,T])$ (and in $L^\infty([0,T])$) so it has a weakly convergent subsequence by [34, Theorem 17, Section 14] whose limit will be denoted by $\bar{w}\colon [0,T] \to \mathbb{R}^p$. Up to a convex combination, the convergence occurs strongly and therefore pointwise almost everywhere by invoking Mazur's Lemma [6, Corollary 3.8]. We deduce that $\bar{w}(t) \in D_\delta(\phi(\bar{x},t))$ for almost all $t \in [0,T]$ using the fact that D_δ has convex values. Combining uniform convergence of V_k to $\bar{V} \in \mathcal{U}(\bar{x})$ and weak convergence of w_k to $\bar{w}$, we have that $d_k \to \int_{t=0}^{t=T} \bar{V}(t)^\top \bar{w}(t) dt \in D_\Delta(\bar{x})$ and $\bar{d} \in D_\Delta(\bar{x})$ by uniqueness of the limit.

From now on, we fix a Borel measurable selection d_Δ such that $d_\Delta(x) \in D_\Delta(x)$ for all $x \in \mathbb{R}^p$. This means that, for all $x \in \mathbb{R}^p$, there is a continuous function $V_x \in \mathcal{U}(x)$ and a measurable function $w_x \in \mathcal{W}(x)$ such that

$$d_\Delta(x) = \int_{t=0}^{t=T} V_x(t)^\top w_x(t) dt.$$

Now, fix an absolutely continuous path $\gamma\colon [0,1] \to \mathbb{R}^p$. Since δ and ϕ are Lipschitz functions, and since γ is absolutely continuous, we have that

$$r \mapsto \Delta(\gamma(r)) := \int_{t=0}^{t=T} \delta(\phi(\gamma(r),t)) dt,$$

is absolutely continuous. By Corollary 2, for all $t \in [0,T]$, for a.e. $r \in [0,1]$,

$$\frac{\partial}{\partial r} \delta(\phi(\gamma(r),t)) = \dot{\gamma}(r)^\top M^\top v, \qquad \forall v \in D_\delta(\phi(\gamma(r),t)),\, \forall M \in U(\phi(\gamma(r),t)). \tag{20}$$

Denote by $E \subset [0,1]\times[0,T]$ the set where (20) holds. Let us show that this set is Lebesgue measurable.

Consider the function

$$f\colon (M,v,r,t) \mapsto \left\| \frac{\partial}{\partial r}\delta(\phi(\gamma(r),t)) - \dot\gamma(r)^T M^T v \right\|^2$$

if $\frac{\partial}{\partial r}\delta(\phi(\gamma(r),t))$ and $\dot\gamma(r)$ are well defined, and 1 otherwise. The function f is jointly Lebesgue measurable in (r,t) for fixed M and v and jointly continuous in (M,v) for fixed $(r,t) \in [0,1]\times\mathbb{R}_+$. Then, it is a Carathéodory function. Therefore, the function

$$\begin{aligned} \tilde f\colon (r,t) \mapsto \max \quad & \left\| \frac{\partial}{\partial r}\delta(\phi(\gamma(r),t)) - \dot\gamma(r)^T M^T v \right\|^2 \\ \text{s.t.} \quad & v \in D_\delta(\phi(\gamma(t),t)) \\ & M \in U(\phi(\gamma(r),t)) \end{aligned}$$

is Lebesgue measurable thanks to [2, Theorem 18.19]. The set $\{(r,t)\in[0,1]\times\mathbb{R}_+, \tilde f(r,t) = 0\}$ is Lebesgue measurable and corresponds to the set where (20) holds. Therefore, (20) holds on a jointly measurable set. Furthermore, since (20) holds for all $t \in [0,T]$ for almost all $r \in [0,1]$, E has actually full measure.

Now consider the set

$$S = \left\{ (r,t) \in [0,1]\times[0,T],\ \frac{\partial}{\partial r}\delta(\phi(\gamma(r),t)) = \dot\gamma(r)^\top V_{\gamma(r)}(t)^\top w_{\gamma(r)}(t) \right\}.$$

Clearly, $E \subset S$ so that $S^c \subset E^c$. Moreover, since E^c has zero measure we deduce that S^c has zero (Lebesgue) measure. Therefore S is measurable jointly in (r,t) and the function $(r,t) \to \dot\gamma(r)V_{\gamma(r)}(t)w_{\gamma(r)}(t)$ is also (Lebesgue) measurable [34, Proposition 3, Section 18.1]. From Lemma 5, we have that $r \mapsto \Delta(\gamma(r))$ is absolutely continuous and for almost all $r \in [0,1]$,

$$\begin{aligned} \frac{d}{dr}\Delta(\gamma(r)) &= \int_{t=0}^{t=T} \frac{\partial}{\partial r}\delta(\phi(\gamma(r),t))dt \\ &= \int_{t=0}^{t=T} \dot\gamma(r)^\top V_{\gamma(r)}(t)^\top w_{\gamma(r)}(t)dt \\ &= \dot\gamma(r)^\top \int_{t=0}^{t=T} V_{\gamma(r)}(t)^\top w_{\gamma(r)}(t)dt \\ &= \dot\gamma(r)^\top d_\Delta(r). \end{aligned}$$

Note that d_Δ was an arbitrary measurable selection in D_Δ. Since D_Δ has a closed graph, it admits a countable Castaing representation ([2, Corollary 18.14] and [2, Theorem 18.20]). Then Lemma 8 applies and conservativity is proved, which leads to the desired result. □

5.3 Path differentiable adjoint method for integral costs

We describe a path differentiable version of the adjoint method for integral costs under ODE constraints.

Corollary 3 *Let $\delta, \delta_T \colon \mathbb{R}^p \mapsto \mathbb{R}$ be locally Lipschitz and path differentiable functions. Let $D_\delta \colon \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ and $D_{\delta_T} \colon \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ be conservative Jacobians for δ and δ_T, respectively where D_δ has convex values.*

For any $x \in \mathbb{R}^p$, any $w \colon [0,T] \to \mathbb{R}^p$ measurable such that $w(t) \in D_\delta(\phi(x,t))$ for all $t \in [0,T]$, any $J \colon [0,T] \to \mathbb{R}^{p \times p}$ measurable such that $J(t) \in J_F(\phi(x,t))$ for all $t \in [0,T]$ and any $u \in D_{\delta_T}(\phi(x,T))$, the unique absolutely continuous solution $\lambda \colon [0,T] \to \mathbb{R}^p$ to the system

$$\begin{aligned} \dot{\lambda}(t) &= -w(t) - J(t)^\top \lambda(t), \\ \lambda(T) &= u \end{aligned} \tag{21}$$

satisfies $\lambda(0) \in D_\Delta(x) + D_T(x)$ where D_Δ and D_T are defined in Corollary 2 and Theorem 2.

Proof : Fix $x \in \mathbb{R}^p$. Fix w and J as in the statement of the corollary. This defines a unique $M \in \mathcal{U}(x)$ by solving (2) $\dot{M}(t) = J(t)M(t)$ with $M(0) = I$ [24, Theorem 2, §1, Chapter 1].

For any absolutely continuous function $\lambda \colon [0,T] \to \mathbb{R}^p$, we have

$$\int_{t=0}^{t=T} M(t)^\top w(t) dt = \int_{t=0}^{t=T} M(t)^\top w(t) + \left(J(t)M(t) - J(t)M(t)\right)^\top \lambda(t) dt$$

Using Lemma 9, we have

$$\int_{t=0}^{t=T} (J(t)M(t))^\top \lambda(t) dt = \int_{t=0}^{t=T} \dot{M}(t)^\top \lambda(t) dt = -\lambda(0) + M(T)^\top \lambda(T) - \int_0^\top M(t)^\top \dot{\lambda}(t) dt.$$

Hence, we have for any $u \in D_T(\phi(x,T))$,

$$\begin{aligned} & M(T)^\top u + \int_{t=0}^{t=T} M(t)^\top w(t) dt \\ = & \int_{t=0}^{t=T} M(t)^\top \left(w(t) + J(t)^\top \lambda(t) + \dot{\lambda}(t) \right) dt + M(0)\lambda(0) + M(T)^\top (u - \lambda(T)) \end{aligned}$$

The latter holds for any absolutely continuous function λ, and in particular, using [24, Theorem 2, §1, Chapter 1], one can choose λ as the unique absolutely continuous solution to

$$\dot{\lambda}(t) = -w(t) - J(t)^\top \lambda(t) \tag{22}$$

$$\lambda(T) = u. \tag{23}$$

Using the fact that $M(0)$ is the identity, one has finally

$$M(T)^\top u + \int_{t=0}^{t=T} M(t)^\top w(t) dt = M(0)\lambda(0) = \lambda(0).$$

The term $\lambda(0)$ being defined as the sum of two specific elements in D_Δ and D_T (see Corollary 2 and Lemma 2), this means that $\lambda(0) \in D_\Delta(x) + D_T(x)$, which concludes the proof. □

Remark 7 *The system* (21) *is typically solved backward in time. Setting* $g\colon s \mapsto \lambda(T-s)$, *we have, for all* $s \in [0,T]$,

$$\begin{aligned} g(0) &= u \\ \dot{g}(s) &= -\dot{\lambda}(T-s) = w(T-s) + J(T-s)\lambda(T-s), \end{aligned}$$

which is the backpropagation equation.

6 Minimization of integral costs with parameterized ODEs constraints

This section is centered around problem (3). The results combine conservative calculus rules with the elements developed in Section 5.

6.1 Problem setting

We consider the optimization problem described in (3) and introduce further notations. First the constraints in (3) relate to the following parametrized ODE, given $T > 0$, for all $t \in [0,T]$

$$\dot{Z}(t) := H(Z(t), \theta), Z(0) = z, \tag{24}$$

where $\theta \in \mathbb{R}^m$ denotes a vector of parameters and $H\colon \mathbb{R}^{p+m} \to \mathbb{R}^p$ is a Lipschitz path differentiable function. We assume that $J_H\colon \mathbb{R}^{p+m} \rightrightarrows \mathbb{R}^p$ is a conservative Jacobian for H and that it is bounded. We denote $\psi(z,\theta,t) \in \mathbb{R}^p$ the flow associated to the ODE (24). Throughout this section, z will be a fixed initial condition in $\mathbb{R}^p$.

We denote by $L_I\colon \mathbb{R}^m \to \mathbb{R}$ the integral part of the loss in (3),

$$L_I\colon \theta \mapsto \int_0^T \ell(\psi(z,\theta,t))dt.$$

Recall that $\ell\colon \mathbb{R}^p \to \mathbb{R}$ is a locally Lipschitz, path differentiable functions from $\mathbb{R}^p$ to $\mathbb{R}$, we further assume that it admits a conservative gradients, $D_\ell\colon \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ with convex values.

Furthermore, we denote by $L_T\colon \mathbb{R}^m \to \mathbb{R}$ the integral part of the loss in (3),

$$L_T\colon \theta \mapsto \ell_T(\psi(z,\theta,T)).$$

Recall again that $\ell_T\colon \mathbb{R}^p \to \mathbb{R}$ a locally Lipschitz, path differentiable function and assume that it admits a conservative gradients $D_{\ell_T}\colon \mathbb{R}^p \rightrightarrows \mathbb{R}^p$.

With a slight abuse of notations, problem (3) can be reformulated equivalently as an unconstrained minimization problem with cost $L := L_I + L_T$ with respect to variable θ, that is,

$$\inf_{\theta\in\mathbb{R}^m} L(\theta) = \inf_{\theta\in\mathbb{R}^m} \int_0^T \ell(\psi(z,\theta,t))dt + \ell_T(\psi(z,\theta,T)). \tag{25}$$

The purpose of this section is to specify the results presented in Section 5 to the parametrized ODE (24) and cost (25) in order to address problem (3). This will result in expressions for conservative gradients, D_I for L_I and D_T for L_T. Setting $D_L = D_I + D_T$, the sum of these conservative gradients, we obtain a conservative gradient for L [7, Corollary 4]. To this end, we see the parametrized flow of (24) as an unparametrized flow in a lifted space and justify formal differentiation operations using conservative calculus rules [7] to make connections with results presented in Section 5. This results in an adjoint method which can in turn be used as an oracle for D_L. This allows to provide a convergence result for the corresponding small step first order optimization method to seek solutions of problem (25).

6.2 Conservative Jacobian of the flow

The following corollary reformulates Theorem 1 in the context of system (24).

Corollary 4 *Let ψ be defined as in equation (24) with $H\colon \mathbb{R}^{p+m} \to \mathbb{R}^p$ a Lipschitz path-differentiable function and $J_H\colon \mathbb{R}^{p+m} \rightrightarrows \mathbb{R}^p$ a bounded conservative Jacobian. Consider the matrix differential inclusion with unknown $M \in \mathbb{R}^{p\times m}$, for almost all $t \in [0,T]$*

$$\begin{aligned} \dot{M}(t) &= J_z(t)M(t) + J_\theta(t) \\ \begin{pmatrix} J_z(t) & J_\theta(t) \end{pmatrix} &\in J_H(\psi(z,\theta,t),\theta), \end{aligned} \tag{26}$$

with initial condition $M(0) = 0_{pm} \in \mathbb{R}^{p\times m}$, where we used block matrix notations in the second line. Then the set $\{M(T),\ M \text{ solution to } (26)\}$ forms a conservative Jacobian for $\theta \mapsto \psi(z,\theta,T)$.

In particular, for any $\ell_T\colon \mathbb{R}^p \to \mathbb{R}$, path differentiable with conservative gradient D_{ℓ_T} and for any $z \in \mathbb{R}^p$, the following set valued map

$$D_T\colon (z,\theta) \rightrightarrows \left\{M(T)^T u,\ \ M \text{ solution to } (26),\ u \in D_{\ell_T}(\psi(z,\theta,T))\right\}$$

is a conservative gradient for the function $L_T\colon \theta \mapsto \ell_T(\psi(z,\theta,T))$.

Proof : We first introduce notations allowing to interpret the system (24) as a system of the form (1) on an extended state space $\mathbb{R}^{p+m}$. We rewrite (24) as follows, for all $t \in [0,T]$:

$$\dot{Z}(t) = H(Z(t),\theta(t)),\ \dot{\theta}(t) = 0, \tag{27}$$

$$Z(0) = z,\ \theta(0) = \theta, \tag{28}$$

We consider $X = \begin{pmatrix} Z \\ \theta \end{pmatrix} \in \mathbb{R}^{p+m}$, the concatenation of the two variables $Z \in \mathbb{R}^p$ and $\theta \in \mathbb{R}^m$. We set, for all such X,

$$F(X) = \begin{pmatrix} H(Z,\theta) \\ 0 \end{pmatrix} \in \mathbb{R}^{p+m}.$$

With these notations, (27) is equivalently rewritten as follows, for all $t \in [0,T]$

$$\dot{X}(t) := F(X(t)),$$

In this case, (27) is in the same form than the one given in (1), and the parameter θ is now seen as an initial condition. Setting

$$J_F(z,\theta)\colon (z,\theta) \rightrightarrows \begin{pmatrix} J_H(z,\theta) \\ 0 \end{pmatrix},$$

we have that J_F is a conservative Jacobian for F since J_H is a conservative Jacobian for H. The variational inclusion (2) for (24) can be written as follows, for a.e. $t \in [0,T]$

$$\dot{M} = \begin{pmatrix} \dot{M}_1 & \dot{M}_2 \\ \dot{M}_3 & \dot{M}_4 \end{pmatrix} \in \begin{pmatrix} J_H(\psi(z,\theta,t),\theta)M \\ 0 \end{pmatrix} \tag{29}$$

where $M_1(0) = I_p \in \mathbb{R}^{p\times p}$, $M_2(0) = 0_{pm} \in \mathbb{R}^{p\times m}$, $M_3(0) = 0_{mp} \in \mathbb{R}^{m\times p}$ and $M_4(0) = I_m \in \mathbb{R}^{m\times m}$. From this equation, M_3 and M_4 remain constant. From [7, Lemma 4] and Theorem 1, the concatenation $\begin{pmatrix} M_1 & M_2 \end{pmatrix}$ for all solutions to (29) forms a conservative Jacobian for $(z,\theta) \mapsto \psi(z,\theta,T)$. Let us express this equation in a simpler form.

For any $t \in [0,T]$, and any $J \in J_H(\psi(z,\theta,t),\theta)$, writing $J = \begin{pmatrix} J_z & J_\theta \end{pmatrix}$ where $J_z \in \mathbb{R}^{p\times p}$ and $J_\theta \in \mathbb{R}^{p\times m}$, with $M_3 = 0_{mp} \in \mathbb{R}^{m\times p}$ and $M_4 = I_m \in \mathbb{R}^{m\times m}$, we have

$$JM = \begin{pmatrix} J_z & J_\theta \end{pmatrix} \begin{pmatrix} M_1 & M_2 \\ M_3 & M_4 \end{pmatrix} = \begin{pmatrix} J_z M_1 & J_z M_2 + J_\theta \end{pmatrix}.$$

Equation (29) is equivalently rewritten, for a.e. $t \in [0,T]$

$$\begin{aligned} \dot{M}_1(t) &= J_z(t) M_1(t) \\ \dot{M}_2(t) &= J_z(t) M_2(t) + J_\theta(t) \\ \begin{pmatrix} J_z(t) & J_\theta(t) \end{pmatrix} &\in J_H(\psi(z,\theta,t),\theta), \end{aligned} \tag{30}$$

with $M_1(0) = I_p \in \mathbb{R}^{p\times p}$ and $M_2(0) = 0_{pm} \in \mathbb{R}^{p\times m}$.

Using Theorem 1, we have proved that concatenations of the form $\begin{pmatrix} M_1(T) & M_2(T) \end{pmatrix}$ where M_1 and M_2 are solutions of (30) form a conservative Jacobian for $(z,\theta) \mapsto \psi(z,\theta,T)$. Focusing on the dependency in θ for fixed z, and invoking Lemma 11, component M_2 form a conservative field for $\theta \mapsto \psi(z,\theta,T)$. This proves the corollary. □

Remark 8 *Using the latter argument, one can also include a dependency of the initial condition in θ, that is $Z(0) = z_0(\theta)$. It suffices to notice that this is a composition of the function $(z,\theta) \to \psi(z,\theta,T)$ for which we have a conservative Jacobian from* (30) *and the function $\theta \to (z_0(\theta),\theta)$ for which we have a conservative Jacobian as long as we know a conservative Jacobian for $\theta \to z_0(\theta)$. We may then apply the composition rule [7, Lemma 5]. It is also possible to include further dependency in θ for ℓ and ℓ_T with similar lifting techniques.*

6.3 Differentiation of integral costs and adjoint method

The following corollary is a reformulation of Theorem 2 in the context of (24), based on Corollary 4.

Corollary 5 *Let ψ be defined as in equation (24) with $H\colon \mathbb{R}^{p+m} \to \mathbb{R}^p$ a Lipschitz path-differentiable function and $J_H\colon \mathbb{R}^{p+m} \rightrightarrows \mathbb{R}^p$ a bounded conservative Jacobian. Let $\ell\colon \mathbb{R}^p \to \mathbb{R}$ be a locally Lipschitz, path differentiable functions with conservative gradients D_ℓ with convex values. For any $z \in \mathbb{R}^p$, consider the function $L_I : \theta \in \mathbb{R}^m \to \int_0^T \ell(\psi(z,\theta,t)dt$ and the set valued field:*

$$D_I\colon (z,\theta) \rightrightarrows \left\{ \int_{t=0}^{t=T} M(t)^\top w(t)dt,\ M \in \mathcal{U}(z,\theta),\ w \in \mathcal{W}(z,\theta) \right\}$$

where $\mathcal{U}(z,\theta)$ is the set of solutions to (26) and $\mathcal{W}(z,\theta)$ is the set of measurable functions $w\colon [0,T] \to \mathbb{R}^p$ such that $w(t) \in D_\ell(\psi(z,\theta,t))$ for all $t \in [0,T]$. For any $z \in \mathbb{R}^p$, $\theta \rightrightarrows D_I(z,\theta)$ is a conservative gradient for L.

Proof : Using Theorem 2 and the expressions given in (29), one knows that the set valued map:

$$(z,\theta) \rightrightarrows \left\{ \int_{t=0}^{t=T} \begin{pmatrix} M_1^\top(t) & 0 \\ M_2^\top(t) & I_p \end{pmatrix} \begin{pmatrix} w(t) \\ 0 \end{pmatrix} dt, M_1 \text{ and } M_2 \text{ solutions to (30)}, w \in D_\ell(\psi(z,\theta,t)) \right\}$$

is a conservative field for the function $(z,\theta) \mapsto \int_0^T \ell(\psi(z,\theta,t))dt$. Using Lemma 11, one can deduce that, for every $z \in \mathbb{R}^p$, the set valued field $\theta \rightrightarrows D_I(z,\theta)$ is a conservative gradient for L_I, which concludes the proof. □

We have the following adaptation of the adjoint of Corollary 3.

Corollary 6 *Let ψ be defined as in equation (24) with $H\colon \mathbb{R}^{p+m} \to \mathbb{R}^p$ a Lipschitz path-differentiable function and $J_H\colon \mathbb{R}^{p+m} \rightrightarrows \mathbb{R}^p$ a bounded conservative Jacobian. Let $\ell\colon \mathbb{R}^p \to \mathbb{R}$ and $\ell_T\colon \mathbb{R}^p \to \mathbb{R}$ be locally Lipschitz, path differentiable functions from $\mathbb{R}^p$ to $\mathbb{R}$ with respective conservative gradients D_ℓ and D_T.*

For any $z \in \mathbb{R}^p$, $\theta \in \mathbb{R}^m$, any $w\colon [0,T] \to \mathbb{R}^p$ measurable such that $w(t) \in D_\ell(\psi(z,\theta,t))$ for all $t \in [0,T]$, any $J_z\colon [0,T] \to \mathbb{R}^{p\times p}$ and $J_\theta\colon [0,T] \to \mathbb{R}^{p\times m}$, measurable such that $\begin{pmatrix} J_z(t) & J_\theta(t) \end{pmatrix} \in J_H(\psi(z,\theta,t),\theta)$ for all $t \in [0,T]$ and any $u \in D_T(\psi(z,\theta,T))$, the unique absolutely continuous solution $\lambda\colon [0,T] \to \mathbb{R}^p$ to the system

$$\begin{aligned} \dot{\lambda}(t) &= -w(t) - J_z(t)^\top \lambda(t), \\ \lambda(T) &= u \end{aligned} \tag{31}$$

satisfies $\int_0^T J_\theta(t)^\top \lambda(t)dt \in D_I(z,\theta) + D_T(z,\theta)$ which is an element of a conservative gradient for the loss function L in (25).

Proof : Fix $z \in \mathbb{R}^p$ and $\theta \in \mathbb{R}^m$. Fix w and J_z as in the statement of the theorem. This defines a unique $M \in \mathcal{U}(z,\theta)$ by solving (31) $\dot{M}(t) = J_z(t)M(t) + J_\theta(t)$ [24, Theorem 2, §1, Chapter 1].

For any absolutely continuous function $\lambda\colon [0,T] \to \mathbb{R}^p$, we have

$$\int_{t=0}^{t=T} M(t)^\top w(t)dt = \int_{t=0}^{t=T} M(t)^\top w(t) + \left(J_z(t)M(t) + J_\theta(t) - J_z(t)M(t) - J_\theta(t)\right)^\top \lambda(t)dt$$

Using Lemma 9, we have

$$\int_{t=0}^{t=T} (J_z(t)M(t) + J_\theta(t))^\top \lambda(t)dt = \int_{t=0}^{t=T} \dot{M}(t)^\top \lambda(t)dt = -M(0)^\top \lambda(0) + M(T)^\top \lambda(T) - \int_0^\top M(t)^\top \dot{\lambda}(t)dt.$$

Hence, since $M(0) = 0$, we have, for any $u \in D_T(\psi(z,\theta,T))$,

$$\begin{aligned} &M(T)^\top u + \int_{t=0}^{t=T} M(t)^\top w(t)dt \\ &= \int_{t=0}^{t=T} M(t)^\top \left(w(t) + J_z(t)^\top \lambda(t) + \dot{\lambda}(t)\right) dt + \int_{t=0}^{t=T} J_\theta(t)^\top \lambda(t)dt + M(T)^\top (u - \lambda(T)) \end{aligned}$$

The latter holds for any absolutely continuous function λ, and in particular, using [24, Theorem 2, §1, Chapter 1], one can choose λ as the unique absolutely continuous solution to

$$\dot{\lambda}(t) = -w(t) - J_z(t)^\top \lambda(t), \tag{32}$$
$$\lambda(T) = u. \tag{33}$$

One has finally

$$M(T)^\top u + \int_{t=0}^{t=T} M(t)^\top w(t)dt = \int_{t=0}^{t=T} J_\theta(t)^\top \lambda(t)dt.$$

The term $\int_{t=0}^{t=T} J_\theta(t)^\top \lambda(t)dt$ being defined as the sum of two specific elements in D_L and D_T (see Corollary 2 and Lemma 2), this means that $\int_{t=0}^{t=T} J_\theta(t)^\top \lambda(t)dt \in D_L(z,\theta) + D_T(z,\theta)$, which concludes the proof. □

6.4 Small step method for optimization

Recall that we are interested in the following problem, for a fixed $z \in \mathbb{R}^p$ and $T > 0$

$$\inf_{\theta \in \mathbb{R}^m} \int_0^T \ell(\psi(z,\theta,t))dt + \ell_T(\psi(z,\theta,T)),$$

where the integral cost is L_I and the terminal cost is L_T and their sum is denoted by L. Given $\theta \in \mathbb{R}^m$, Corollary 6 allows to obtain an element of $D_L(\theta) = D_I(\theta) + D_T(\theta)$.

This constitutes relevant first order information. Indeed, for example the results borrowed from [7, Theorem 1, Corollary 1] ensure the following properties

$$\begin{aligned} \partial^c L(\theta) &\subset \mathrm{conv}\{D_L(\theta)\}, && \text{for all } \theta \in \mathbb{R}^m, \\ D_L(\theta) &= \{\nabla L(\theta)\}, && \text{for Lebesgue almost all } \theta \in \mathbb{R}^m, \end{aligned}$$

where ∂^c denotes the Clarke subgradient [18, Chapter 2]. As detailed in [7, Section 6], elements of D_L can be used in place of gradients in an optimization context. Given a sequence of positive step sizes $(\alpha_k)_{k\in\mathbb{N}}$ and $\theta_0 \in \mathbb{R}^m$, one can iterate the following recursion

$$\theta_{k+1} = \theta_k - \alpha_k g_k \tag{34}$$

$$g_k \in D_L(\theta_k). \tag{35}$$

We insist on the fact that g_k can be obtained for example using Corollary 6. Recall that the set of accumulation points of the sequence $(\theta_k)_{k\in\mathbb{N}}$ is the set of $\bar\theta$ such that, for all $r > 0$, the set $\{i \in \mathbb{N}, \|\theta_i - \bar\theta\| < r\}$ is infinite. The following result is a consequence of [9, Theorem 6] about bounded sequences of the form (34). This uses a weaker notion of accumulation point to characterize the fact that the sequence is essentially attracted by critical points, that is points which comply with the necessary optimality condition $0 \in \mathrm{conv}\{D_L(\theta)\}$.

Corollary 7 *Assume that $\alpha_k \to 0$ and $\sum_{i=0}^k \alpha_i \to +\infty$ as $k \to +\infty$. Assume furthermore that the sequence $(\theta_k)_{k\in\mathbb{N}}$ given by (34) remains bounded. Then the set*

$$\Omega = \left\{ \bar\theta \in \mathbb{R}^m,\ \forall r > 0, \limsup_{N\to\infty} \frac{\sum_{0\le i\le N,\, \|\theta_i - \bar\theta\| < r} \alpha_i}{\sum_{0\le i\le N} \alpha_i} > 0 \right\}$$

is non empty and satisfies $\Omega \subset \mathrm{crit}_L$, where crit_L is the set of $\theta \in \mathbb{R}^m$ complying with the optimality condition $0 \in \mathrm{conv}\{D_L(\theta)\}$.

In the latter corollary, Ω is termed the "essential accumulation set" of the sequence [9]. This is a subset of the set of usual accumulation points of the sequence, corresponding to those accumulation points for which the sequence spends a significant amount of time, as measured with respect to the sum of neighboring step sizes. This result illustrates the minimizing behavior of the sequence (34). The result is quite weak, but provides a solid ground regarding the use of the proposed conservative adjoint method for finite dimensional optimization under ODE constraints. At this level of generality, stronger assumptions, such as Sard type conditions related to the loss L, would be required to obtain stronger statements [33]. This will be a topic of future research.

7 Conclusion

In this article, we have proved that flows of ODEs expressed with vector fields that are path differentiable are also path differentiable. The proof of this results stands on the fact that the set valued mapping obtained from the variational inclusion is a conservative

Jacobian. This allows to develop a conservative calculus for integral costs, similar as one would have in the smooth case. This culminates with a conservative version of the adjoint method to propagate derivatives backward and obtain gradients of integral costs, with a considerable reduction of the size of the differential inclusion to be solved. A consequence of these results is the fact small step methods of gradient type for minimizing integral costs are attracted by solutions of a certain optimality condition for such problems with path differentiable data.

The developments provided in this work could be extended to parametric partial differential equations (PDEs), this was actually one of the original motivations for the proposed developments. The question of path differentiability of PDEs could be considered under regularity assumption by exhibiting a conservative Jacobian in a way similar to what is proposed for ODE flows. One should probably start with specific sub-classes of PDEs, for instance hyperbolic ones [20].

Acknowledgements

Edouard Pauwels acknowledges the support of AI Interdisciplinary Institute ANITI funding, through the French "Investing for the Future - PIA3" program under the Grant agreement ANR-19-PI3A0004, Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant numbers FA9550-19-1-7026, FA9550-18-1-0226, and ANR MaSDOL 19-CE23-0017-01.

A Technical results

This appendix is devoted to the statement and the proof of some crucial results for our analysis.

The following result is about the Borel measurability of partial derivatives. Its proof is inspired by [23, Theorem 3.2].

Lemma 3 (Measurability of partial derivatives) *Consider a function $G : (x, y) \in \mathbb{R}^n \times \mathbb{R} \to G(x, y) \in \mathbb{R}^m$. Suppose that, for all $x \in \mathbb{R}^n$, $y \mapsto G(x, y)$ is absolutely continuous, and for all y, $x \mapsto G(x, y)$ is Borel measurable. Then, the function $(x, y) \mapsto \frac{\partial}{\partial y} G(x, y)$ defined on a set of full Lebesgue measure and is measurable. Futhermore, for all x, the function $y \mapsto \frac{\partial}{\partial y} G(x, y)$ exists for almost all t.*

Proof :

As a Carathéodory function, G is jointly Borel measurable [2, Lemma 4.51] and therefore it is Lebesgue measurable. Consider the functions:

$$G_y^u(x, y) := \lim_{h \to 0} \sup \frac{G(x, y + h) - G(x, y)}{h} \tag{36}$$

and

$$G_y^l(x,y) := \liminf_{h\to 0} \frac{G(x,y+h)-G(x,y)}{h}, \tag{37}$$

with $h \in \mathbb{R}$.

Using the continuity of G in its second argument, one has the equivalent definition (36) by

$$G_y^u(x,y) = \lim_{k\to+\infty} \sup_{0<|h|<\frac{1}{k}, h\in\mathbb{Q}} \frac{G(x,y+h)-G(x,y)}{h}, \tag{38}$$

For all $k \geq 1$, the set $\{h \in \mathbb{Q} \mid 0 < |h| < \frac{1}{k}\}$ is countable and therefore, using Lemma [34, Corollary 7, Section 18.1], the supremum in (38) is Lebesgue measurable as the countable supremum of measurable functions. This implies that the sequence

$$\left\{ \sup_{0<|h|<\frac{1}{k}} \frac{G(x,y+h)-G(x,y)}{h} \right\}_{k\in\mathbb{N}}$$

is a bounded, decreasing sequence of measurable functions and it has therefore a pointwise limit everywhere. Using [34, Chapter 18, Corollary 7], the pointwise limit of measurable functions is a measurable function. This implies that G_y^u defined in (38) or equivalently in (36) is Lebesgue measurable. By a similar argument, one can show that (37) is also measurable.

It remains to prove that the function $(x,y) \mapsto \frac{\partial}{\partial y}G(x,y)$ exists Lebesgue almost everywhere. To do so, consider the Lebesgue measurable set

$$A := \left\{(x,y) \in \mathbb{R}^{n+1}, G_y^u(x,y) = G_y^l(x,y),\, G_y^l(x,y) \neq \pm\infty\right\}$$

Its complement A^c is the subset of $\mathbb{R}^{n+1}$ where $G_y^u(x,y) \neq G_y^l(x,y)$ or $G_y^l(x,y) = \pm\infty$. By Fubini's theorem [34, Theorem 16 Section 20.2],

$$\int_{(x,y)} \mathbb{I}_A(x,y)dxdy = \int_x \left(\int_y \mathbb{I}_A(x,y)dy \right) dx = 0.$$

where $\mathbb{I}_A$ is the function such that $\mathbb{I}_A(x,y) = 1$ if $(x,y) \in A$ and 0 otherwise. By Lebesgue integration theorem [34, Theorem 10, Section 6.5], the inner integral is zero because of the absolute continuity in y for fixed x. This shows that the function $(x,y) \mapsto \frac{\partial}{\partial y}G(x,y)$ exists for Lebesgue almost all $(x,y) \in \mathbb{R}^{n+1}$. Furthermore, up to an arbitrary measurable choice outside of its domain of definition, it is a Lebesgue measurable function. This concludes the proof of Lemma 3. □

We also state a useful result which states that every lower semicontinuous functions are Borel measurable.

Lemma 4 *Given X a metric space, let $f : X \to \mathbb{R}$ be a lower semicontinuous real function. Then, it is Borel measurable.*

Proof : The function f being lower semicontinuous, the set $\{(x,c) \in X \times \mathbb{R} \mid c \geq f(x)\}$ is closed. It is in particular a Borel set, which implies that f is Borel measurable. This concludes the proof. □

The following result concerns derivatives of integrals. More precisely, it states and proves that the operators derivatives and integrals can be permutated. This result is closely related to the well-known Leibniz rule, but it concerns in our case absolutely continuous functions. It can be seen then as a general Leibniz rule.

Lemma 5 (General Leibniz rule) *Consider a Lipschitz function $F : \mathbb{R}^p \times X \to \mathbb{R}^p$ where X is a bounded interval of $\mathbb{R}$. Consider furthermore an absolutely continuous function $\gamma : [0,1] \to \mathbb{R}^p$. Then, $r \mapsto \int_X F(\gamma(r), s)ds$ is absolutely continuous and for a.e $r \in [0,1]$:*

$$\frac{d}{dr}\int_X F(\gamma(r), s)ds = \int_X \frac{\partial}{\partial r}F(\gamma(r), s)ds. \tag{39}$$

Proof : Since F is Lipchitz continuous and γ is absolutely continuous, then, for all $s \in X$, the function $r \mapsto F(\gamma(r), s)$ is absolutely continuous as the composition of a Lipschitz function with an absolutely continuous function. In particular, it is differentiable for a.e. $r \in [0,1]$.

Furthermore, since the function $(r, s) \mapsto F(\gamma(r), s)$ is continuous, one can prove that, due to Lemma 3, the function:

$$(r, s) \mapsto \frac{\partial}{\partial r}F(\gamma(r), s) \tag{40}$$

is well defined for all $s \in X$ and for a.e. $r \in [0,1]$. It is also jointly measurable in (r, s) (up to arbitrary values outside of its full measure domain of definition). Denoting by L a Lipschitz constant of F, we have we have for all s and almost all r

$$\left\| \frac{\partial}{\partial r}F(\gamma(r), s) \right\| \leq L\|\dot{\gamma}(r)\| \tag{41}$$

by definition of the derivative.

Then, using again the fact that, for all $s \in X$, $r \mapsto F(\gamma(r), s)$ is absolutely continuous, then one has, for all $s \in X$ and $r \in [0,1]$,

$$F(\gamma(r), s) - F(\gamma(0), s) = \int_0^r \frac{\partial}{\partial q}F(\gamma(q), s)dq. \tag{42}$$

Integrating the previous equation over the domain X leads to

$$\int_X [F(\gamma(r), s) - F(\gamma(0), s)]ds = \int_X \int_0^r \frac{\partial}{\partial q}F(\gamma(q), s)dqds.$$

Since the function given by (40) is jointly measurable, Fubini's theorem [2, Theorem 11.27] applies and one has for all $r \in [0,1]$

$$\int_X [F(\gamma(r), s) - F(\gamma(0), s)]ds = \int_0^r \int_X \frac{\partial}{\partial q}F(\gamma(q), s)dsdq.$$

This proves the desired result because (39) is a consequence of Lebesgue differentiation theorem [34, Section 6.5], for all q,

$$\left\| \int_X \frac{\partial}{\partial q}F(\gamma(q), s)ds \right\| \leq L\|\dot{\gamma}(q)\| \times \int_X ds$$

where right hand side is integrable because X is a bounded interval and γ is absolutely continuous (its derivative is integrable). This concludes the proof. □

Next, we provide a generalization of the Grönwall's inequality for absolutely continuous functions. The proof is inspired by the proof of [24, Theorem 2].

Lemma 6 (Grönwall's Lemma for absolutely continuous functions) *Let K be a constant and $f : \mathbb{R}_+ \to \mathbb{R}_+$ be absolutely continuous on $[0, T]$ such that, for a.e. $t \in [0, T]$:*

$$\frac{d}{dt} f(t) \leq K f(t). \tag{43}$$

Then, $f(t) \leq \exp(Kt) f(0)$ for all $t \in [0, T]$.

Proof : From the inequality, one has, for a.e. $t \in [0, T]$,

$$\left(\frac{d}{dt} f(t) - K f(t) \right) \exp(-Kt) = \frac{d}{dt} (f(t) \exp(-Kt)) \leq 0.$$

The function $t \mapsto f(t) \exp(-Kt)$ is absolutely continuous as a product of absolutely continuous functions with bounded domain. One deduces therefore that, for all $t \in [0, T]$, $f(t) \exp(-Kt) f(t) - f(0) \leq 0$. Then, for all $t \in [0, T]$, $f(t) \leq f(0) \exp(Kt)$, which concludes the proof of the lemma. □

Associated to this inequality, one may deduce a Grönwall inequality for integrable functions, as stated in the following lemma.

Lemma 7 (Grönwall's inequality for integrable functions) *Let K be a positive constant and $f : \mathbb{R}_+ \to \mathbb{R}_+$ be integrable on $[0, T]$, such that for all $t \in [0, T]$*

$$f(t) \leq K \int_0^t f(s) ds.$$

Then $f(t) = 0$ for all $t \in [0, T]$.

Proof : Let $G : t \mapsto \int_0^t f(s) ds$. This function is absolutely continuous, nonnegative (since f is nonnegative), and for almost all $t \in [0, T]$, one has $\frac{d}{dt} G(t) \leq K G(t)$. Then, one can apply Lemma 6 and deduce that, for all $t \in [0, T]$, $G(t) \leq \exp(Kt) G(0)$. Therefore, since $G(0) = 0$ one has $G(t) = 0$ for all $t \geq 0$, which implies that $f(t) = 0$ since $0 \leq f(t) \leq K G(t)$. This concludes the proof of the lemma. □

Finally, we provide a density result showing that for a set valued map D to be conservative for a given function f, it is sufficient to prove a conservativity relation for each element of a Castaing representation of D.

Lemma 8 (Measurable selections and density) *Consider a set-valued map $D : x \in \mathbb{R}^p \rightrightarrows D(x) \in \mathbb{R}^p$ that is locally bounded and has non empty values and a closed graph. Consider a sequence of measurable selections in D, denoted by $\{M_i\}_{i \in \mathbb{N}}$, such that, for all x*

$$D(x) = \overline{\{M_i(x)\}_{i \in \mathbb{N}}}.$$

Then, for a given function f; $\mathbb{R}^p \to \mathbb{R}$, if, for all $i \in \mathbb{N}$ and for all absolutely continuous path $\gamma : [0,1] \to \mathbb{R}^p$, one has, for all $t \in [0,1]$

$$f(\gamma(t)) - f(\gamma(0)) = \int_0^t M_i(\gamma(s))\dot{\gamma}(s)ds, \tag{44}$$

then D is a conservative Jacobian for f.

Proof : Fix an absolutely continuous path γ. By assumption, since D is locally bounded, we have that f is locally Lipschitz and therefore $f \circ \gamma$ is absolutely continous.

For each $i \in \mathbb{N}$, by Lebesgue integration theorem,

$$\frac{d}{dt} f(\gamma(t)) = M_i(\gamma(t))\dot{\gamma}(t), \quad \text{for a.e. } t \in [0,1]. \tag{45}$$

For $i \in \mathbb{N}$, consider the set defined by

$$E_i := \{t \in [0,1] \text{ such that (45) holds}\} \subset [0,1].$$

The set E_i has full measure for each $i \in \mathbb{N}$. Then, we define $E := \cap_{i\in\mathbb{N}} E_i$. Since E is a countable intersection of full measure sets, its complement E^c has a Lebesgue measure zero.

Therefore, for all $t \in E$, (45) holds for any $i \in \mathbb{N}$. Since $\overline{\{M_i(\gamma(t))\}_{i\in\mathbb{N}}} = D(\gamma(t))$, we have

$$\frac{d}{dt} f(\gamma(t)) = W\gamma(t), \ \forall W \in D(\gamma(t)), \tag{46}$$

for all $t \in E$. Since E has full measure and γ is an arbitrary absolutely continuous path, this shows that D is a conservative Jacobian for f and then concludes the proof of the Lemma. □

Another important result is the integration by parts formula for absolutely continuous functions, that is stated as follows.

Lemma 9 (Integration by parts) *Consider two absolutely continuous functions $f, g\colon [0,T] \to \mathbb{R}^p$. Then, the following integration by parts formula holds*

$$\int_0^T f(t)\dot{g}(t)dt = f(T)g(T) - f(0)g(0) - \int_0^T \dot{f}(t)g(t)dt. \tag{47}$$

Proof : Since f and g are absolutely continuous with bounded domains, then the product fg is absolutely continuous. By definition of absolutely continuous functions, one has

$$\int_0^T \frac{d}{dt}(fg)(t)dt = f(T)g(T) - f(0)g(0).$$

Noticing that, for a.e. $t \in [0,T]$, $\frac{d}{dt}(fg)(t) = \dot{f}(t)g(t) + f(t)\dot{g}(t)$, one obtains the desired result. □

We also state and prove a result stating that the space of Lipschitz functions whose domain is a given bounded interval is σ-compact.

Lemma 10 *The space $\mathcal{L}$ of L Lipschitz functions from $[0,T]$ to $\mathbb{R}^q$, equipped with the suppremum norm $\|\cdot\|_\infty$, is σ-compact and Hausdorff.*

Proof : For any $\bar{u} \in \mathcal{L}$ and $\bar{v} \in \mathcal{L}$, with $\bar{u} \neq \bar{v}$ we have $\|\bar{u} - \bar{v}\|_\infty > 0$ and therefore $U = \{u \in \mathcal{L}, \|u - \bar{u}\|_\infty < \|\bar{u} - \bar{v}\|_\infty/4\}$ and $V = \{v \in \mathcal{L}, \|v - \bar{v}\|_\infty < \|\bar{u} - \bar{v}\|_\infty/4\}$ form two disjoint neighborhoods of $\bar{u}$ and $\bar{v}$ and we have Hausdorff separation condition. Furthermore, denoting by $\mathcal{B}_\infty(s)$ the ball centered at 0 of radius $s > 0$ in $\mathcal{L}$, we have $\mathcal{L} = \cup_{i\in\mathbb{N}} \mathcal{L} \cap \mathcal{B}_\infty(i)$. Since all the functions of this set are L-Lipschitz and bounded, each element of the union is sequentially compact using Arzelà-Ascoli Theorem [6, Theorem 4.25]. Thus $\mathcal{L}$ is a countable union of compact subspaces, meaning that the space $\mathcal{L}$ is σ-compact. □

We now state and prove a result dealing with the projection of conservative Jacobians.

Lemma 11 *Let $G(x,y) : \mathbb{R}^{p+m} \to \mathbb{R}^n$ be a path-differentiable function whose conservative Jacobian is denoted by $J_G : \mathbb{R}^{p+m} \rightrightarrows \mathbb{R}^{n\times(p+m)}$. Consider*

$$\Pi_y J_G(x,y) := \{M_2 \in \mathbb{R}^{n\times m}, \exists M_1 \in \mathbb{R}^{n\times p}, (M_1, M_2) \in J_G(x,y)\}.$$

Then, for all $x \in \mathbb{R}^p$, $\Pi_y J_G(x,y)$ is conservative for the function $y \mapsto F(x,y)$.

Proof : Consider an absolute continuous function $\gamma : [0,1] \to \mathbb{R}^m$. Then, the function

$$\begin{aligned} \tilde{\gamma} : [0,1] &\to \mathbb{R}^{p+m} \\ t &\mapsto \begin{pmatrix} 0 \\ \gamma(t) \end{pmatrix} \end{aligned}$$

is also an absolute continuous function. Then, for every $x \in \mathbb{R}^p$, it is clear that

$$J_F(\tilde{\gamma}(t))\dot{\tilde{\gamma}}(t) = \Pi_y J_F(x, \gamma(t))\dot{\gamma}(t).$$

From this identity, and by definition of the conservativity, one can show that, for every $x \in \mathbb{R}^p$, $\Pi_y J_F(x, \gamma(t))$ is conservative for $y \mapsto G(x,y)$, which concludes the proof. □

References

[1] Acary, V. and Brogliato, B. (2008) Numerical methods for nonsmooth dynamical systems: applications in mechanics and electronics. Springer Science & Business Media

[2] Aliprantis C.D., Border K.C. (2005) Infinite Dimensional Analysis (3rd edition) Springer

[3] Aubin, J. P., and Cellina, A. (1984). Differential inclusions: set-valued maps and viability theory (Vol. 264). Springer.

[4] Barton, P.I. and Khan, K.A. and Stechlinski, P. and Watson, H. A.J. (2018) Computationnally relevant generalized derivatives: theory, evaluation and applications. Optimization Methods and Software, 33(4-6), 1030–1072.

[5] Benaïm M., Hofbauer J. and Sorin S. (2005). Stochastic approximations and differential inclusions. SIAM Journal on Control and Optimization, 44(1), 328-348.

[6] H. Brezis (2010) Functional analysis, Sobolev spaces and partial differential equations Springer Science & Business Media

[7] Bolte, J. and Pauwels, E. (2020). Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning. Mathematical Programming.

[8] Bolte, J. and Pauwels, E. (2020). A mathematical model for automatic differentiation in machine learning. In Conference on Neural Information Processing Systems.

[9] Bolte, J. and Pauwels, E. and Rios-Zertuche, R. (2020) Long term dynamics of the subgradient method for Lipschitz path differentiable functions arXiv preprint arXiv:2006.00098

[10] Bonnet, B. and Frankowska, H. (2021). Differential inclusions in wasserstein spaces: The cauchy-lipschitz framework. Journal of Differential Equations, 271, 594-637.

[11] Bottou, L. and Curtis, F.E. and Nocedal, J. (2018) Optimization methods for large-scale machine learning. SIAM review, 60(2), 223–311.

[12] Borwein J. M. and Moors, W. B. (1998). A chain rule for essentially smooth Lipschitz functions. SIAM Journal on Optimization, 8(2), 300-308

[13] Borwein J., Moors W. and Wang, X. (2001). Generalized subdifferentials: a Baire categorical approach. Transactions of the American Mathematical Society, 353(10), 3875-3893.

[14] Borwein J. M. (2017). Generalisations, Examples, and Counter-examples in Analysis and Optimisation. Set-Valued and Variational Analysis, 25(3), 467-479.

[15] Cao, Y. and Li, S. and Petzold, L. and Serban, R. (2003) Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution. SIAM journal on scientific computing, 24(3), 1076–1089.

[16] Chen, R. T.Q. and Rubanova Y. and Bettencourt J. and Duvenaud D. (2018) Neural ordinary differential equations In Proceedings of the 32nd International Conference on Neural Information Processing Systems (pp. 6572-6583).

[17] Clarke, F. H. (1975). The Euler-Lagrange differential inclusion. Journal of Differential Equations, 19(1), 80-90.

[18] Clarke F. H. (1983). Optimization and nonsmooth analysis. SIAM.

[19] Clarke, F. H., Ledyaev, Y. S., Stern, R. J. (1998). Asymptotic stability and smooth Lyapunov functions. Journal of differential Equations, 149(1), 69-114.

[20] C.M Dafermos (2005). Hyperbolic conservation laws in continuum physics (3). Springer.

[21] Dontchev, A. and Lempio, F. (1992). Difference methods for differential inclusions: A survey. SIAM review, 34(2), 263-294.

[22] Dupont, E., Doucet, A. and Teh, Y. W. (2019). Augmented neural ODEs. In Proceedings of the 33rd International Conference on Neural Information Processing Systems (pp. 3140-3150).

[23] Evans L. C. and Gariepy R. F. (2015). Measure theory and fine properties of functions. Revised Edition. Chapman and Hall/CRC.

[24] Filippov, A. F. (1988). Differential equations with discontinuous righthand sides: control systems (Vol. 18). Springer Science & Business Media.

[25] Hirsch, M. W., Smale, S. and Devaney, R. L. (2012). Differential equations, dynamical systems, and an introduction to chaos. Academic press.

[26] A. Kechris (2012) Classical descriptive set theory Springer Science & Business Media (156)

[27] Khan, K. A. and Barton, P. I. (2014). Generalized derivatives for solutions of parametric ordinary differential equations with non-differentiable right-hand sides. Journal of Optimization Theory and Applications, 163(2), 355-386.

[28] Lewis, J. M., Lakshmivarahan, S. and Dhall, S. (2006). Dynamic data assimilation: a least squares approach (Vol. 13). Cambridge University Press.

[29] Liberzon, D. (2003). Switching in systems and control. Springer Science & Business Media

[30] Nesterov, Y. (2005). Lexicographic differentiation of nonsmooth functions. Mathematical programming, 104(2), 669-700.

[31] Pang, J. S. and Stewart, D. E. (2009). Solution dependence on initial conditions in differential variational inequalities. Mathematical Programming, 116(1), 429-460.

[32] R-E Plessix (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophysical Journal International, 167(2), 495-503.

[33] Rios-Zertuche, R. (2020). Examples of pathological dynamics of the subgradient method for Lipschitz path-differentiable functions. arXiv preprint arXiv:2007.11699.

[34] H. Royden, P. Fitzpatrick (2010) Real Analysis Prentice Hall

[35] Sahlodin, A.M. and Watson, H. AJ, Barton, P. I. (2016) Nonsmooth model for dynamic simulation of phase changes AIChE Journal, 62(9), 3334–3351.

[36] Stechlinski, P. G. and Barton, P. I. (2016). Generalized derivatives of differential-algebraic equations. Journal of Optimization Theory and Applications, 171(1), 1-26.

[37] Stechlinski, P. G. and Barton, P. I. (2017). Dependence of solutions of non-smooth differential-algebraic equations on parameters. Journal of Differential Equations, 262(3), 2254-2285.

[38] Valadier M. (1989). Entraînement unilatéral, lignes de descente, fonctions lipschitziennes non pathologiques. Comptes rendus de l'Académie des Sciences, 308, 241-244.

[39] Wang X. (1995). Pathological Lipschitz functions in $\mathbb{R}^n$. Master Thesis, Simon Fraser University.