# Higher order Moreau's sweeping process: Mathematical formulation and numerical simulation

Vincent Acary, Bernard Brogliato, Daniel Goeleven

**To cite this version:**

Vincent Acary, Bernard Brogliato, Daniel Goeleven. Higher order Moreau's sweeping process: Mathematical formulation and numerical simulation. Mathematical Programming, Springer Verlag, 2008, 113 (1), pp.133-217. 10.1007/s10107-006-0041-0 . inria-00423566

**HAL Id: inria-00423566**

**https://hal.inria.fr/inria-00423566**

Submitted on 29 Oct 2017

# Higher order Moreau's sweeping process: mathematical formulation and numerical simulation

**Vincent Acary · Bernard Brogliato · Daniel Goeleven**

**Abstract** In this paper we present an extension of Moreau's sweeping process for higher order systems. The dynamical framework is carefully introduced, qualitative, dissipativity, stability, existence, regularity and uniqueness results are given. The time-discretization of these nonsmooth systems with a time-stepping algorithm is also presented. This differential inclusion can be seen as a mathematical formulation of complementarity dynamical systems with arbitrary dimension and arbitrary relative degree between the complementary-slackness variables. Applications of such high-order sweeping processes can be found in dynamic optimization under state constraints and electrical circuits with ideal diodes.

**Keywords** Convex analysis · Measure differential inclusions · Variational inequalities · Complementarity systems · Numerical simulation · Zero dynamics · Relative degree · Schwartz distributions · Electrical circuits · Time-stepping algorithm · Dissipative systems

V. Acary · B. Brogliato
INRIA Rhône-Alpes, Bipop project, ZIRST Montbonnot, 655 avenue de l'Europe,
38334 Saint Ismier cedex, France
e-mail: Bernard.Brogliato@inrialpes.fr

V. Acary
e-mail: Vincent.Acary@inrialpes.fr

D. Goeleven
IREMIA, Université de La Réunion, 97715 Saint-Denis, France
e-mail: goeleven@univ-reunion.fr

## 1 Introduction

The so-called *sweeping process* is a particular differential inclusion of the general form

$$-\dot{z}(t) \in N_{K(t)}(z(t)) \quad (t \geqslant 0), \;\; z(0) \in K(0), \tag{1}$$

where $K(t)$ is a nonempty closed convex time-dependent set, and $N_{K(t)}(z(t))$ is the normal cone to $K(t)$ at $z(t)$. Such evolution problems have been introduced by Moreau in 1971 [45–48]. These models prove to be quite useful in nonsmooth mechanics (elastoplasticity), convex optimization, mathematical economics, queuing theory, etc. Generalizations of the sweeping process have been the object of many studies, see e.g. [5,19,32–34,36,41,67,68,72] and more references in [18,35,42]. Recently it was also shown that quite similar formalisms apply to nonsmooth electrical networks as well as some problems of absolute stability [10,23].

Moreau (see also Schatzman [63,64]) introduced [50,51] an extension of the sweeping process for Lagrangian systems subject to frictionless unilateral constraints. He termed the resulting equations Measure Differential Inclusions. Roughly, such evolution equations are of the form

$$-\mathrm{d}v \in N_{V(q(t))}(v(t)), \quad v(t) \in V(q(t)),$$

where $V(q(t))$ is a closed convex multivalued function depending on $q(\cdot)$, and $v(\cdot)$ is the first derivative of $q(\cdot)$. The term $dv$ denotes the measure that is associated to what one could call the second derivative of $q(\cdot)$. At this stage, we would just like to point out that this evolution problem is of second order, whereas the one in (1) is of first order. This is crucial because it generally implies that the solutions $z(\cdot)$ of the former are absolutely continuous, whereas $v(\cdot)$ in the second formalism is of bounded variation (consequently may possess jumps). In this paper we shall see that the order (which is directly related to some *relative degree* $r$ between two complementary-slackness variables $w$ and $\lambda$) indeed is a fundamental parameter which determines the nature of the solutions and their degree as distributions.

Most importantly, numerical time integration schemes have been introduced for both first and second order sweeping processes in [48,51] which correspond to what is now called *time-stepping* schemes [15]. The term "time-stepping" is used to underline that there is no explicit procedure to take into account events, such as activation of constraints. For the first order case (1), the algorithm $-z(t_{k+1}) + z(t_k) \in N_{K(t_{k+1})}(z(t_{k+1}))$ is called the *catching up* algorithm because of its geometrical interpretation. It has been extensively used both for well-posedness studies [35,40,42] and for numerical simulation [15,52]. For Lagrangian systems subject to unilateral constraints, the time integration method is built in the same way and called the Non Smooth Contact Dynamics method [30,51,53]. Convergence and well-posedness studies has been also given for this second order case [35,40,42,70]. The algorithm that is proposed in this paper, is inspired from these time integration schemes as will be explained in Sect. 5. One motivation of the proposed study, is to provide us with an efficient numerical scheme for systems with unilateral constraints and with relative degree larger than 3. It is shown in [16] and recalled in the Sect. 5.5.1 that applying directly a backward Euler scheme leads to some inconsistencies due to the fact that distributional solutions are not correctly approximated. We aim at bridging this gap.

Recently, dynamical complementarity systems (introduced by Moreau in 1963 in the framework of Lagrangian mechanical systems [43,44], see also [39]) have been the object of interest in the control literature [9,11,14,16,17,28,38, 57,71,73–75], because of many applications in various fields such as nonsmooth electrical networks, optimal control with state and/or input constraint, Lagrangian mechanical systems subject to unilateral constraints, etc [9,73]. It is well-known that complementarity problems, variational inequalities and inclusions (or generalized equations) are closely related [22]. Similarly, complementarity systems, evolution variational inequalities and differential inclusions as the above ones (which we could name *unbounded DI*) are related [9,10,23,24]. Equivalences between various unilateral dynamics formalisms are established in [12]. For instance it is easy to see that the first order sweeping process can also be rewritten as the evolution variational inequality $\langle \dot{z}(t), v - z(t) \rangle \geqslant 0$, for all $v \in K(t)$, $z(0) \in K(0)$, $z(t) \in K(t)$ for all $t \geqslant 0$. Such relationships between various formalisms will be important for the developments in this paper, especially for the design of a numerical algorithm. In parallel results have been obtained for classes of piecewise affine systems [29], however the link with complementarity is not yet clear except in some particular cases [27, Sect. 4.2.2]. The time-discretization of linear complementarity systems with implicit Euler algorithms has been considered in [16], as indicated above.

In [55] the so-called differential variational inequalities with index $\leqslant 2$ (i.e. with relative degree $\leqslant 1$ in the language of this paper) are studied in a rather detailed manner: well-posedness results are given, and numerical schemes are proposed and shown to converge.

**Objective of the paper** The starting point is to consider a dynamical system of the form $\dot{x}(t) = Ax(t) + B\lambda(t)$ whose trajectories have to evolve in a domain of the form $Cx(t) \geqslant 0$. The dynamics is embedded into a distributional differential inclusion (named the higher order sweeping process) which allows us to integrate the system while respecting the unilateral constraint on the state. In the proposed formalism the Lagrange multiplier $\lambda$ is a distribution which possesses a specific decomposition into measures which satisfy inclusions into a family of convex cones. A time-stepping numerical algorithm is constructed and its properties are analyzed. Though the convergence towards solutions of the continuous-time system is not yet complete, the analysis shows however that the numerical scheme possesses strong properties.

The paper is organized as follows. Some mathematical tools are presented in Sect. 2. A special canonical state space representation which is useful for the subsequent developments is introduced in Sect. 3. The corresponding differential inclusion formalism is introduced and motivated in Sect. 4, where important properties and well-posedness are shown. Section 5 is devoted to the design and the analysis of a time-stepping numerical scheme. In Sect. 6 an application of the extended sweeping process is presented. Conclusions close the paper in Sect. 7.

**Notation** The following notation is used: $\mathbb{R}$ is the real line, $\mathbb{R}^+$ is the nonnegative real line, $x_i$ is the $i$th component of a vector $x \in \mathbb{R}^n$. The relative degree between two signals $w$ and $\lambda$ defining the output and the input of a system, is denoted as $r$. The indicator function of a set $K$ is denoted as $\psi_K(\cdot)$, and $\partial$ is the convex analysis subdifferential operator. The closed convex hull of a set $K$ is denoted by $\overline{\text{co}}(K)$. When $K$ is a nonempty closed convex set then the normal cone to $K$ is denoted as $N_K(\cdot) = \partial\psi_K(\cdot)$. Lexicographical inequalities are denoted as $\succeq$. For a vector $x$, $x \succeq 0$ means that all entries $x_i = 0$ or the first nonzero entry is positive. $x \geqslant 0$ means that all entries $x_i \geqslant 0$. $I_n$ denotes the $n \times n$ identity matrix, $0^n$ denotes the vector $(0, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^n$ and $0_n = (0^n)^{\mathrm{T}}$. Let $M \in \mathbb{R}^{n \times n}$ be a symmetric and positive definite matrix, the proximation operator $prox_M[K; .]$ is defined by $prox_M[K; x] = \text{argmin}_{z \in K}(z - x)^{\mathrm{T}} M(z - x)$. If $M = I_n$, we set $prox[K; .] := prox_{I_n}[K; .]$. The notation $\langle x, y\rangle := x^{\mathrm{T}}y$ and $\|x\| := \sqrt{x^{\mathrm{T}}x}$ will also be used. For a matrix $A \in \mathbb{R}^{n \times n}$, we set $|||A||| := \sup_{x \neq 0}\{\frac{\|Ax\|}{\|x\|}\}$. $A_{ij}$ denotes the entry of the $i$th row and $j$th column, $A_i$ denotes the $i$th row of the matrix $A$.

**Acronyms** The following acronyms are used throughout the paper: Complementarity Problem (CP), Linear Complementarity Problem (LCP), Variational Inequality (VI), Ordinary Differential Equation (ODE), Differential Algebraic Equation (DAE), Measure Differential Inclusion (MDI), Complementarity System (CS), Linear Complementarity System (LCS).

## 2 Mathematical tools

In this section we present fundamental analysis tools which will be helpful in settling the higher order sweeping process formalism. In particular a class of distributions that is a potential class of solutions of the dynamics is presented in detail.

**Radon measure** Let us denote by $\mathcal{B}(\mathbb{R})$ the Borel $\sigma$ algebra and let $\mathrm{d}R$ be a $\mathbb{R}^n$-valued Radon measure, i.e. a Borel regular measure such that $\mathrm{d}R_i(K) < +\infty$ $(1 \leqslant i \leqslant n)$ for every compact set $K \subset \mathbb{R}$ (see e.g. [21]).

Let $A \in \mathcal{B}(\mathbb{R})$ be given, we say that $\{A_i\}_{i=1}^m$ is a finite partition of $A$ if $A_i \in \mathcal{B}(\mathbb{R})$, $A_i \cap A_j = \emptyset, i \neq j$ $(1 \leqslant i, j \leqslant n)$ and $\cup_{i=1}^m A_i = A$. Let us now denote by $\mathcal{P}(A)$ the set of finite partitions of $A$. The modulus measure of $dR$ is defined by (see e.g. [65]):

$$|\mathrm{d}R|(A) = \sup_{\{A_i\}_{i=1}^m \in \mathcal{P}(A)} \sum_{i=1}^m \|dR(A_i)\|, \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

Let $\mathrm{d}\mu$ be a real-valued Radon measure. Let us denote by $L^1_{\mathrm{loc}}(\mathbb{R}, \mathrm{d}\mu; \mathbb{R}^n)$ the space of $\mathrm{d}\mu$-locally integrable $\mathbb{R}^n$-valued functions. One says that $\mathrm{d}R$ has a density relative to $\mathrm{d}\mu$ provided that there exists a (class of) function $R'_\mu \in L^1_{\mathrm{loc}}(\mathbb{R}, \mathrm{d}\mu; \mathbb{R}^n)$ such that $\mathrm{d}R = R'_\mu \mathrm{d}\mu$, i.e.

$$\mathrm{d}R(A) = \int_A R'_\mu \mathrm{d}\mu, \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

If $\mathrm{d}\mu$ is nonnegative and $\mathrm{d}R$ is absolutely continuous with respect to $\mathrm{d}\mu$, i.e. $A \in \mathcal{B}(\mathbb{R}), \mathrm{d}\mu(A) = 0 \Rightarrow \mathrm{d}R(A) = 0$, then a classical direct consequence of the Lebesgue–Radon–Nikodym Theorem (see Theorem 5.10.22 in [65]) ensures the existence of a unique (class of) function $R'_\mu \in L^1_{\mathrm{loc}}(\mathbb{R}, \mathrm{d}\mu; \mathbb{R}^n)$ such that $\mathrm{d}R = R'_\mu \mathrm{d}\mu$. Here

$$R'_\mu(t) = \frac{\mathrm{d}R}{\mathrm{d}\mu}(t),$$

where $\mathrm{d}R/\mathrm{d}\mu$ denotes the derivative (density) of $\mathrm{d}R$ with respect to $\mathrm{d}\mu$. In particular, since $|\mathrm{d}R|$ is nonnegative and $\|\mathrm{d}R(A)\| \leqslant |\mathrm{d}R|(A), \forall A \in \mathcal{B}(\mathbb{R})$, there exists a unique (class of) function $\theta_R \in L^1_{\mathrm{loc}}(\mathbb{R}, |\mathrm{d}R|; \mathbb{R}^n)$ such that $\mathrm{d}R = \theta_R |\mathrm{d}R|$.

**Differential measure process** Let $I$ be a real non-degenerate interval (not empty nor reduced to a singleton), and let $\{K(t); t \in I\} \subset \mathbb{R}^n$ be a family of non-empty closed convex cones. Suppose that $\mathrm{d}\mu$ is nonnegative and $\mathrm{d}R$ is absolutely continuous with respect to $\mathrm{d}\mu$. By convention, we shall write

$$\mathrm{d}R = R'_\mu \mathrm{d}\mu \in K(t) \quad \text{on } I \tag{2}$$

to mean that

$$R'_\mu(t) \in K(t), \quad \mathrm{d}\mu - a.e. \;\; t \in I. \tag{3}$$

**Proposition 1** *If the relation in* (3) *holds then, for every nonempty bounded set* $A \in \mathcal{B}(\mathbb{R})$, $A \subset I$, *we have:*

$$\mathrm{d}R(A) \in \overline{\mathrm{co}}(\cup_{\tau \in A} K(\tau)). \tag{4}$$

*Proof* Suppose that (3) is satisfied. If $A \in \mathcal{B}(\mathbb{R})$, $A \subset I$ and $\mathrm{d}\mu(A) = 0$ then $\mathrm{d}R(A) = 0$ and it is clear that (4) holds since $0 \in \overline{\mathrm{co}}(\cup_{\tau \in A} K(\tau))$. Let $A \in \mathcal{B}(\mathbb{R})$, $A \subset I$ such that $0 < \mathrm{d}\mu(A) < +\infty$. Then $R'_\mu(A) \subset \overline{\mathrm{co}}(\cup_{\tau \in A} K(\tau))$ and thus (see Theorem 5.7.35 in [65]):

$$\frac{1}{\mathrm{d}\mu(A)} \int_A R'_\mu \mathrm{d}\mu \in \overline{\mathrm{co}}(\cup_{\tau \in A} K(\tau))$$

since $\overline{\mathrm{co}}(\cup_{\tau\in A}K(\tau))$ is closed and convex. It results that $\mathrm{d}R(A) = \int_A R'_\mu \mathrm{d}\mu \in \overline{\mathrm{co}}(\cup_{\tau\in A}K(\tau))$ since $\mathrm{d}\mu(A) > 0$ and $\overline{\mathrm{co}}(\cup_{\tau\in A}K(\tau))$ is a closed convex cone. □

The following result is also important for the study of the differential measure process (see e.g. [42]).

**Proposition 2** *Let*

$$\mathrm{d}R = R'_t \mathrm{d}t + \mathrm{d}\nu$$

*be the Lebesgue–Radon–Nikodym decomposition of* d$R$ *with respect to the Lebesgue measure* d$t$*. Then the relation in* (3) *holds if and only if*

$$R'_t(t) \in K(t), \quad \mathrm{d}t - \text{a.e. } t \in I \tag{5}$$

*and*

$$\frac{\mathrm{d}\nu}{|\mathrm{d}\nu|}(t) \in K(t), \quad |\mathrm{d}\nu| - a.e. \ t \in I. \tag{6}$$

**Functions of bounded variation** Let $I$ denote a non-degenerate real interval (not empty nor reduced to a singleton).

- By $u \in \mathbf{BV}(\mathbf{I}; \mathbb{R}^{\mathbf{n}})$ it is meant that $u$ is a $\mathbb{R}^n$-valued function of bounded variation if there exists a constant $C > 0$ such that for all finite sequences $t_0 < t_1 < \cdots < t_N$ ($N$ arbitrary) of points of $I$, we have

$$\sum_{i=1}^{N} \|u(t_i) - u(t_{i-1})\| \leqslant C.$$

  Let $J$ be a subinterval of $I$. The real number

$$\mathrm{var}(u, J) := \sup \sum_{i=1}^{N} \|u(t_i) - u(t_{i-1})\|,$$

  where the supremum is taken with respect to all the finite sequences $t_0 < t_1 < \cdots < t_N$ ($N$ arbitrary) of points of $J$, is called the variation of $u$ in $J$.

Any BV function has a countable set of discontinuity points and is almost everywhere differentiable. A BV function defined on $[a, b] \subset I$ possesses left-limits in $]a, b]$ and right-limits in $[a, b[$. Moreover, the functions $t \mapsto u(t^+) := \lim_{s\to t, s>t} u(s)$ and $t \mapsto u(t^-) := \lim_{s\to t, s<t} u(s)$ are both BV functions.

Let us also recall here a classical form of the Gronwall–Bellman Lemma (see e.g. [7]).

**Lemma 1** *Let* $0 < T < +\infty$ *be given. Let* $c_1 \geqslant 0, c_2 \geqslant 0$ *and* $u \in \mathbf{BV}([\mathbf{0},\mathbf{T}];\mathbb{R})$ *such that:*

$$0 \leqslant u(t) \leqslant c_1 + c_2 \int_0^t u(s)\mathrm{d}s, \quad \forall\, t \in [0,T].$$

*Then*

$$u(t) \leqslant c_1 e^{c_2 t}, \quad \forall\, t \in [0,T].$$

- We denote by $\mathbf{LBV}(\mathbf{I};\mathbb{R}^{\mathbf{n}})$ the space of functions of locally bounded variation, i.e. of bounded variation on every compact subinterval of $I$.
- We denote by $\mathbf{RCLBV}(\mathbf{I};\mathbb{R}^{\mathbf{n}})$ the space of right-continuous functions of locally bounded variation. It is known that if $u \in \mathbf{RCLBV}(\mathbf{I};\mathbb{R}^{\mathbf{n}})$ and $[a,b]$ denotes a compact subinterval of $I$, then $u$ can be represented in the form (see e.g. [66]):

  $$u(t) = \mathcal{J}_u(t) + [u](t) + \zeta_u(t), \quad \forall\, t \in [a,b],$$

  where $\mathcal{J}_u$ is a jump function, $[u]$ is an absolutely continuous function and $\zeta_u$ is a singular function. Here $\mathcal{J}_u$ is a jump function in the sense that $\mathcal{J}_u$ is right-continuous and given any $\varepsilon > 0$, there exist finitely many points of discontinuity $t_1, \ldots, t_N$ of $\mathcal{J}_u$ such that $\sum_{i=1}^N \|\mathcal{J}_u(t_i) - \mathcal{J}_u(t_i^-)\| + \varepsilon > \operatorname{var}(\mathcal{J}_u, [a,b])$, $[u]$ is an absolutely continuous function in the sense that for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\sum_{i=1}^N \|[u](\beta_i) - [u](\alpha_i)\| < \varepsilon$, for any collection of disjoint subintervals $]\alpha_i, \beta_i] \subset [a,b] (1 \leqslant i \leqslant N)$ such that $\sum_{i=1}^N (\beta_i - \alpha_i) < \delta$, and $\zeta_u$ is a singular function in the sense that $\zeta_u$ is a continuous and bounded variation function on $[a,b]$ such that $\dot{\zeta}_u = 0$ almost everywhere on $[a,b]$.
- By $u \in \mathbf{RCSLBV}(\mathbf{I};\mathbb{R}^{\mathbf{n}})$ it is meant that $u$ is a right-continuous function of special locally bounded variation, i.e. $u$ is of bounded variation and can be written as the sum of a jump function and an absolutely continuous function on every compact subinterval of $I$. So, if $u \in \mathbf{RCSLBV}(\mathbf{I};\mathbb{R}^{\mathbf{n}})$ then

  $$u = [u] + \mathcal{J}_u \tag{7}$$

  where $[u]$ is a locally absolutely continuous function called the absolutely continuous component of $u$ and $\mathcal{J}_u$ is uniquely defined up to a constant by

  $$\mathcal{J}_u(t) = \sum_{t \geqslant t_n} u(t_n^+) - u(t_n^-) = \sum_{t \geqslant t_n} u(t_n) - u(t_n^-) \tag{8}$$

  where $t_1, t_2, \ldots, t_n, \ldots$ denote the countably many points of discontinuity of $u$ in $I$.

**Stieltjes measure** Let $u \in \mathbf{LBV}(\mathbf{I};\mathbb{R}^{\mathbf{n}})$ be given. We denote by d$u$ the Stieltjes measure generated by $u$ (see e.g. [65] and [42]). Recall that for $a \leqslant b$, $a, b \in I$:

$$\begin{aligned} \mathrm{d}u([a,b]) &= u(b^+) - u(a^-), \\ \mathrm{d}u([a,b[) &= u(b^-) - u(a^-), \\ \mathrm{d}u(]a,b]) &= u(b^+) - u(a^+), \\ \mathrm{d}u(]a,b[) &= u(b^-) - u(a^+). \end{aligned}$$

In particular, we have

$$\mathrm{d}u(\{a\}) = u(a^+) - u(a^-).$$

For $u \in \mathbf{LBV}(\mathbf{I};\mathbb{R}^{\mathbf{n}})$, $u^+$ and $u^-$ denote the functions defined by

$$u^+(t) = u(t^+) = \lim_{s \to t, s > t} u(s), \quad \forall\, t \in I,\ \ t < \sup\{I\},$$

and

$$u^-(t) = u(t^-) = \lim_{s \to t, s < t} u(s), \quad \forall\, t \in I,\ \ t > \inf\{I\}.$$

(where $\sup\{I\}$ (resp. $\inf\{I\}$) denotes the supremum (resp. infimum) of the set $I$). If $u, v \in \mathbf{LBV}(\mathbf{I};\mathbb{R}^{\mathbf{n}})$ then $u^T v \in \mathbf{LBV}(\mathbf{I};\mathbb{R})$ and

$$\mathrm{d}(u^{\mathrm{T}} v) = (v^-)^{\mathrm{T}} \mathrm{d}u + (u^+)^{\mathrm{T}} \mathrm{d}v = (v^+)^{\mathrm{T}} \mathrm{d}u + (u^-)^{\mathrm{T}} \mathrm{d}v. \qquad (9)$$

Let us also recall that

$$2(u^-)^{\mathrm{T}} \mathrm{d}u \leqslant \mathrm{d}(u^{\mathrm{T}} u) = (u^+ + u^-)^{\mathrm{T}} \mathrm{d}u \leqslant 2(u^+)^{\mathrm{T}} \mathrm{d}u. \qquad (10)$$

Finally, we recall that if $J$ denotes a bounded subinterval of $I$, then:

$$|\mathrm{d}u|(J) = \mathrm{var}(u, J). \qquad (11)$$

**Measure differential inclusions** The material given in this section can be used to formulate measure differential inclusions. For example, let $F : I \times \mathbb{R}^n \to \mathbb{R}^n$ be a locally $L^1$-Caratheodory function and let $C \subset \mathbb{R}^n$ be a nonempty closed convex set. We consider the measure differential inclusion: Find $u \in \mathbf{RCLBV}(\mathbf{I};\mathbb{R}^{\mathbf{n}})$ such that:

$$\mathrm{d}u + F(t, u(t))\mathrm{d}t \in -N_C(u(t)). \qquad (12)$$

The sense of (12) is given by the existence of a nonnegative Radon measure d$\mu$ such that the measures d$u$ and d$t$ are absolutely continuous with respect to d$\mu$ and

$$\frac{\mathrm{d}u}{\mathrm{d}\mu}(t) + F(t, u(t)) \frac{\mathrm{d}t}{\mathrm{d}\mu}(t) \in -N_C(u(t)), \quad \mathrm{d}\mu - a.e.\ \ t \in I. \qquad (13)$$

Note that the concept of solution does not depend of the choice of the nonnegative Radon measure $\mathrm{d}\mu$ since the right-hand side of (13) is a cone, and the densities can be obtained one from other by multiplication with a nonnegative function. It results from Proposition 1 that

$$\mathrm{d}u(A) + \int_A F(\tau, u(\tau))\mathrm{d}\tau \in \overline{\mathrm{co}}\left(\bigcup_{\tau\in A} -N_C(u(\tau))\right), \tag{14}$$

for every nonempty $A \in \mathcal{B}(\mathbb{R})$, $A \subset I$, such that $\mathrm{d}\mu(A) < +\infty$.

Note that from Proposition 2, we get also:

$$u(t^+) - u(t^-) \in -N_C(u(t)), \quad \forall\, t \in I \tag{15}$$

and

$$u_t'(t) + F(t, u(t)) \in -N_C(u(t)), \quad a.e.\ \ t \in I. \tag{16}$$

According to the notation introduced above, $u_t'$ denotes the density of $\mathrm{d}u$ with respect to the Lebesgue measure $\mathrm{d}t$.

**Distributions generated by RCSLBV functions** Let $I$ be the real interval given by

$$I = [\alpha, \beta[,$$

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R} \cup \{+\infty\}$.

The support $\mathrm{supp}\{\varphi\}$ of a function $\varphi : I \to \mathbb{R}$ is defined by $\mathrm{supp}\{\varphi\} := \overline{\{t \in I : \varphi(t) \neq 0\}}$. We denote by $C_0^\infty(I)$ the space of real-valued $C^\infty(I)$-mappings with compact support contained in the open interval $]\alpha, \beta[$ and $\mathcal{D}'(I)$ is the space of Schwartz distributions on $I$, i.e. the space of linear continuous forms on $C_0^\infty(I)$. Recall that for $T \in \mathcal{D}'(I)$, the (generalized) derivative of $T$ is defined by

$$\langle \mathrm{D}T, \varphi\rangle = -\langle T, \dot{\varphi}\rangle, \quad \forall\, \varphi \in C_0^\infty(I).$$

The (generalized) derivative of order $n$ is then given by

$$\mathrm{D}^n T = \mathrm{D}(\mathrm{D}^{n-1} T)\ (n \geqslant 2),$$

that is

$$\langle \mathrm{D}^n T, \varphi\rangle = (-1)^n \langle T, \varphi^{(n)}\rangle, \quad \forall\, \varphi \in C_0^\infty(I).$$

For $a \in I$, we denote by $\delta_a$ the Dirac distribution at $a$, defined by

$$\langle \delta_a, \varphi\rangle = \varphi(a), \quad \forall\, \varphi \in C_0^\infty(I).$$

Note that $\delta_a = D\mathcal{H}(. - a)$ where $\mathcal{H}$ is the Heaviside function:

$$\mathcal{H}(t) = \begin{cases} 1 & \text{if } t \geqslant 0 \\ 0 & \text{if } t < 0. \end{cases} \tag{17}$$

The support $\text{supp}\{T\}$ of a distribution $T \in \mathcal{D}'(I)$ is defined by $\text{supp}\{T\} := I \backslash \mathcal{O}$ where $\mathcal{O} \subset I$ denotes the largest open set in $I$ on which $T$ vanishes in the sense that $\langle T, \varphi \rangle = 0, \forall\, \varphi \in C_0^\infty(I)$ with support contained in $\mathcal{O}$.

- Let $h \in \mathbf{RCSLBV}(\mathbf{I};\mathbb{R})$ be given. We will denote by $E_0(h)$ the countable set of points of discontinuity $t_1, t_2, \ldots, t_k, \ldots$ of $h$. As seen above, $h$ can be written as the sum of a locally absolutely continuous function $[h]$ and the locally jump function $\mathcal{J}_h$ given by

  $$\mathcal{J}_h(t) = \sum_{t \geqslant t_k} \sigma_h(t_k),$$

  where for $t \in I$,

  $$\sigma_h(t) := h(t^+) - h(t^-) = h(t) - h(t^-)$$

  denotes the jump of $h$ at $t$. It is clear that if $t \in I \backslash E_0(h)$ then $\sigma_h(t) = 0$.

We will denote by $\hat{h}^{(1)}(t)$ the right derivative (if it exists) of the absolutely continuous part $[h]$ of $h \in \mathbf{RCSLBV}(\mathbf{I};\mathbb{R})$ at $t$, i.e.

$$\hat{h}^{(1)}(t) := \frac{\mathrm{d}^+[h]}{\mathrm{d}t}(t) = \lim_{\sigma \to 0^+} \frac{[h](t+\sigma) - [h](t)}{\sigma}.$$

We have thus:

$$h = [h] + \mathcal{J}_h \tag{18}$$

and

$$\mathrm{d}h = \hat{h}^{(1)}\mathrm{d}t + \mathrm{d}\mathcal{J}_h. \tag{19}$$

The measure $\mathrm{d}\mathcal{J}_h$ is atomic as a measure concentrated on the set $E_0(h)$ of countably many points of discontinuity of $h$ in $I$, i.e. $\mathrm{d}\mathcal{J}_h(A) = 0, \forall\, A \in \mathcal{B}(\mathbb{R}), A \subset I \backslash E_0(h)$.

Let us now set

$$\begin{cases} \mathcal{F}_0(I;\mathbb{R}) = RCSLBV(I;\mathbb{R}) \\ \mathcal{F}_1(I;\mathbb{R}) = \{h \in \mathcal{F}_0(I;\mathbb{R}) : \hat{h}^{(1)} \in RCSLBV(I;\mathbb{R})\} \\ \mathcal{F}_2(I;\mathbb{R}) = \{h \in \mathcal{F}_1(I;\mathbb{R}) : \hat{h}^{(2)} := \frac{\mathrm{d}^+}{\mathrm{d}t}[\hat{h}^{(1)}] \in RCSLBV(I;\mathbb{R})\} \\ \cdots\cdots \\ \mathcal{F}_k(I;\mathbb{R}) = \{h \in \mathcal{F}_{k-1}(I;\mathbb{R}) : \hat{h}^{(k)} := \frac{\mathrm{d}^+}{\mathrm{d}t}[\hat{h}^{(k-1)}] \in RCSLBV(I;\mathbb{R})\} \end{cases}$$

and

$$\mathcal{F}_\infty(I;\mathbb{R}) = \bigcap_{k\in\mathbb{N}} \mathcal{F}_k(I;\mathbb{R}).$$

We standardize the notation by setting $\hat{h}^{(0)} := h$. Note that $\hat{h}^{(\alpha)} \in \mathbf{RCSLBV(I;}\mathbb{R})$ $(\alpha \geqslant 1)$ means that the absolutely continuous function $[\hat{h}^{(\alpha-1)}]$ admits a right derivative $\hat{h}^{(\alpha)}(t) = \frac{\mathrm{d}^+}{\mathrm{d}t}[\hat{h}^{(\alpha-1)}](t)$ at each $t \in I$ and $\hat{h}^{(\alpha)}$ is of special local bounded variation over $I$.

*Remark 1*
i) Note that if $\frac{\mathrm{d}^+}{\mathrm{d}t}[\hat{h}^{(\alpha-1)}]$ exists and is of special local bounded variation on $I$ then necessarily $\frac{\mathrm{d}^+}{\mathrm{d}t}[\hat{h}^{(\alpha-1)}]$ is right-continuous. Properties $\hat{h}^{(\alpha)} \in \mathbf{RCLBV(I;}\mathbb{R})$ and $\hat{h}^{(\alpha)} \in \mathbf{LBV(I;}\mathbb{R})$ are thus equivalent. The "RC" requirement stands in the definition of the vector space $\mathcal{F}_\alpha(I;\mathbb{R})$ for pedagogical reasons.

ii) Note also that for *a.e.t* $\in I$, we have $\frac{\mathrm{d}^+}{\mathrm{d}t}[\hat{h}^{(\alpha-1)}](t) = \frac{\mathrm{d}^+}{\mathrm{d}t}\hat{h}^{(\alpha-1)}(t) = h_t^{(\alpha-1)}(t)$, where, according to the notation introduced above, $h_t^{(\alpha-1)}$ denotes the density of $\mathrm{d}h^{(\alpha-1)}$ with respect to the Lebesgue measure $\mathrm{d}t$.

*Example 1* Set $I = [0, +\infty[$ and let $u : I \to \mathbb{R}$ be the function given by

$$u(t) = |\sin(t)|, \quad \forall\, t \geqslant 0.$$

It is clear that $\hat{u}^{(0)} := u \in \mathbf{RCSLBV(I;}\mathbb{R})$ since $u$ is Lipschitz-continuous. Then we obtain

$$\hat{u}^{(1)}(t) := \frac{\mathrm{d}^+}{\mathrm{d}t}[\hat{u}^{(0)}](t) = \frac{\mathrm{d}^+u}{\mathrm{d}t}(t) = \cos(t - k\pi) \quad \text{if } t \in [k\pi, (k+1)\pi[, \;\; k \in \mathbb{N}.$$

We see that $E_0(\hat{u}^{(1)}) = \{k\pi;\ k \in \mathbb{N}\backslash\{0\}\}$ and

$$\hat{u}^{(1)} = [\hat{u}^{(1)}] + \mathcal{J},$$

where

$$[\hat{u}^{(1)}](t) = -2k + \cos(t - k\pi) \quad \text{if } t \in [k\pi, (k+1)\pi[, \;\; k \in \mathbb{N},$$

and

$$\mathcal{J}(t) = 2k \quad \text{if } t \in [k\pi, (k+1)\pi[, \;\; k \in \mathbb{N}.$$

Thus $\hat{u}^{(1)} \in \mathbf{RCSLBV(I;}\mathbb{R})$. Then

$$\hat{u}^{(2)}(t) := \frac{\mathrm{d}^+}{\mathrm{d}t}[\hat{u}^{(1)}](t) = -|\sin(t)|$$

so that $\hat{u}^{(2)} \in \mathbf{RCSLBV}(\mathbf{I};\mathbb{R})$. And so on, we see that

$$\hat{u}^{(k)}(t) = \begin{cases} (-1)^m \hat{u}^{(0)}(t) & \text{if } \ k = 2m \\ \\ (-1)^m \hat{u}^{(1)}(t) & \text{if } \ k = 2m+1 \end{cases}, \quad m \in \mathbb{N},$$

so that $\hat{u}^{(k)} \in \mathbf{RCSLBV}(\mathbf{I};\mathbb{R}),\ \forall k \in \mathbb{N}$, and thus $u \in \mathcal{F}_\infty(I;\mathbb{R})$.

*Example 2* Let $\alpha < 0$ be given and set $I = [\alpha, +\infty[$. One says that $u : I \to \mathbb{R}$ is a Bohl function (see e.g. [17,28]) if there exist $N \in \mathbb{N}$ and matrices $H \in \mathbb{R}^{1\times N}, U \in \mathbb{R}^{N\times N}$ and $G \in \mathbb{R}^{N\times 1}$ such that:

$$u(t) = \begin{cases} 0 & \text{if} \quad \alpha \leqslant t < 0 \\ \\ He^{Ut}G & \text{if} \quad t \geqslant 0. \end{cases}$$

Then, for all $k \in \mathbb{N}$, we get

$$\hat{u}^{(k)}(t) = \begin{cases} 0 & \text{if} \quad \alpha \leqslant t < 0 \\ \\ HU^k e^{Ut}G & \text{if} \quad t \geqslant 0 \end{cases}$$

and it is clear that $u \in \mathcal{F}_\infty(I;\mathbb{R})$.

Let $h \in \mathcal{F}_\infty(I;\mathbb{R})$ be given. One remarks that (generalized) derivatives of $h$ are easy to handle. Indeed, let us here denote by $T_h$ the regular distribution generated by $h$, i.e.

$$\langle T_h, \varphi\rangle = \int_I \varphi h \, \mathrm{d}t, \quad \forall\, \varphi \in C_0^\infty(I).$$

Let $\varphi \in C_0^\infty(I)$ be given. Using (9) and (19), we obtain (here $h^+ = h$ and $\varphi^- = \varphi$):

$$\langle \mathrm{D}T_h, \varphi\rangle = -\int_I \dot{\varphi} h \, \mathrm{d}t = -\int_I h \, \mathrm{d}\varphi = -\int_I \mathrm{d}(h\varphi) + \int_I \varphi \, \mathrm{d}h$$

$$= \int_I \varphi \, \mathrm{d}h = \int_I \varphi \, \hat{h}^{(1)} \mathrm{d}t + \sum_{t_k \in E_0(h) \cap \operatorname{supp}\{\varphi\}} (h(t_k^+) - h(t_k^-)) \langle \delta_{t_k}, \varphi\rangle$$

and so on, for any $\alpha \geqslant 2$, we have:

$$\begin{aligned}
&\langle \mathrm{D}^\alpha T_h, \varphi\rangle = \\
&\quad = \int\limits_I \varphi \, \mathrm{d}\hat{h}^{(\alpha-1)} + \sum_{i=2}^{\alpha} \left( \sum_{t_k \in E_0(\hat{h}^{(\alpha-i)}) \cap \operatorname{supp}\{\varphi\}} (\hat{h}^{(\alpha-i)}(t_k^+) - \hat{h}^{(\alpha-i)}(t_k^-)) \langle \delta_{t_k}^{(i-1)}, \varphi \rangle \right) \\
&\quad = \int\limits_I \varphi \hat{h}^{(\alpha)} \mathrm{d}t + \sum_{i=1}^{\alpha} \left( \sum_{t_k \in E_0(\hat{h}^{(\alpha-i)}) \cap \operatorname{supp}\{\varphi\}} (\hat{h}^{(\alpha-i)}(t_k^+) - \hat{h}^{(\alpha-i)}(t_k^-)) \langle \delta_{t_k}^{(i-1)}, \varphi \rangle \right).
\end{aligned}$$

*Example 3* Let $u : I \to \mathbb{R}$ be the function considered in Example 1. Let us now consider the distribution $T$ defined by

$$\langle T, \varphi \rangle = \int\limits_I |\sin(t)| \varphi(t) \mathrm{d}t, \quad \forall\, \varphi \in C_0^\infty(I).$$

Then for a given function $\varphi \in C_0^\infty(I)$, we see that:

$$\langle \mathrm{D}T, \varphi \rangle = \int\limits_I \hat{u}^{(1)}(t)\varphi(t)\mathrm{d}t = \sum_{k \in \mathbb{N} \cap \operatorname{supp}\{\varphi\}} \int\limits_{k\pi}^{(k+1)\pi} \cos(t - k\pi)\varphi(t)\mathrm{d}t,$$

and

$$\begin{aligned}
\langle \mathrm{D}^2 T, \varphi \rangle &= \int\limits_I \hat{u}^{(2)}(t)\varphi(t)\mathrm{d}t + \sum_{k \in \mathbb{N}\setminus\{0\} \cap \operatorname{supp}\{\varphi\}} (\hat{u}^{(1)}(k\pi^+) - \hat{u}^{(1)}(k\pi^-)) \langle \delta_{k\pi}, \varphi \rangle \\
&= -\int\limits_I |\sin(t)|\varphi(t)\mathrm{d}t + 2 \sum_{k \in \mathbb{N}\setminus\{0\} \cap \operatorname{supp}\{\varphi\}} \langle \delta_{k\pi}, \varphi \rangle,
\end{aligned}$$

and so on.

- Let $n \in \mathbb{N}$ be given.

**Definition 1** *We say that a Schwartz distribution $T \in \mathcal{D}'(I)$ is of class $\mathcal{T}_n$ on $I$ provided that there exists a function $F \in \mathcal{F}_\infty(I; \mathbb{R})$ such that $T = D^n F$.*

Let us now denote by $\mathcal{T}_n(I)$ the set of all distributions of class $\mathcal{T}_n$ on $I$, i.e.

$$\mathcal{T}_n(I) = \{T \in \mathcal{D}'(I) : \exists F \in \mathcal{F}_\infty(I; \mathbb{R}) \ \text{ such that } \ T = D^n F\ \}.$$

It is clear that

$$\mathcal{T}_0(I) = \mathcal{F}_\infty(I; \mathbb{R}).$$

If $T \in \mathcal{T}_1(I)$ then there exists $F \in \mathcal{F}_\infty(I;\mathbb{R})$ such that

$$\langle T,\varphi\rangle = \int_I \varphi \mathrm{d}\hat{F}^{(0)} = \int_I \varphi \mathrm{d}F, \quad \forall\, \varphi \in C_0^\infty(I).$$

More generally, if $T \in \mathcal{T}_n(I)$ for some $n \geqslant 2$, then there exists $F \in \mathcal{F}_\infty(I;\mathbb{R})$ such that, for all $\varphi \in C_0^\infty(I)$:

$$\begin{aligned}\langle T,\varphi\rangle &= \\ &= \int_I \varphi\ \mathrm{d}\hat{F}^{(n-1)} + \sum_{i=2}^{n}\left(\sum_{t_k\in E_0(\hat{F}^{(n-i)})\cap\mathrm{supp}\{\varphi\}} \left(\hat{F}^{(n-i)}(t_k^+) - \hat{F}^{(n-i)}(t_k^-)\right)\langle \delta_{t_k}^{(i-1)},\varphi\rangle\right) \\ &= \int_I \varphi\hat{F}^{(n)}\mathrm{d}t + \sum_{i=1}^{n}\left(\sum_{t_k\in E_0(\hat{F}^{(n-i)})\cap\mathrm{supp}\{\varphi\}} \sigma_{\hat{F}^{(n-i)}}(t_k)\langle \delta_{t_k}^{(i-1)},\varphi\rangle\right). \end{aligned} \tag{20}$$

*Example 4* Let $\alpha < 0$ be given and set $I = [\alpha,+\infty[$. One says that $U \in \mathcal{D}'(I)$ is an impulsive-smooth distribution (see e.g. [17,28]) if there exist $l \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$ $(0 \leqslant i \leqslant l)$ and $\tilde{u} \in C^\infty(I;\mathbb{R})$ such that

$$U = \sum_{i=0}^{l} \alpha_i D^{(i)}\delta_0 + \mathcal{H}\tilde{u}$$

where $\mathcal{H}$ is the Heaviside function defined in (17). Note that if $\tilde{u}$ is a Bohl function (see Example 2) then $U$ is called a Bohl distribution (see e.g. [17,28]). Let us now remark that

$$U \in \mathcal{T}_{l+1}(I).$$

Indeed, let $F : I \to \mathbb{R}$ be the function defined by

$$F(t) = \sum_{i=0}^{l} \alpha_i \mathcal{H}(t)\frac{t^{l-i}}{(l-i)!} + I_{l+1}(t), \quad \forall\, t \in I,$$

where $I_{l+1}$ is the function defined on $I$ by recurrence as follows:

$$\begin{cases} I_1(t) &= \int_0^t \mathcal{H}(s)\tilde{u}(s)\mathrm{d}s \\ I_{k+1}(t) &= \int_0^t I_k(s)\mathrm{d}s, \qquad k = 1,\ldots,l. \end{cases}$$

It is easy to check that $F \in \mathcal{F}_\infty(I;\mathbb{R})$. Moreover $\mathrm{D}^{(l+1)}F = U$ and thus $U \in \mathcal{T}_{l+1}(I)$.

For a distribution $T \in \mathcal{T}_n(I)$, as expressed in (20), we may clearly identify the "function part" $\{T\}$ and the "measure part" $\langle\langle T \rangle\rangle$ respectively by

$$\{T\} = \hat{F}^{(n)} \tag{21}$$

and (if $n \geqslant 1$)

$$\langle\, \langle\langle T \rangle\rangle, \varphi \rangle = \int_I \varphi \,\mathrm{d}\hat{F}^{(n-1)}, \quad \forall\, \varphi \in C_0^\infty(I). \tag{22}$$

We will also use the notation $\mathrm{d}T$ to denote the Stieltjes measure $\mathrm{d}\hat{F}^{(n-1)}$ generated by $\hat{F}^{(n-1)} \in \mathbf{RCSLBV}(\mathbf{I}; \mathbb{R})$. Here $\{T\}$ is a **RCSLBV** function and $\mathrm{d}\langle\!\langle \mathrm{T} \rangle\!\rangle$ is a Stieltjes measure. For pedagogical reasons, we use the two different notation $\mathrm{d}\langle\!\langle \mathrm{T} \rangle\!\rangle$ and $\langle\langle T \rangle\rangle$ to denote respectively the Radon measure defined on the Borel sets and the corresponding distribution, i.e.

$$\langle\, \langle\langle T \rangle\rangle, \varphi \rangle = \int_I \varphi \,\mathrm{d}\langle\!\langle \mathrm{T} \rangle\!\rangle, \quad \forall\, \varphi \in C_0^\infty(I).$$

It will be also convenient to use the notation $\{T^{(k)}\}$ to denote the "function part" of $\mathrm{D}^k T$, i.e.

$$\{T^{(k)}\} = \{\mathrm{D}^k T\} = \hat{F}^{(n+k)}.$$

**Definition 2** *We say that a Schwartz distribution $T \in \mathcal{D}'(I)$ is of class $\mathcal{T}_\infty$ on $I$ provided that there exist $n \in \mathbb{N}$ and a function $F \in \mathcal{F}_\infty(I; \mathbb{R})$ such that $T = \mathrm{D}^n F$.*

Defining $\mathcal{T}_\infty$ therefore allows one to encompass all distributions of class $\mathcal{T}_n$, $n \in \mathbb{N}$. The set of Schwartz distributions of class $\mathcal{T}_\infty$ on $I$ will be denoted by $\mathcal{T}_\infty(I)$. It is clear that

$$\mathcal{T}_\infty(I) = \bigcup_{n \in \mathbb{N}} \mathcal{T}_n(I).$$

For $T \in \mathcal{T}_\infty(I)$, we define the degree "deg($T$)" of $T$ in the following way: Let $n$ be the smallest integer such that $T \in \mathcal{T}_n(I)$, we set

$$\deg(T) = \begin{cases} n+1 & \text{if } n \geqslant 1 \\ 1 & \text{if } n = 0 \text{ and } E_0(\{T\}) \neq \emptyset \\ 0 & \text{if } n = 0 \text{ and } E_0(\{T\}) = \emptyset. \end{cases} \tag{23}$$

*Remark 2* The distributions of degree 0 are the continuous $\mathcal{F}_\infty$-functions while the distributions of degree 1 are the discontinuous $\mathcal{F}_\infty$-functions. The right-continuous Heaviside function is of degree 1, the Dirac distribution $\delta_a$ ($a \in I$) is of degree 2, the distribution $\mathrm{D}^{(n)}\delta_a$ ($a \in I$) is of degree $n+1$.

*Example 5* Let us here consider the distribution $T$ of Example 3. We have:

$$\begin{aligned}
&T \equiv \{T\} = \hat{u}^{(0)} = |\sin(.)|, \quad \deg(T) = 0,\\
&\mathrm{D}T \equiv \{T^{(1)}\} = \{\mathrm{D}T\} = \hat{u}^{(1)} = \cos(. - k\pi) \quad \text{on } [k\pi, (k+1)\pi) \ (k \in \mathbb{N}),\\
&\deg(\mathrm{D}T) = 1,\\
&\mathrm{D}^2 T \equiv \langle\langle \mathrm{D}^2 T\rangle\rangle = -|\sin(.)| + 2 \sum_{k \in \mathbb{N}\setminus\{0\}} \delta_{k\pi}, \ \deg(\mathrm{D}^2 T) = 2,\\
&\{T^{(2)}\} = \{\mathrm{D}^2 T\} = \hat{u}^{(2)} = -|\sin(.)|,\\
&\mathrm{d}\langle\langle \mathrm{D}^2 T\rangle\rangle = \mathrm{d}\hat{u}^{(1)}.
\end{aligned}$$

## 3 The ZD canonical representation

In this section a canonical state space representations is derived, which will prove to be useful to formalize the extended sweeping process.

### 3.1 Canonical state space representation

Our treatment here is, of necessity informal. The scope of this section is not to get all the smoothness hypotheses worked out but is merely to illustrate a canonical state representation that is generally used by researchers from Systems and Control.

Let us set $I = [0, T[$ for some $T > 0$, $T \in \mathbb{R} \cup \{+\infty\}$. We consider the following dynamical problem: Find functions $x : I \to \mathbb{R}^n; t \mapsto x(t)$, $\lambda : I \to \mathbb{R}; t \mapsto \lambda(t)$ and $w : I \to \mathbb{R}; t \mapsto w(t)$ such that:

$$\begin{cases} \dot{x}(t) = Ax(t) + B\lambda(t) \ (t \in I) \\ x(0) = x_0 \\ w(t) = Cx(t) \geqslant 0 \qquad (t \in I), \end{cases} \tag{24}$$

where $x_0 \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$ and $C \in \mathbb{R}^{1\times n}$.

The transfer function $H : \mathbb{C} \to \mathbb{C}$ of the system is given by

$$H(s) = C(sI_n - A)^{-1}B.$$

We may write:

$$H(s) = \frac{N(s)}{D(s)}$$

where $D$ is a polynomial of degree $n$ and $N$ is a polynomial of degree $l < n$. The relative degree $r$ of the triple $(A, B, C)$ is defined as the difference between the degrees of the denominator and numerator polynomials of $H$, i.e. $r = n - l$. Note that $1 \leqslant r \leqslant n$.

We assume that $H$ is not zero. Equivalently, $CA^{r-1}B \neq 0$ while $CA^{i-1}B = 0$ for all $1 \leqslant i \leqslant r-1$. Then there exists a full-rank matrix $W \in \mathbb{R}^{n\times n}$ such that (see e.g. [60]):

$$WB = \begin{pmatrix} 0^{r-1} \\ CA^{r-1}B \\ 0^{n-r} \end{pmatrix},$$

$$CW^{-1} = \begin{pmatrix} 1 & 0_{n-1} \end{pmatrix}$$

and

$$WAW^{-1} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0_{n-r} \\ 0 & 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & \dots & 0 & 1 & 0_{n-r} \\ d_1 & d_2 & d_3 & \dots & d_r & d_\xi^{\mathrm{T}} \\ B_\xi & 0^{n-r} & 0^{n-r} & \dots & 0^{n-r} & A_\xi \end{pmatrix}, \tag{25}$$

where $A_\xi \in \mathbb{R}^{(n-r)\times(n-r)}, B_\xi \in \mathbb{R}^{(n-r)\times 1}$, and $(d^{\mathrm{T}}, d_\xi^{\mathrm{T}}) = (CA^rW^{-1})^{\mathrm{T}}$ with $d^{\mathrm{T}} = (d_1, \dots, d_r)$.

Actually, the framework that is presented next is essentially linked to systems with $r \geqslant 1$. The existence of a relative degree allows one to perform a state space transformation, with new state vector $z = Wx$,

$$z^{\mathrm{T}} = (z_1, z_2, \dots, z_r, \xi^{\mathrm{T}}) = \left(\bar{z}^{\mathrm{T}}, \xi^{\mathrm{T}}\right), \quad \xi \in \mathbb{R}^{n-r} \tag{26}$$

such that the new state space representation is (see [60]):

$$\begin{cases} \dot{z}(t) = WAW^{-1}z(t) + WB\lambda(t) & (t \in I) \\ z(0) = Wx_0 \\ w(t) = CW^{-1}z(t) \geqslant 0 & (t \in I) \end{cases} \tag{27}$$

that is

$$\begin{cases} \dot{z}_1(t) = z_2(t) & (t \in I) \\ \dot{z}_2(t) = z_3(t) & (t \in I) \\ \dot{z}_3(t) = z_4(t) & (t \in I) \\ \vdots \\ \dot{z}_{r-1}(t) = z_r(t) & (t \in I) \\ \dot{z}_r(t) = CA^rW^{-1}z(t) + CA^{r-1}B\lambda(t) & (t \in I) \\ \dot{\xi}(t) = A_\xi\xi(t) + B_\xi z_1(t) & (t \in I) \\ w(t) = z_1(t) \geqslant 0 & (t \in I) \\ z(0) = z_0. \end{cases} \tag{28}$$

Moreover

$$CA^rW^{-1}z = d^{\mathrm{T}}\bar{z} + d_\xi^{\mathrm{T}}\xi. \tag{29}$$

In Systems and Control theory, the dynamics $\dot{\xi} = A_\xi \xi + B_\xi z_1$ is called the *zero dynamics*, so we shall denote the state space form in (28) the ZD representation.

*Remark 3* [The multivariable case] We note that the formalism in (28) continues to hold if $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{m\times n}$, $m \geqslant 2$. One says that the triple $(A, B, C)$ has a vector relative degree $\bar{r} \triangleq [r, \ldots, r]^{\mathrm{T}} \in \mathbb{R}^m$ if there exists an integer $r \geqslant 1$ such that $mr \leqslant n$ and the following conditions are satisfied: $CA^iB = 0$ for all $i = 0, 1, \ldots, r-2$ and the matrix $CA^{r-1}B \in \mathbb{R}^{m\times m}$ is nonsingular. In this case, one gets the same expression as in (28) but all $z_i$, $1 \leqslant i \leqslant r$, are $m$-dimensional, $\xi$ is $n - mr$ dimensional, $CA^{r-1}B$ is an $m \times m$ matrix and $CA^rW^{-1}$ is an $m \times n$ matrix. Moreover $A_\xi$ is a $(n - mr) \times (n - mr)$ matrix and $B_\xi$ is a $(n-mr)\times m$ matrix, see [60] for a proof. We shall say that the system has a vector relative degree $\bar{r}$ in such a case. Such conditions on the triple $(A, B, C)$ are closely related to the invertibility of the system with input $\lambda(\cdot)$ and output $w(\cdot)$ (actually they are sufficient conditions for invertibility [59, Theorem 3]). Other similar canonical forms have been derived by Sannuti and co-workers for the sake of control applications [58,61,62], however in this paper we shall content ourselves with the assumption of a relative degree $r$.

*Example 6* Let us consider the state representation in (24) with

$$A = \begin{pmatrix} 2 & 7 & 3-2\alpha & -2\beta+2 \\ -1 & -3 & -1+\alpha & \beta-1 \\ 0 & 0 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 2 & 0 & 0 \end{pmatrix},$$

where $\alpha, \beta \in \mathbb{R}$. The transfer function of this system is given by

$$H(s) = \frac{s^2 - 1}{s^4 + s^3 - (1+\alpha)s - 1 - \beta}.$$

Then the transformation matrix

$$W = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

allows one to transform the system $(A, B, C)$ into the ZD canonical state space representation $(WAW^{-1}, WB, CW^{-1})$ where

$$WAW^{-1} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & -1 & \alpha & \beta \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \quad WB = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad CW^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix},$$

so that the dynamics in (28) is given by

$$\begin{cases} \dot{z}_1(t) = z_2(t) & (t \in I) \\ \dot{z}_2(t) = \lambda(t) - z_1(t) - z_2(t) + \alpha\xi_1(t) + \beta\xi_2(t) & (t \in I) \\ \dot{\xi}_1(t) = \xi_2(t) & (t \in I) \\ \dot{\xi}_2(t) = \xi_1(t) + z_1(t) & (t \in I) \\ w(t) = z_1(t) & (t \in I). \end{cases} \tag{30}$$

### 3.2 Distributional dynamics model

Starting from this section on, we set $I = [0, T[$ for some $T > 0$, $T \in \mathbb{R} \cup \{+\infty\}$, and only functions from the class $\mathbf{RCSLBV}(\mathbf{I}; \mathbb{R})$ and distributions from the class $\mathcal{T}_\infty(I)$ are used. Moreover, in order to simplify the presentation of our problem, we shall assume that $m = 1$. When the statements or results also obviously hold for the multivariable case $m \geqslant 2$ with vector relative degree $\bar{r}$ this will be pointed out.

In this section, we present our model. Recalling that all concrete mathematical models were born without mathematical maturity, we confess that it is also the case of the following one. We need thus in this section to stay beyond some standards required in the mathematical literature. Happily, this will be brief, and our aim in the following sections will be precisely to bring our model to come to its mathematical maturity.

It is of utmost importance to notice that in general, the possible solutions of (28) (equivalently of (24)) cannot be defined in a class of smooth functions. Consider for instance the initial data $z_{0,i} \leqslant -\delta$ for some $\delta > 0$ and all $1 \leqslant i \leqslant r$. Then, since the unilateral constraint $z_1 \geqslant 0$ must be satisfied on $I$, it is necessary that $z_1(0^+) \geqslant 0$, i.e. $z_1$ needs to "jump" to some non-negative value. It results that $z_1$ cannot be continuous and the derivatives in (28) must be considered in the sense of distributions. At this stage we can just say that a jump mapping is needed. Its form will depend on the type of system one handles (in Mechanics, this is the realm of impact mechanics [8]). If one considers (28) as an equality of distributions of class $\mathcal{T}_\infty(I)$, then we can rewrite it as

$$\begin{cases} Dz_1 = z_2 \\ Dz_2 = z_3 \\ Dz_3 = z_4 \\ \vdots \\ Dz_{r-1} = z_r \\ Dz_r = CA^rW^{-1}z + CA^{r-1}B\lambda \\ D\xi = A_\xi\xi + B_\xi z_1. \end{cases} \tag{31}$$

Consider the above initial conditions on $\{z_i\}$ $(1 \leqslant i \leqslant r)$. Then $Dz_1$ is a distribution of degree 2 and we get $Dz_1 = \{\dot{z}_1\} + \sigma_{z_1}(0)\delta_0 = z_2$. Consequently $Dz_2$ is a distribution of degree 3 and $Dz_2 = D^2z_1 = D\{\dot{z}_1\} + \sigma_{z_1}(0)D\delta_0 = \{\dot{z}_2\} + \sigma_{\{\dot{z}_1\}}(0)\delta_0 + \sigma_{z_1}(0)D\delta_0 = z_3$, and $\{\dot{z}_1\} = \{z_2\}$. Then $Dz_3$ is a distribution of degree 4, and we get $Dz_3 = D\{\dot{z}_2\} + \sigma_{\{\dot{z}_1\}}(0)D\delta_0 + \sigma_{z_1}(0)D^2\delta_0 = \{\dot{z}_3\} + \sigma_{\{\dot{z}_2\}}(0)\delta_0 + \sigma_{\{\dot{z}_1\}}(0)D\delta_0 + \sigma_{z_1}(0)D^2\delta_0 = z_4$, and $\{\dot{z}_2\} = \{z_3\}$, and so on. Thus $\sigma_{\{\dot{z}_1\}}(0) = \{z_2\}(0^+) - \{z_2\}(0^-)$, $\sigma_{\{\dot{z}_2\}}(0) = \{z_3\}(0^+) - \{z_3\}(0^-)$, and so on. Until now we have decomposed only the left hand side of the dynamics as distributions of some degrees. Now let us get back to the distributional dynamics in (31). Starting from $Dz_1 = z_2$, one deduces that the right hand side has to be of the same degree than the left hand side. This means that the right hand side is equal to $\{z_2\} + \nu_1$, where $\nu_1$ is a distribution of degree 2, i.e. a measure. Similarly from $Dz_2 = z_3$ one deduces that $z_3 = \{z_3\} + \tilde{\nu}_2$, where $\tilde{\nu}_2$ has degree 3 and can therefore further be decomposed as $\nu_2 + \tilde{\nu}_1$, with $\deg(\nu_2) = 2$ and $\deg(\tilde{\nu}_1) = 3$. It is not difficult to see that $\tilde{\nu}_1 = D\nu_1$. Therefore $Dz_2 = \{z_3\} + \nu_2 + D\nu_1$. The variables $\nu_1$ and $\nu_2$ are slack variables (or Lagrange multipliers), and are measures of the form $\nu_i = \int_I \mathrm{d}\nu_i$, where $\mathrm{d}\nu_i$ is a Stieltjes measure generated by a $\mathcal{F}_\infty(I;\mathbb{R})$- function. Continuing the reasoning until $Dz_r$, we obtain $Dz_r = CA^rW^{-1}\{z\} + CA^{r-1}B\lambda$ where $\deg(\lambda) = \deg(Dz_r) = r + 1$. Consequently from (31) one gets

$$\begin{cases} Dz_1 = \{z_2\} + \nu_1 \\ Dz_2 = \{z_3\} + D\nu_1 + \nu_2 \\ Dz_3 = \{z_4\} + D^2\nu_1 + D\nu_2 + \nu_3 \\ \vdots \\ Dz_i = \{z_{i+1}\} + D^{(i-1)}\nu_1 + D^{(i-2)}\nu_2 + \ldots + D\nu_{i-1} + \nu_i \\ \vdots \\ Dz_{r-1} = \{z_r\} + D^{(r-2)}\nu_1 + \ldots + D\nu_{r-2} + \nu_{r-1} \\ \\ Dz_r = CA^rW^{-1}\{z\} + CA^{r-1}B\lambda. \end{cases} \tag{32}$$

We keep the notation $\lambda$ for the multiplier which appears in the last line. One sees that $\lambda$ in (32) can be given a meaning as

$$\lambda = (CA^{r-1}B)^{-1}[D^{(r-1)}\nu_1 + \cdots + D\nu_{r-1}] + \nu_r \tag{33}$$

provided $CA^{r-1}B \neq 0$ (invertible in the multivariable case $m \geqslant 2$ with relative degree $\bar{r}$). Then $\lambda$ is uniquely defined as in (33).

It is important at this stage to realize that $\lambda$ is the unique source of higher degree distributions in the system, which will allow the state to jump. Therefore the measures $\nu_i$ have themselves to be considered as sub-multipliers. In (32) we have separated the regular (functions) parts denoted as $\{\cdot\}$ (see Sect. 2) and the atomic distributional parts.

Only the Dirac measures $\nu_i$ and time functions are signed. Consequently imposing $\lambda \geqslant 0$ is meaningless in general. Another point of view is to assert that $\lambda \geqslant 0$ implies that $\lambda$ is a measure. However this is not sufficient to assure $z_1 \geqslant 0$ along the time integration. Consequently one has to resort to higher degree distributions to give a reasonably general meaning to the dynamics in (24).

Our aim is now to propose a mathematical formulation applicable to the study of our problem. Let us here also suppose that $CA^{r-1}B \neq 0$ (invertible in the multivariable case $m \geqslant 2$ with vector relative degree $\bar{r}$).

**Distributional formalism** A mathematical problem that appears suitable for the study of the dynamics in (31), (32), (33) consists to find $z_1, \dots, z_r \in \mathcal{T}_\infty(I)$ and $\xi_i \in \mathcal{T}_\infty(I)$ $(1 \leqslant i \leqslant n-r)$ satisfying the distributional equations

$$\begin{cases} Dz_1 - z_2 = 0 \\ Dz_2 - z_3 = 0 \\ Dz_3 - z_4 = 0 \\ \vdots \\ Dz_{r-1} - z_r = 0 \\ Dz_r - CA^rW^{-1}\{z\} = CA^{r-1}B\lambda \\ D\xi = A_\xi \xi + B_\xi z_1 \end{cases} \tag{34}$$

together with some constraints (that will be specified and discussed in the following section) on the distributions (of degree $\leqslant 2$) $\nu_1, \dots, \nu_r$ defined by

$$\begin{cases} \nu_1 := \langle\langle Dz_1 - \{z_2\}\rangle\rangle \\ \nu_2 := \langle\langle Dz_2 - \{z_3\}\rangle\rangle \\ \nu_3 := \langle\langle Dz_3 - \{z_4\}\rangle\rangle \\ \vdots \\ \nu_{r-1} := \langle\langle Dz_{r-1} - \{z_r\}\rangle\rangle \\ \nu_r := \langle\langle Dz_r - CA^rW^{-1}\{z\}\rangle\rangle (= CA^{r-1}B\langle\langle\lambda\rangle\rangle) \end{cases} \tag{35}$$

and

$$\lambda = (CA^{r-1}B)^{-1}[D^{(r-1)}\nu_1 + \dots + D\nu_{r-1}] + \nu_r. \tag{36}$$

Each distribution $\nu_i$ is of the form

$$\langle \nu_i, \varphi \rangle = \int_I \varphi \, \mathrm{d}\nu_i, \quad \forall\, \varphi \in C_0^\infty(I) \tag{37}$$

where $\mathrm{d}\nu_i$ is a Stieltjes measure generated by a $\mathcal{F}_\infty(I;\mathbb{R})$-function.

Let us now denote by $\mathrm{d}\{z_i\}$ $(1 \leqslant i \leqslant r)$ the Stieltjes measures given by:

$$\langle \langle\langle Dz_i \rangle\rangle, \varphi \rangle = \int_I \varphi \, \mathrm{d}\{z_i\}, \quad \forall\, \varphi \in C_0^\infty(I), \tag{38}$$

that is also the Stieltjes measure generated by $\{z_i\}$. We see that

$$\mathrm{d}\nu_i = \mathrm{d}\langle\langle Dz_i - \{z_{i+1}\}\rangle\rangle = \mathrm{d}\{z_i\} - \{z_{i+1}\}(t)\mathrm{d}t, \quad (1 \leqslant i \leqslant r-1) \tag{39}$$

and

$$\mathrm{d}\nu_r = \mathrm{d}\langle\langle Dz_r - CA^rW^{-1}\{z\}\rangle\rangle = \mathrm{d}\{z_r\} - CA^rW^{-1}\{z\}(t)\mathrm{d}t. \tag{40}$$

If $K$ denotes a closed convex cone then the expression $d\nu_i \in K$ has a sense that has been specified in Sect. 2 (see (2) and (3)). The constraints will be specified and discussed in the following section.

*Remark 4* The structure of the system in (34) ensures that necessarily $z_i$ is a distribution of degree $\leqslant i$ while $\lambda$ is a distribution of degree $\leqslant r+1$. It results that $z_1 \equiv \{z_1\}$ and $\deg(z_1) \leqslant 1$. Moreover $\xi \equiv \{\xi\}$, $\deg(\xi) = 0$ and $\xi$ is continuous. The zero-dynamics reduces to

$$\mathrm{d}\xi = (A_\xi \xi(t) + B_\xi z_1(t))\mathrm{d}t.$$

**Measure differential formalism** The distributional formalism contains all the information on the dynamics. However, it would be convenient to end up with a "weaker" formalism for (31), (32) by requiring that the solutions $z_i$ of the system in (31), (32) are regular distributions $z_i$ generated by right continuous functions of special locally bounded variation. More precisely, our problem consists to find $z_1, \ldots, z_r, \xi_1, \ldots, \xi_{n-r} \in \mathcal{F}_\infty(I;\mathbb{R})$ such that

$$\begin{cases} \mathrm{d}z_1 = z_2(t)\mathrm{d}t + \mathrm{d}\nu_1 \\ \mathrm{d}z_2 = z_3(t)\mathrm{d}t + \mathrm{d}\nu_2 \\ \mathrm{d}z_3 = z_4(t)\mathrm{d}t + \mathrm{d}\nu_3 \\ \vdots \\ \mathrm{d}z_i = z_{i+1}(t)\mathrm{d}t + \mathrm{d}\nu_i \\ \vdots \\ \mathrm{d}z_{r-1} = z_r(t)\mathrm{d}t + \mathrm{d}\nu_{r-1} \\ \mathrm{d}z_r = CA^rW^{-1}z(t)\mathrm{d}t + CA^{r-1}B\mathrm{d}\nu_r \\ \mathrm{d}\xi = (A_\xi\xi(t) + B_\xi z_1(t))\mathrm{d}t \end{cases} \tag{41}$$

where $\mathrm{d}\nu_i$ denotes a Radon measure and $\mathrm{d}z_i$ $(1 \leqslant i \leqslant r)$ is the Stieltjes measure generated by $z_i$. If $K$ denotes a closed convex cone then the expression $\mathrm{d}\nu_i \in K$ $(1 \leqslant i \leqslant r)$ can be defined as in Sect. 2.

Here, all distributions, let us say $T$, invoked in the system (31), (32) are of degree $\leqslant 1$ and thus $\langle\langle T\rangle\rangle = T$. We have also $\{z_i\} \equiv z_i$ $(1 \leqslant i \leqslant r)$.

*Remark 5* In Mechanics one has $r = 2$ and $\lambda$ is of degree 2 because $\nu_1 = 0$ (the position $z_1(\cdot)$ is locally absolutely continuous, see e.g. [42,70]). In dissipative electrical circuits with $r = 1$ then possible inconsistent initial data on $z_1$ may lead to $\nu_1(= \lambda)$ of degree 2 [17]. The measures $\nu_i$ and the distribution $\lambda$ in (32) play a similar role to the Lagrange multiplier in Mechanics with unilateral contact. The idea of the extended Moreau's sweeping process is to represent these atomic distributions in a way similar to what is done in Mechanics and to subsequently take advantage of this formalism to derive a time-stepping numerical algorithm. Moreover, viewing the dynamics as an equality of distributions as in (32) paves the way towards time-discretization with time-stepping algorithms, i.e. numerical schemes working without event detection procedures and with constant time-step.

*Remark 6* The idea of observing the derivatives of the variable $w$ in order to analyze unilaterally constrained systems is certainly not new, and has been often used previously [28,71]. Also it has been long well known in DAE theory that higher degree distributions can occur due to inconsistent initial state [37,20]. Therefore introducing such ingredients in dynamical systems is by far not new. It is however clear that the gap between DAE and differential inclusions, is not trivial.

## 4 The extended Moreau's sweeping process

### 4.1 Preliminaries

Let us first recall that in order to simplify the presentation we shall continue to assume in many places that $m = 1$.

Let $D$ be a nonempty closed convex subset of $\mathbb{R}$. We denote by $T_D(x)$ the tangent cone of $D$ at $x \in \mathbb{R}$ defined by

$$T_D(x) = \overline{\mathrm{cone}}(D - \{x\}) \tag{42}$$

where $\mathrm{cone}(D - \{x\})$ denotes the cone generated by $D - \{x\}$ and $\overline{\mathrm{cone}}(D - \{x\})$ denotes the closure of $\mathrm{cone}(D - \{x\})$, i.e. $\overline{\mathrm{cone}}(D - \{x\}) = \overline{\mathrm{cone(D - \{x\})}}$. The definition in (42) allows us to take into account constraints violations. Note that

$$T_{\mathbb{R}^+}(x) = \begin{cases} \mathbb{R} & \text{if } x > 0 \\ \mathbb{R}^+ & \text{if } x \leqslant 0 \end{cases}$$

and

$$T_{\mathbb{R}}(x) = \mathbb{R}.$$

Let us now set

$$\Phi := \mathbb{R}^+. \tag{43}$$

For $z \in \mathbb{R}^r$, we set

$$Z_i = (z_1, z_2, \ldots, z_i), \quad (1 \leqslant i \leqslant r). \tag{44}$$

By convention, we set $Z_0 = 0$ and

$$T^0_\Phi(Z_0) = \Phi$$

and we define

$$\begin{cases} T^1_\Phi(Z_1) = T_\Phi(z_1), \\ T^2_\Phi(Z_2) = T_{T^1_\Phi(Z_1)}(z_2), \\ \qquad \vdots \\ T^r_\Phi(Z_r) = T_{T^1_\Phi(Z_{r-1})}(z_r), \end{cases}$$

that is

$$T^i_\Phi(Z_i) = T_{T^{i-1}_\Phi(Z_{i-1})}(z_i), \quad 1 \leqslant i \leqslant r.$$

*Remark 7* In the multivariable case $m \geqslant 2$ with vector relative degree $\bar{r}$, we have $\Phi = (\mathbb{R}^+)^m, Z^l_0 = 0, Z^l_i = \left(z^l_1, z^l_2, \ldots, z^l_i\right), 1 \leqslant i \leqslant r, 1 \leqslant l \leqslant m$, and

$$T^i_\Phi(Z_i) = \times^m_{l=1} T^i_\Phi\left(Z^l_i\right), \quad 1 \leqslant i \leqslant r.$$

Recalling that $I = [0, T[$ for some $0 < T \leqslant +\infty$, and starting from (31), (32) the extended sweeping process is written as follows

$$\mathrm{d}v_i \in -\partial\psi_{T^{i-1}_\Phi(\{Z_{i-1}\}(t^-))}(\{z_i\}(t^+)) \quad \text{on } I, \ \ (1 \leqslant i \leqslant r), \tag{45}$$

with $\mathrm{d}v_i$ in (37)–(40). Here $\{z_i\}(0^-)$ $(1 \leqslant i \leqslant r)$ will be given (by convention) so as to define some initial conditions for the process (see (52) and Remark 9). The sets $\partial\psi_{T^{i-1}_\Phi(\{Z_{i-1}\}(t^-))}(\{z_i\}(t^+))$ $(1 \leqslant i \leqslant r)$ are nonempty closed convex cones and the sense of the inclusions in (45) is given (see Sect. 2) by the existence of nonnegative real-valued Radon measures $d\mu_i$ $(1 \leqslant i \leqslant r)$ relative to which $\mathrm{d}v_i$ possess densities $\mathrm{d}v_i/\mathrm{d}\mu_i$ such that:

$$\frac{\mathrm{d}v_i}{\mathrm{d}\mu_i}(t) \in -\partial\psi_{T^{i-1}_\Phi(\{Z_{i-1}\}(t^-))}(\{z_i\}(t^+)), \quad \mathrm{d}\mu_i - a.e. \ t \in I \ \ (1 \leqslant i \leqslant r). \tag{46}$$

The sets $T^{i-1}_\Phi(\{Z_{i-1}\}(t^-))$ $(1 \leqslant i \leqslant r)$ are nonempty closed convex cones and thus the system in (46) can also be written equivalently as follows:

$$\begin{cases} \{z_i\}(t^+) \in T^{i-1}_\Phi(\{Z_{i-1}\}(t^-)), & \mathrm{d}\mu_i - \text{a.e. } t \in I \ \ (1 \leqslant i \leqslant r), \\ \left\langle \frac{\mathrm{d}v_i}{\mathrm{d}\mu_i}(t), \{z_i\}(t^+) \right\rangle = 0, & \mathrm{d}\mu_i - \text{a.e. } t \in I \ \ (1 \leqslant i \leqslant r), \\ \left\langle \frac{\mathrm{d}v_i}{\mathrm{d}\mu_i}(t), v \right\rangle \geqslant 0, & \forall\, v \in T^{i-1}_\Phi(\{Z_{i-1}\}(t^-)), \\ & \mathrm{d}\mu_i - \text{a.e. } t \in I \ \ (1 \leqslant i \leqslant r). \end{cases} \tag{47}$$

*Remark 8*

i) Using the notation and conventions specified above, the first inclusion in (45) reads

$$\mathrm{d}\nu_1 \in -\partial\psi_\Phi(\{z_1\}(t^+)).$$

ii) Recall that if $z_i \in T_\Phi^{i-1}(Z_{i-1})$ then

$$\partial\psi_{T_\Phi^{i-1}(Z_{i-1})}(z_i) = \{w \in \mathbb{R} : \langle w, v - z_i\rangle \leqslant 0, \forall v \in T_\Phi^{i-1}(Z_{i-1})\}$$

is the outward normal cone to $T_\Phi^{i-1}(Z_{i-1})$ at $z_i$.

iii) Note that

$$\begin{aligned}
&T_\Phi^{i-1}(Z_{i-1}) = \mathbb{R} \;\Rightarrow\; \partial\psi_{T_\Phi^{i-1}(Z_{i-1})}(z_i) = \{0\},\\
&T_\Phi^{i-1}(Z_{i-1}) = \mathbb{R}^+ \text{ and } z_i > 0 \;\Rightarrow\; \partial\psi_{T_\Phi^{i-1}(Z_{i-1})}(z_i) = \{0\},\\
&T_\Phi^{i-1}(Z_{i-1}) = \mathbb{R}^+ \text{ and } z_i \leqslant 0 \;\Rightarrow\; \partial\psi_{T_\Phi^{i-1}(Z_{i-1})}(z_i) = \mathbb{R}^-.
\end{aligned}$$

iv) Starting from (24) and (28) one is tempted to write the inclusion "$Dz_r - CA^rW^{-1}z \in -CA^{r-1}B\;\partial\psi_\Phi(\{z_1\}(t^+))$" which makes sense only if $\lambda$ is a measure since $\partial\psi_\Phi(\{z_1\}(\cdot^+))$ is a cone. This inclusion is replaced by

$$\mathrm{d}\nu_r \in -\;\partial\psi_{T_\Phi^{r-1}(\{Z_{r-1}\}(t^-))}(\{z_r\}(t^+)) \quad \text{on } I,$$

in (45). The positivity of $\lambda$ is now understood as the positivity of $\nu_r$ (see also Theorem 1 below).

v) It is then important to see that the distributional dynamics

$$D^r z_1 = Dz_r = CA^rW^{-1}\{z\} + CA^{r-1}B\lambda \tag{48}$$

with $\lambda$ in (45), (33), is equivalent to (32), (45). Notice that (32), (45), (48) is the same as (31), (33), (45).

vi) It is noteworthy that though the formalism is presented in the coordinates in (24), its implementation requires the knowledge of $A$, $B$, $C$ and $x(\cdot)$ only. We however believe that the state space representation in (28) allows one to better understand the underlying dynamics.

## 4.2 Mathematical formalisms

Let us now complete the formalisms introduced in Sect. 3.2. Let $T > 0$, $T \in \mathbb{R} \cup \{+\infty\}$ be given and set $I = [0, +T[$.

Let

$$z_0^{\mathrm{T}} = \left(\bar{z}_0^{\mathrm{T}}, \xi_0^{\mathrm{T}}\right)$$

be given in $\mathbb{R}^n$ with $\bar{z}_0 \in \mathbb{R}^r$ and $\xi_0 \in \mathbb{R}^{n-r}$.

**Distributional formalism** Using (34), (35), (36), (37), (38), (39), (40) and (45), our mathematical problem reads:

✠ **PROBLEM SP**($z_0$**;** $I$) ✠ Find $z_1, \ldots, z_r \in \mathcal{T}_\infty(I)$ and $\xi_i \in \mathcal{T}_\infty(I)$ $(1 \leqslant i \leqslant n-r)$ satisfying the distributional equations:

$$\begin{cases} Dz_1 - z_2 = 0 \\ Dz_2 - z_3 = 0 \\ Dz_3 - z_4 = 0 \\ \\ \vdots \\ \\ Dz_{r-1} - z_r = 0 \\ Dz_r - CA^r W^{-1}\{z\} = CA^{r-1}B\lambda \\ \\ D\xi = A_\xi \xi + B_\xi z_1 \end{cases} \tag{49}$$

$$\lambda = (CA^{r-1}B)^{-1} \left[ \sum_{i=1}^{r-1} D^{(r-i)} \langle\langle Dz_i - \{z_{i+1}\} \rangle\rangle \right] + \langle\langle Dz_r - CA^r W^{-1}\{z\} \rangle\rangle, \tag{50}$$

the measure differential inclusions on $]0, T[$:

$$\begin{cases} \mathrm{d}\{z_1\} - \{z_2\}(t)\mathrm{d}t \;\in -\partial\psi_\Phi(\{z_1\}(t^+)), \\ \mathrm{d}\{z_2\} - \{z_3\}(t)\mathrm{d}t \in -\partial\psi_{T^1_\Phi(\{Z_1\}(t^-))}(\{z_2\}(t^+)), \\ \vdots \\ \mathrm{d}\{z_i\} - \{z_{i+1}\}(t)\mathrm{d}t \in -\partial\psi_{T^{i-1}_\Phi(\{Z_{i-1}\}(t^-))}(\{z_i\}(t^+)), \\ \vdots \\ \mathrm{d}\{z_{r-1}\} - \{z_r\}(t)\mathrm{d}t \in -\partial\psi_{T^{r-2}_\Phi(\{Z_{r-2}\}(t^-))}(\{z_{r-1}\}(t^+)), \\ (CA^{r-1}B)^{-1}[\mathrm{d}\{z_r\} - CA^r W^{-1}\{z\}(t)\mathrm{d}t] \in -\partial\psi_{T^{r-1}_\Phi(\{Z_{r-1}\}(t^-))}(\{z_r\}(t^+)) \end{cases} \tag{51}$$

and the initial conditions:

$$\begin{cases} \{z_1\}(0^+) - z_{0,1} \;\in -\partial\psi_\Phi(\{z_1\}(0^+)), \\ \{z_2\}(0^+) - z_{0,2} \in -\partial\psi_{T^1_\Phi(Z_{0,1})}(\{z_2\}(0^+)), \\ \vdots \\ \{z_i\}(0^+) - z_{0,i} \in -\partial\psi_{T^{i-1}_\Phi(Z_{0,i-1})}(\{z_i\}(0^+)), \\ \vdots \\ \{z_{r-1}\}(0^+) - z_{0,r-1} \in -\partial\psi_{T^{r-2}_\Phi(Z_{r-2})}(\{z_{r-1}\}(0^+)), \\ (CA^{r-1}B)^{-1}[\{z_r\}(0^+) - z_{0,r}] \in -\partial\psi_{T^{r-1}_\Phi(Z_{0,r-1})}(\{z_r\}(0^+)) \end{cases} \tag{52}$$

and

$$\{\xi\}(0^+) = \xi_0. \tag{53}$$

*Remark 9* If $z$ denotes a solution of problem SP($z_0$; $I$) then we will write by convention that

$$\{\bar{z}\}(0^-) = \bar{z}_0,\ \{\xi\}(0^-) = \xi_0.$$

Then the relations in (51) formulated on $]0, T[$ together with the initial conditions in (52) reduce to the relations in (51) formulated on $I = [0, T[$. Moreover, recalling that $\xi$ is here necessarily a continuous function (see Remark 4), the condition in (53) reads:

$$\xi(0) = \xi_0 \tag{54}$$

and with our convention, we see that the last relation in (49) together with the initial condition (53) reduce to the measure differential equation (see Remark 4):

$$\mathrm{d}\xi - (A_\xi \xi(t) + B_\xi z_1(t))\mathrm{d}t = 0 \quad \text{on} \ \ I.$$

*Remark 10* The relations given in (51), formulated on $I$, have to be interpreted in the following sense: Find nonnegative real-valued Radon measures $\mathrm{d}\mu_i$ $(1 \leqslant i \leqslant r)$ relative to which the Lebesgue measure $\mathrm{d}t$ and the Stieltjes measure $\mathrm{d}\{z_i\}$ possess densities $\frac{\mathrm{d}t}{\mathrm{d}\mu_i}$ and $\frac{\mathrm{d}\{z_i\}}{\mathrm{d}\mu_i}$ respectively such that:

$$\frac{\mathrm{d}\{z_i\}}{\mathrm{d}\mu_i}(t) - \{z_{i+1}\}(t)\frac{\mathrm{d}t}{\mathrm{d}\mu_i}(t) \in -\partial\psi_{T_\Phi^{i-1}(\{Z_{i-1}\}(t^-))}(\{z_i\}(t^+)),$$

$$\mathrm{d}\mu_i - a.e. \quad t \in I \ \ (1 \leqslant i \leqslant r-1) \tag{55}$$

and

$$(CA^{r-1}B)^{-1}\left[\frac{\mathrm{d}\{z_r\}}{\mathrm{d}\mu_r}(t) - CA^r W^{-1}\{z\}(t)\frac{\mathrm{d}t}{\mathrm{d}\mu_r}(t)\right]$$

$$\times \in -\partial\psi_{T_\Phi^{r-1}(\{Z_{r-1}\}(t^-))}(\{z_r\}(t^+)), \quad \mathrm{d}\mu_r - a.e. \ t \in I. \tag{56}$$

*Remark 11* The dynamical system represented in problem **SP**($z_0$; $I$) will be called the "Higher Order Moreau's Sweeping Process".

**Measure differential formalism** Using (41), (45) as well as the conventions given in Remark 9, our mathematical problem reads:

✠ **PROBLEM MP(**$z_0$**;** $I$) ✠ Find $z_i \in \mathcal{F}_\infty(I;\mathbb{R})$ ($1 \leqslant i \leqslant r$) and $\xi_i \in \mathcal{F}_\infty(I;\mathbb{R})$ ($1 \leqslant i \leqslant n-r$) such that

$$\mathrm{d}z_i - z_{i+1}(t)\mathrm{d}t \in -\partial\psi_{T_\Phi^{i-1}(Z_{i-1}(t^-))}(z_i(t^+)) \quad \text{on } I \ \ (1 \leqslant i \leqslant r-1) \tag{57}$$

$$(CA^{r-1}B)^{-1}[\mathrm{d}z_r - CA^rW^{-1}z(t)\mathrm{d}t] \in -\partial\psi_{T_\Phi^{r-1}(Z_{r-1}(t^-))}(z_r(t^+)) \quad \text{on } I \tag{58}$$

and

$$\mathrm{d}\xi - (A_\xi\xi(t) + B_\xi z_1(t))\mathrm{d}t = 0 \quad \text{on } I. \tag{59}$$

The system in (57) and (58) has to be interpreted in the following sense: find nonnegative real-valued Radon measure $\mathrm{d}\mu_i$ relative to which the Lebesgue measure $\mathrm{d}t$ and the Stieltjes measure $\mathrm{d}z_i$ possess densities $\frac{\mathrm{d}t}{\mathrm{d}\mu_i}$ and $\frac{\mathrm{d}z_i}{\mathrm{d}\mu_i}$ respectively such that

$$\frac{\mathrm{d}z_i}{\mathrm{d}\mu_i}(t) - z_{i+1}(t)\frac{\mathrm{d}t}{\mathrm{d}\mu_i}(t) \in -\partial\psi_{T_\Phi^{i-1}(Z_{i-1}(t^-))}(z_i(t^+)), \ \ \mathrm{d}\mu_i\text{-}a.e. \ t \in I \ \ (1 \leqslant i \leqslant r-1) \tag{60}$$

and

$$(CA^{r-1}B)^{-1}\left[\frac{\mathrm{d}z_r}{\mathrm{d}\mu_r}(t) - CA^rW^{-1}z(t)\frac{\mathrm{d}t}{\mathrm{d}\mu_r}(t)\right] \\ \times \in -\partial\psi_{T_\Phi^{r-1}(Z_{r-1}(t^-))}(z_r(t^+)), \quad \mathrm{d}\mu_r - a.e. \ t \in I. \tag{61}$$

*Remark 12* i) The model in (60)–(61) is the same that the one given in (55)–(56) since here $\{z_i\} \equiv z_i$.
ii) If $z_1$ is piecewise continuous then the solution $\xi$ of (59) is given by

$$\xi(t) = \mathrm{e}^{A_\xi t}\xi_0 + \int_0^t \mathrm{e}^{A_\xi(t-\tau)}B_\xi z_1(\tau)\mathrm{d}\tau.$$

iii) It is noteworthy that the differential inclusions which are considered in this paper, are specific DI (which could be named *unbounded* DI), which cannot be analyzed using the tools for "standard" DIs of the form " $\dot{x}(t) \in F(x(t))$" as in [3]. Especially the basic assumptions for standard DIs are that $F(x)$ is compact for each $x$, and that a linear growth condition holds. Such assumptions do not hold for the inclusions considered here, since the sets in the right hand sides are normal cones, hence unbounded sets. Such a kind of differential inclusions gave rise to all the mathematical and numerical studies on the sweeping process and mechanical systems subject to unilateral constraints [4,5,15,18,19,28,32–36,40,42,45–48,63,64,70].

**Variational inequalities** The system in (55)–(56) (and consequently the one in (60)–(61) too) can be written as the evolution variational inequalities:

$$\begin{cases} \left\langle \frac{\mathrm{d}\{z_i\}}{\mathrm{d}\mu_i}(t) - \{z_{i+1}\}(t)\frac{\mathrm{d}t}{\mathrm{d}\mu_i}(t), v - \{z_i\}(t^+)\right\rangle \geqslant 0, \quad \forall\, v \in T_\Phi^{i-1}(\{Z_{i-1}\}(t^-)), \\ \mathrm{d}\mu_i - a.e. \quad t\in I \ \ (1\leqslant i \leqslant r-1), \\ \langle (CA^{r-1}B)^{-1}\left[\frac{\mathrm{d}\{z_r\}}{\mathrm{d}\mu_r}(t) - CA^rW^{-1}\{z\}(t)\frac{\mathrm{d}t}{\mathrm{d}\mu_r}(t)\right], \quad v - \{z_r\}(t^+)\rangle \geqslant 0, \\ \forall \quad v\in T_\Phi^{r-1}(\{Z_{r-1}\}(t^-)), \quad \mathrm{d}\mu_r - a.e.\ t\in I. \end{cases} \tag{62}$$

**Complementarity systems** The sets $T_\Phi^{i-1}(\{Z_{i-1}\}(t^-))$ $(1 \leqslant i \leqslant r)$ are nonempty closed convex cones and thus the system in (62) can also be written equivalently as the evolution complementarity systems:

$$\begin{cases} \{z_i\}(t^+) \in T_\Phi^{i-1}(\{Z_{i-1}\}(t^-)) \quad \mathrm{d}\mu_i - a.e.\ t\in I \ \ (1\leqslant i\leqslant r), \\ \left\langle \frac{\mathrm{d}\{z_i\}}{\mathrm{d}\mu_i}(t) - \{z_{i+1}\}(t)\frac{\mathrm{d}t}{\mathrm{d}\mu_i}(t), \{z_i\}(t^+)\right\rangle = 0, \quad \mathrm{d}\mu_i - a.e.\ t\in I \ \ (1\leqslant i\leqslant r-1), \\ \langle (CA^{r-1}B)^{-1}\left[\frac{\mathrm{d}\{z_r\}}{\mathrm{d}\mu_r}(t) - CA^rW^{-1}\{z\}(t)\frac{\mathrm{d}t}{\mathrm{d}\mu_r}(t)\right], \ \{z_r\}(t^+)\rangle = 0,\ \mathrm{d}\mu_r - a.e.\ t\in I, \\ \left\langle \frac{\mathrm{d}\{z_i\}}{\mathrm{d}\mu_i}(t) - \{z_{i+1}\}(t)\frac{\mathrm{d}t}{\mathrm{d}\mu_i}(t), v\right\rangle \geqslant 0, \quad \forall\, v\in T_\Phi^{i-1}(\{Z_{i-1}\}(t^-)), \\ \quad \mathrm{d}\mu_i - a.e.\ t\in I \ \ (1\leqslant i\leqslant r-1), \\ \langle (CA^{r-1}B)^{-1}\left[\frac{\mathrm{d}\{z_r\}}{\mathrm{d}\mu_r}(t) - CA^rW^{-1}\{z\}(t)\frac{\mathrm{d}t}{\mathrm{d}\mu_r}(t)\right], v\rangle \geqslant 0, \\ \quad \forall\, v\in T_\Phi^{r-1}(\{Z_{r-1}\}(t^-)),\ \mathrm{d}\mu_r - a.e.\ t\in I. \end{cases}$$

The following example and proposition show the link between the distributional formalism and the measure differential formalism. It is a direct consequence of the structure of the models (see also Remark 4).

*Example 7* Suppose that $m = 1$, $r = n = 3$ and let $w_1, w_2, w_3 \in \mathcal{F}_\infty(I;\mathbb{R})$ such that

$$\begin{cases} \mathrm{d}w_1 = w_2\mathrm{d}t + \mathrm{d}\nu_1 \\ \mathrm{d}w_2 = w_3\mathrm{d}t + \mathrm{d}\nu_2 \\ \mathrm{d}w_3 = CA^3W^{-1}w\mathrm{d}t + CA^2B\mathrm{d}\nu_3 \end{cases} \tag{63}$$

where $\mathrm{d}\nu_1, \mathrm{d}\nu_2$ are atomic Radon measures and $\mathrm{d}\nu_3$ is a Radon measure whose Lebesgue–Radon–Nikodym decomposition is given by $\mathrm{d}\nu_3 = \nu'_{3,t}\mathrm{d}t + \mathrm{d}\sigma_3$. Suppose also that $E_0(w_1) = \{a_1\}$, $E_0(w_2) = \{a_2\}$ and $E_0(w_3) = \{a_3\}$, $(a_1, a_2, a_3 \in I)$. We have

$$\begin{cases} \mathrm{d}w_1 = \hat{w}_1^{(1)}\mathrm{d}t + \mathrm{d}\mathcal{J}_1 \\ \mathrm{d}w_2 = \hat{w}_2^{(1)}\mathrm{d}t + \mathrm{d}\mathcal{J}_2 \\ \mathrm{d}w_3 = \hat{w}_3^{(1)}\mathrm{d}t + \mathrm{d}\mathcal{J}_3 \end{cases} \tag{64}$$

where $\mathcal{J}_i(.) \equiv \mathcal{H}_0(. - a_i)$ (see (17)).

Let us now define the distributions $z_1, z_2, z_3 \in \mathcal{T}_\infty(I)$ by:

$$\begin{cases} z_1 = w_1 \\ z_2 = w_2 + \sigma_{w_1}(a_1)\delta_{a_1} \\ z_3 = w_3 + \sigma_{w_2}(a_2)\delta_{a_2} + \sigma_{w_1}(a_1)D\delta_{a_1}. \end{cases} \tag{65}$$

From (63) and (64), we deduce that

$$\begin{cases} \hat{w}_1^{(1)} = w_2 \\ Dz_1 = \hat{w}_1^{(1)} + \sigma_{w_1}(a_1)\delta_{a_1} = w_2 + \sigma_{w_1}(a_1)\delta_{a_1} = z_2 \\ \nu_1 = \int_I \mathrm{d}\nu_1 = \int_I \mathrm{d}\mathcal{J}_1 = \sigma_{w_1}(a_1)\delta_{a_1} \\ \hat{w}_2^{(1)} = w_3 \\ Dz_2 = \hat{w}_2^{(1)} + \sigma_{w_2}(a_2)\delta_{a_2} + \sigma_{w_1}(a_1)D\delta_{a_1} = w_3 + \sigma_{w_2}(a_2)\delta_{a_2} + \sigma_{w_1}(a_1)D\delta_{a_1} = z_3 \\ \nu_2 = \int_I \mathrm{d}\nu_2 = \int_I \mathrm{d}\mathcal{J}_2 = \sigma_{w_2}(a_2)\delta_{a_2} \\ \hat{w}_3^{(1)} = CA^3W^{-1}w + CA^2B\nu'_{3,t} \\ \nu_3 = \int_I \mathrm{d}\nu_3 = (CA^2B)^{-1}\int_I \mathrm{d}\mathcal{J}_3 + \nu'_{3,t} = (CA^2B)^{-1}\sigma_{w_3}(a_3)\delta_{a_3} + \nu'_{3,t}. \end{cases}$$

Then we get

$$\begin{aligned} Dz_3 &= \hat{w}_3^{(1)} + \sigma_{w_3}(a_3)\delta_{a_3} + \sigma_{w_2}(a_2)D\delta_{a_2} + \sigma_{w_1}(a_1)D^{(2)}\delta_{a_1} \\ &= CA^3W^{-1}w + CA^2B(\nu'_{3,t} + (CA^2B)^{-1}\sigma_{w_3}(a_3)\delta_{a_3}) + D\nu_2 + D^{(2)}\nu_1 \\ &= CA^3W^{-1}w + CA^2B\nu_3 + D\nu_2 + D^{(2)}\nu_1 = CA^3W^{-1}w + CA^2B\lambda \end{aligned}$$

with

$$\lambda = (CA^2B)^{-1}(D\nu_2 + D^{(2)}\nu_1) + \nu_3.$$

So, if the measures $d\nu_1, d\nu_2, d\nu_3$ satisfy the extended Moreau's Process in (45), then distributions $z_1, z_2, z_3$ satisfying problem $\mathbf{SP}(z_0; I)$ can be obtained from the functions $w_1, w_2, w_3$ by using the relations in (65).

More generally, we have:

**Proposition 3**

*i) Let $(z_1, \ldots, z_r, \xi) \in (\mathcal{T}_\infty(I))^n$ be a solution of Problem* $\mathbf{SP}$*$(z_0; I)$. Then*

$$deg(z_i) \leqslant i \quad (1 \leqslant i \leqslant r),$$
$$z_1 = \{z_1\} \in \mathcal{F}_\infty(I;\mathbb{R}), \quad \xi = \{\xi\} \in (\mathcal{F}_\infty(I;\mathbb{R}))^{n-r} \cap (C^0(I;\mathbb{R}))^{n-r}$$

*and* $(\{z_1\}, \dots, \{z_r\}, \xi)$ *is a solution of Problem* **MP**$(z_0; I)$.

*ii)* *Let* $(w_1, \dots, w_r, \xi) \in (\mathcal{F}_\infty(I;\mathbb{R}))^n$ *be a solution of Problem* **MP**$(z_0; I)$ *such that, for each* $1 \leqslant i \leqslant r-1$, *the measure* $\mathrm{d}w_i - w_{i+1}\mathrm{d}t$ *is atomic. Let* $z_1, \dots, z_r$ *be defined by*

$$z_1 := w_1$$

*and*

$$z_i := w_i + \sum_{j=1}^{i-1} \left( \sum_{t_k \in E_0(w_j)} (w_j(t_k^+) - w_j(t_k^-))\delta_{t_k}^{(i-j-1)} \right) \quad (2 \leqslant i \leqslant r).$$

*Then* $(z_1, \dots, z_r, \xi) \in (\mathcal{T}_\infty(I))^n$ *and is a solution of Problem SP*$(z_0; I)$.

### 4.3 The state jump mapping

In this section, we examine some properties of the solution of Problem **SP**$(z_0; I)$ (and consequently of problem **MP**$(z_0; I)$). The existence of such a solution is thus assumed for the time being. Let us first remark that the following is true

**Lemma 2** *Let* $z_1, \dots, z_r \in \mathbb{R}$ *be given and let* $\sigma_1, \dots, \sigma_r \in \mathbb{R}$ *such that*

$$\sigma_i \in -\partial\psi_{T_\Phi^{i-1}(Z_{i-1})}(z_i) \quad \textit{for all } \ 1 \leqslant i \leqslant r.$$

*Then the inclusion*

$$\partial\psi_{T_\Phi^{r-1}(Z_{r-1})}(z_r) \subseteq \partial\psi_\Phi(z_1)$$

*holds. Moreover,*

$$\begin{cases} z_1 > 0 \implies \sigma_r = 0 \\ z_1 = 0 \implies \sigma_r \geqslant 0. \end{cases}$$

*Proof* If $z_1 > 0$ then $T_\Phi(z_1) = \mathbb{R}$ and then for all $2 \leqslant i \leqslant r$, we have

$$T_\Phi^{i-1}(Z_{i-1}) = \mathbb{R}.$$

Thus for all $1 \leqslant i \leqslant r$, we obtain

$$\partial\psi_{T_\Phi^{i-1}(Z_{i-1})}(z_i) = \{0\}.$$

In particular we deduce that $\partial\psi_{T_\Phi^{r-1}(Z_{r-1})}(z_r) = \partial\psi_\Phi(z_1)$ and $\sigma_r = 0$.

If $z_1 \leqslant 0$ then $\partial\psi_{\mathbb{R}^+}(z_1) = \mathbb{R}^-$. Assume that $z_1 \leqslant 0, z_2 \leqslant 0, \ldots, z_j \leqslant 0$ and $z_{j+1} > 0$ for some $1 \leqslant j \leqslant r-1$. Then

$$T_\Phi^0(Z_0) = T_\Phi^1(Z_1) = T_\Phi^2(Z_2) = T_\Phi^3(Z_3) = \cdots = T_\Phi^j(Z_j) = \mathbb{R}^+$$

and

$$T_\Phi^{j+1}(Z_{j+1}) = T_{\mathbb{R}^+}(z_{j+1}) = \mathbb{R}.$$

Then

$$T_\Phi^{j+2}(Z_{j+2}) = \cdots = T_\Phi^{r-1}(Z_{r-1}) = \mathbb{R}.$$

Consequently $\partial\psi_{T_\Phi^i(Z_i)}(z_{i+1}) = \mathbb{R}^-$ for all $0 \leqslant i \leqslant j$, whereas $\partial\psi_{T_\Phi^i(Z_i)}(z_{i+1}) = \{0\}$ for $j+1 \leqslant i \leqslant r-1$. We conclude that under such conditions $\sigma_i \geqslant 0$ for all $1 \leqslant i \leqslant j+1$, and $\sigma_i = 0$ for all $j+2 \leqslant i \leqslant r$. Consequently $\sigma_r \geqslant 0$ when $z_1 = 0$. The inclusion is also proved.

Finally, if $z_1, \ldots, z_r \in \mathbb{R}^-$ then

$$T_\Phi^0(Z_0) = T_\Phi^1(Z_1) = \cdots = T_\Phi^{r-1}(Z_{r-1}) = \mathbb{R}^+,$$

and thus $\partial\psi_{T_\Phi^i(Z_i)}(z_{i+1}) = \mathbb{R}^-$ for all $0 \leqslant i \leqslant r-1$, and the conclusion follows. □

**Lemma 3** *Let $z_1, z_2, \ldots, z_r \in \mathbb{R}$ be given. The following inclusion holds*

$$\partial\psi_{T_\Phi^{i-1}(Z_{i-1})}(z_i) \subseteq \partial\psi_{T_\Phi^{i-2}(Z_{i-2})}(z_{i-1}), \tag{66}$$

*for all $2 \leqslant i \leqslant r$.*

*Proof* The result has already been shown in the proof of Lemma 2. □

Let $t \in I$ be given. A direct consequence of the higher order sweeping process appears as soon as one measures the set $\{t\}$ with the Radon measure $\mathrm{d}\nu_i$. Indeed, as seen in Sect. 2, we get from (45):

$$\mathrm{d}\nu_i(\{t\}) \in -\partial\psi_{T_\Phi^{i-1}(\{Z_{i-1}\}(t^-))}(\{z_i\}(t^+)). \tag{67}$$

Then

$$\mathrm{d}\nu_i(\{t\}) = \mathrm{d}\{z_i\}(\{t\}) - \{z_{i+1}\}(t)\mathrm{d}t(\{t\}) = \mathrm{d}\{z_i\}(\{t\}) = \{z_i\}(t^+) - \{z_i\}(t^-)$$

for all $1 \leqslant i \leqslant r-1$, and

$$\begin{aligned}(CA^{r-1}B)\mathrm{d}\nu_r(\{t\}) &= \mathrm{d}\{z_r\}(\{t\}) - CA^rW^{-1}\{z\}(t)\mathrm{d}t(\{t\}) \\ &= \mathrm{d}\{z_r\}(\{t\}) = \{z_r\}(t^+) - \{z_r\}(t^-).\end{aligned}$$

Obviously, the same result holds in the framework of the measure differential formalism (in this case $z_i \equiv \{z_i\}$).

These results ensure that the extended sweeping process inclusion defines a well-posed state jump mapping and the following holds.

**Proposition 4** *Let $z$ be a solution of problem* $\mathbf{SP}(z_0; I)$. *Then, for each $t \in I$, and for all $1 \leqslant i \leqslant r-1$, we have:*

$$\{z_i\}(t^+) - \{z_i\}(t^-) \in -\partial\psi_{T_\Phi^{i-1}(\{Z_{i-1}\}(t^-))}(\{z_i\}(t^+)), \tag{68}$$

*and*

$$\{z_r\}(t^+) - \{z_r\}(t^-) \in -CA^{r-1}B\ \partial\psi_{T_\Phi^{r-1}(\{Z_{r-1}\}(t^-))}(\{z_r\}(t^+)). \tag{69}$$

□

*Remark 13* Note that if $CA^{r-1}B > 0$ then (69) holds if and only if

$$\{z_r\}(t^+) - \{z_r\}(t^-) \in -\partial\psi_{T_\Phi^{r-1}(\{Z_{r-1}\}(t^-))}(\{z_r\}(t^+)) \tag{70}$$

since $T_\Phi^{r-1}(\{Z_{r-1}\}(t^-))$ is a nonempty closed convex cone. The expression in (69) is however essential for the treatment of the multivariable case $m \geqslant 2$ with relative degree $\bar{r}$.

Note that, with our convention, the result of Proposition 4 for $t = 0$ reads

$$\{z_i\}(0^+) - z_{0,i} \in -\partial\psi_{T_\Phi^{i-1}(Z_{0,i-1})}(\{z_i\}(0^+)), \quad (1 \leqslant i \leqslant r-1)$$

and

$$\{z_r\}(0^+) - z_{0,r} \in -CA^{r-1}B\ \partial\psi_{T_\Phi^{r-1}(Z_{0,r-1})}(\{z_r\}(0^+)).$$

where $z_{0,i}$ is the $i$-th component of $z_0$ and $Z_{0,i} = (z_{0,1}, \ldots, z_{0,i})$.

Another direct consequence of (45) is that, for each $t \in I$:

$$\{z_i\}(t^+) \in T_\Phi^{i-1}(\{Z_{i-1}\}(t^-)), \quad (1 \leqslant i \leqslant r). \tag{71}$$

**Theorem 1** *Let $z$ be a solution of problem* $\mathbf{SP}(z_0; I)$. *Then, for each $t \in I$, we have:*

$$0 \leqslant z_1(t^+) \perp \mathrm{d}\nu_r(\{t\}) \geqslant 0 \tag{72}$$

*Proof* Let us first recall (see Remark 4) that $z_1 = \{z_1\}$. Moreover $z_1(t^+) \in \Phi$ and thus $z_1(t^+) \geqslant 0$. From Lemma 2, we obtain that if $z_1(t^+) > 0$ then $\mathrm{d}\nu_r(\{t\}) = 0$ while if $z_1(t^+) = 0$ then $\mathrm{d}\nu_r(\{t\}) \geqslant 0$. We deduce that if $\mathrm{d}\nu_r(\{t\}) > 0$ then necessarily $z_1(t^+) = 0$. The result follows. □

*Remark 14* We may write $\mathrm{d}\nu_i$ as

$$\mathrm{d}\nu_i = \chi_i(t)\mathrm{d}t + \mathrm{d}\mathcal{J}_i, \tag{73}$$

where $\chi_i \in \mathcal{F}_\infty(I;\mathbb{R})$ and $\mathrm{d}\mathcal{J}_i$ is an atomic measure with countable set of atoms generated by a right continuous jump function $\mathcal{J}_i$. Let $1 \leqslant i \leqslant r-1$ be given. We know that $Dz_i = z_{i+1}$ and thus $\{Dz_i\} = \{z_{i+1}\}$. Thus $\nu_i = \langle\langle Dz_i - \{Dz_i\}\rangle\rangle$. It results that $\mathrm{d}\nu_i$ is an atomic measure and thus

$$\chi_i(t) = 0, \quad a.e.\ t \in I, \ (1 \leqslant i \leqslant r-1). \tag{74}$$

The condition in (45) on $\mathrm{d}\nu_r$ implies that (see Proposition 2):

$$\chi_r(t) \in -\partial\psi_{T_\Phi^{r-1}(\{Z_{r-1}\}(t^-))}(\{z_r\}(t^+)), \quad a.e.\ t \in I. \tag{75}$$

Using (75) and Lemma 2 we get also

$$0 \leqslant z_1(t^+) \perp \chi_r(t) \geqslant 0, \quad a.e.\ t \in I. \tag{76}$$

The following result gives some equivalent characterizations of the relations given in Proposition 4 and Remark 13. The proof is straightforward.

**Proposition 5** *Let $z$ be a solution of problem* $\mathbf{SP}(z_0;\ I)$*. Then, for each $t \in I$ and for all $1 \leqslant i \leqslant r-1$, we have:*

$$\{z_i\}(t^+) - \{z_i\}(t^-) \in -\partial\psi_{T_\Phi^{i-1}(\{Z_{i-1}\}(t^-))}(\{z_i\}(t^+))$$

*if and only if*

$$\{z_i\}(t^+) = prox\left[T_\Phi^{i-1}(\{Z_{i-1}\}(t^-)); \{z_i\}(t^-)\right].$$

*If $CA^{r-1}B > 0$ then*

$$\{z_r\}(t^+) - \{z_r\}(t^-) \in -CA^{r-1}B\ \partial\psi_{T_\Phi^{r-1}(\{Z_{r-1}\}(t^-))}(\{z_r\}(t^+))$$

*if and only if*

$$\{z_r\}(t^+) = prox\left[T_\Phi^{r-1}(\{Z_{r-1}\}(t^-)); \{z_r\}(t^-)\right].$$

□

This shows that "jumps" are automatically taken into account by the dynamics as it is written in (51) (or in (57)–(58)). We notice also that the lower triangular structure of the tangent cones which appear in (45) merely reflects the way the measures $d\nu_i$ appear in (51) (or in (57)–(58)).

*Remark 15*

i) In the multivariable case $m \geqslant 2$ with vector relative degree $\bar{r}$, if the matrix $CA^{r-1}B$ is symmetric and positive definite then the second equivalence in Proposition 5 can be generalized as follows:

$$\{z_r\}(t^+) - \{z_r\}(t^-) \in -CA^{r-1}B\; \partial\psi_{T_\Phi^{r-1}(\{Z_{r-1}\}(t^-))}(\{z_r\}(t^+))$$

if and only if

$$\{z_r\}(t^+) = \mathrm{prox}_{(CA^{r-1}B)^{-1}}\left[T_\Phi^{r-1}(\{Z_{r-1}\}(t^-)); \{z_r\}(t^-)\right].$$

ii) If $(CA^{r-1}B)^{-1} > 0$ then the initial conditions in (52) can be written as follows:

$$\begin{cases} \{z_1\}(0^+) = \mathrm{prox}\left[\mathbb{R}^+; z_{0,1}\right], \\ \{z_2\}(0^+) = \mathrm{prox}\left[T_\Phi^1(Z_{0,1}); z_{0,2}\right], \\ \quad\vdots \\ \{z_i\}(0^+) = \mathrm{prox}\left[T_\Phi^{i-1}(Z_{0,i-1}); z_{0,i}\right], \\ \quad\vdots \\ \{z_{r-1}\}(0^+) = \mathrm{prox}\left[T_\Phi^{r-1}(Z_{0,r-1}); z_{0,r-1}\right], \\ \{z_r\}(0^+) = \mathrm{prox}\left[T_\Phi^{r-1}(Z_{0,r-1}); z_{0,r}\right]. \end{cases} \tag{77}$$

In the multivariable case $m \geqslant 2$ with vector relative degree $\bar{r}$, if the matrix $CA^{r-1}B$ is symmetric and positive definite, then the last relation in (77) must be written as follows:

$$\{z_r\}(0^+) = \mathrm{prox}_{(CA^{r-1}B)^{-1}}\left[T_\Phi^{r-1}(Z_{0,r-1}); z_{0,r}\right].$$

iii) We note that from Proposition 5, the measures $\mathrm{d}\nu_i$ are uniquely determined at jump instants. The formalism in (45) corresponds to some kind of "plastic" impacts since the high-order inconsistent derivatives jump to zero. It is quite possible to incorporate different jumps by modifying the right-hand sides and changing the argument in the subdifferentials from $\{z_i\}(t^+)$ to $\frac{\{z_i\}(t^+)+e_i\{z_i\}(t^-)}{1+e_i}$ [40,51]. Then

$$\{z_i\}(t^+) - \{z_i\}(t^-) \in -\partial\psi_{T_\Phi^{i-1}(\{Z_{i-1}\}(t^-))}\left(\frac{\{z_i\}(t^+) + e_i\{z_i\}(t^-)}{1+e_i}\right)$$

if and only if

$$\{z_i\}(t^+) = -e_i\{z_i\}(t^-) + (1+e_i)\text{prox}\left[T_\Phi^{i-1}(\{Z_{i-1}\}(t^-)); \{z_i\}(t^-)\right].$$

The choice of the coefficients $e_i$ depends on the application. It follows that if $T_\Phi^i(\{Z_i\}(t^-)) = (\mathbb{R}^+)^m$ and if $\{z_{i+1}\}(t^-) < 0$, then $\{z_{i+1}\}(t^+) = -e_{i+1}\{z_{i+1}\}(t^-)$.

## 4.4 Qualitative properties

In this section, we state some important qualitative properties of a (possible) solution of problem $\mathbf{SP}(z_0;\ I)$.

Let $0 < T \leqslant +\infty$ be given. Recall that we set $I = [0, T[$. Let us here also set

$$\Lambda := |||WAW^{-1}||| = \sup_{x\in\mathbb{R}^n, x\neq 0} \frac{\|WAW^{-1}x\|}{\|x\|}. \tag{78}$$

**Theorem 2** *Suppose that $CA^{r-1}B > 0$. Let $z$ be a solution of problem $SP(z_0;\ I)$. Then*

*i)* $\|\{z\}(t)\| \leqslant \|z_0\|e^{\Lambda t}, \ \ \forall\, t \in I.$
*ii)* *If $T < +\infty$ then* $\text{var}(\{z\}, I) < +\infty$.
*iii)* *If $T < +\infty$ then, for all $1 \leqslant i \leqslant n$, $\{z_i\}(T^-)$ exists and is finite.*

*Proof* i) We may write

$$\mathrm{d}\{z\} = WAW^{-1}\{z\}\mathrm{d}t + N\mathrm{d}\nu$$

where

$$\mathrm{d}\nu := (\mathrm{d}\nu_1, \ldots, \mathrm{d}\nu_{r-1}, \mathrm{d}\nu_r, 0_{n-r})^{\mathrm{T}}$$

and $N$ is the $n \times n$ diagonal matrix:

$$N = \begin{pmatrix} I_{r-1} & 0^{r-1} & 0_{(r-1)\times(n-r)} \\ 0_{r-1} & CA^{r-1}B & 0_{n-r} \\ 0_{(n-r)\times(r-1)} & 0^{n-r} & I_{n-r} \end{pmatrix}. \tag{79}$$

Thus

$$(\{z\}^+)^{\mathrm{T}}\mathrm{d}\{z\} = (\{z\}^+)^{\mathrm{T}}(WAW^{-1}\{z\}^+)\mathrm{d}t + (\{z\}^+)^{\mathrm{T}}N\mathrm{d}\nu.$$

Here

$$(\{z\}^+)^{\mathrm{T}}N\mathrm{d}\nu = \sum_{i=1}^{r-1}\{z_i\}^+\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\mathrm{d}\mu_i + CA^{r-1}B\frac{\mathrm{d}\nu_r}{\mathrm{d}\mu_r}\mathrm{d}\mu_r$$

for some non-negative real-valued Radon measures $\mathrm{d}\mu_i$ $(1 \leqslant i \leqslant r)$ (see (46)). We have $(\{z\}^+)^{\mathrm{T}}\mathrm{d}\nu \equiv 0$ on $[0, T[$ since (see (47)) $\{z_i\}(t^+)\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}(t) = 0$ for $\mathrm{d}\mu_i - a.e.$ $t \in [0, T[$. Consequently

$$(\{z\}^+)^{\mathrm{T}}\mathrm{d}\{z\} = (\{z\}^+)^{\mathrm{T}}(WAW^{-1}\{z\})\mathrm{d}t.$$

Then using (10) and recalling that $\{z\}$ is right continuous, we get

$$\mathrm{d}(\{z\}^{\mathrm{T}}\{z\}) \leqslant 2(\{z\})^{\mathrm{T}}(WAW^{-1}\{z\})\mathrm{d}t.$$

Thus, for any $t \in [0, T[$, we get

$$\mathrm{d}(\{z\}^{\mathrm{T}}\{z\})([0,t]) \leqslant 2\int_0^t (\{z\}(s))^{\mathrm{T}}(WAW^{-1}\{z\})(s)\mathrm{d}s$$

so that

$$\|\{z\}(t)\|^2 \leqslant \|z_0\|^2 + 2\Lambda\int_0^t \|\{z\}(s)\|^2\mathrm{d}s.$$

From Lemma 2, we get

$$\|\{z\}(t)\|^2 \leqslant \|z_0\|^2\mathrm{e}^{2\Lambda t}$$

and the result follows.

ii) Here $T < +\infty$. Let $1 \leqslant i \leqslant n$ be given. We have

$$\mathrm{d}\nu_i = \frac{1}{N_{ii}}(-(WAW^{-1})_i\{z\}\mathrm{d}t + \mathrm{d}\{z_i\})$$

where $N_{ii} = 1$ for $i \neq r$ and $N_{rr} = CA^{r-1}B$, and thus, for all $0 \leqslant s_1 < s_2 < T$, we get

$$\begin{aligned}\mathrm{d}\nu_i(]s_1, s_2]) &= \frac{1}{N_{ii}}\left(-\int_{s_1}^{s_2}(WAW^{-1})_i\{z\}(s)\mathrm{d}s + \{z_i\}(s_2) - \{z_i\}(s_1)\right)\\ &\leqslant \frac{\Lambda}{N_{ii}}\int_{s_1}^{s_2}\|\{z\}(s)\|\mathrm{d}s + \frac{1}{N_{ii}}(|\{z_i\}(s_2)| + |\{z_i\}(s_1)|).\end{aligned}$$

Thus

$$\mathrm{d}\nu_i(]s_1, s_2]) \leqslant \frac{\Lambda}{N_{ii}}\|z_0\|\mathrm{e}^{\Lambda T}(s_2 - s_1) + \frac{2}{N_{ii}}\|z_0\|\mathrm{e}^{\Lambda T} \leqslant \frac{(2 + \Lambda T)}{N_{ii}}\|z_0\|\mathrm{e}^{\Lambda T}.$$

The measure $\mathrm{d}\nu_i$ is nonnegative and thus (see e.g. Theorem 5.9.4 in [65]):

$$|\mathrm{d}\nu_i|([0,T[) = \sup_{\mathcal{A}\in\mathcal{B}(\mathbb{R}),\mathcal{A}\subset[0,T[} \mathrm{d}\nu_i(\mathcal{A}).$$

Then

$$|\mathrm{d}\nu_i|([0,T[) \leqslant \frac{(2+\Lambda T)}{N_{ii}}\|z_0\|\mathrm{e}^{\Lambda T}.$$

Moreover (see e.g. Theorem 5.10.23 in [65]):

$$|(WAW^{-1})_i\{z\}\mathrm{d}t|([0,T[) = \int_0^T |(WAW^{-1})_i\{z\}(s)|\mathrm{d}s \leqslant \Lambda T\|z_0\|\mathrm{e}^{\Lambda T}.$$

Then (see (11))

$$\begin{aligned}\mathrm{var}(\{z_i\},[0,T[) = |\mathrm{d}\{z_i\}|([0,T[) &\leqslant |(WAW^{-1})_i\{z\}\mathrm{d}t|([0,T[) + |N_{ii}||\mathrm{d}\nu_i|([0,T[)\\ &\leqslant 2(1+\Lambda T)\|z_0\|\mathrm{e}^{\Lambda T} < +\infty\end{aligned}$$

and the result follows.
iii) Here $T < +\infty$. Let $1 \leqslant i \leqslant n$ be given. From (ii), we have

$$K_i := \mathrm{var}(\{z_i\},[0,T[) < +\infty.$$

We claim that the limit

$$\{z_i\}(T^-) = \lim_{t\to T, t<T}\{z_i\}(t)$$

exists and is finite. Indeed, we may write

$$\{z_i\}(t) = \Phi_i(t) - \Psi_i(t)$$

where

$$\Phi_i(t) = \mathrm{var}(\{z_i\},[0,t]), \quad \Psi_i(t) = \mathrm{var}(\{z_i\},[0,t]) - \{z_i\}(t).$$

The mappings $t \mapsto \Phi_i(t)$ and $t \mapsto \Psi_i(t)$ are monotone nondecreasing and from (ii), we get

$$\Phi_i(t) \leqslant K_i, \quad \forall\, t \in [0,T[.$$

Using (i), we get also

$$\Psi_i(t) \leqslant K_i + \mathrm{e}^{\Lambda T}\|z_0\|, \quad \forall\, t \in [0,T[.$$

Thus

$$\lim_{t\to T, t<T} \Phi_i(t) = \sup_{t\in[0,T[} \Phi_i(t) \leqslant K_i$$

and

$$\lim_{t\to T, t<T} \Psi_i(t) = \sup_{t\in[0,T[} \Psi_i(t) \leqslant K_i + \mathrm{e}^{\Lambda T}\|z_0\|.$$

It results that $\{z_i\}(T^-)$ exists and is finite. □

**Proposition 6** *Suppose that $CA^{r-1}B > 0$ and let $z$ be a solution of problem* $\mathbf{SP}(z_0;\ I)$. *Let $t \in I$ be given. Then:*

*i)* *If $\{z_1\}(t^-) > 0$ then for all $1 \leqslant i \leqslant r$, we have*

$$\{z_i\}(t^+) = \{z_i\}(t^-).$$

*ii)* *If for some $1 \leqslant j \leqslant r$, $\{z_i\}(t^-) \leqslant 0$ for all $1 \leqslant i \leqslant j$ and $\{z_{j+1}\}(t^-) > 0$, then*

$$1 \leqslant i \leqslant j \ \Rightarrow \ \{z_i\}(t^+) = 0$$

*and*

$$j+1 \leqslant i \leqslant r \ \Rightarrow \ \{z_i\}(t^+) = \{z_i\}(t^-).$$

*iii)* *If $\{z_i\}(t^-) \leqslant 0$ for all $1 \leqslant i \leqslant r$ then, for all $1 \leqslant i \leqslant r$, we have*

$$\{z_i\}(t^+) = 0.$$

*Proof* The values of the cones in the inclusions in (45) have been computed in the proof of Lemma 2. The result follows from Propositions 5 and (45).

i) If $\{z_1\}(t^-) > 0$ then

$$\{z_1\}(t^+) = \operatorname{prox}\left[\mathbb{R}^+; \{z_1\}(t^-)\right] = \{z_1\}(t^-).$$

Then

$$\{z_2\}(t^+) = \operatorname{prox}\left[T_{\mathbb{R}^+}(\{z_1\}(t^-)); \{z_2\}(t^-)\right] = \operatorname{prox}\left[\mathbb{R}; \{z_2\}(t^-)\right] = \{z_2\}(t^-)$$

and then, for all $3 \leqslant i \leqslant r$, we get

$$\{z_i\}(t^+) = \operatorname{prox}\left[\mathbb{R}; \{z_i\}(t^-)\right] = \{z_i\}(t^-).$$

ii) Here we have, for all $1 \leqslant i \leqslant j$,

$$\{z_i\}(t^+) = \mathrm{prox}\left[\mathbb{R}^+; \{z_i\}(t^-)\right] = 0,$$
$$\{z_{j+1}\}(t^+) = \mathrm{prox}\left[\mathbb{R}^+; \{z_{j+1}\}(t^-)\right] = \{z_{j+1}\}(t^-).$$

and, for all $j+2 \leqslant i \leqslant r$,

$$\{z_i\}(t^+) = \mathrm{prox}\left[\mathbb{R}; \{z_i\}(t^-)\right] = \{z_i\}(t^-).$$

iii) Here we have, for all $1 \leqslant i \leqslant r$,

$$\{z_i\}(t^+) = \mathrm{prox}\left[\mathbb{R}^+; \{z_i\}(t^-)\right] = 0.$$

□

*Remark 16* i) Proposition 6 entails that if $\{z_j\}(t^-) > 0$ for some $1 \leqslant j \leqslant r$ then the RCSLBV mappings $\{z_j\}, \{z_{j+1}\}, \ldots, \{z_r\}$ are continuous at $t$. ii) A consequence of Proposition 6 is that

$$\{\bar{z}\}(t^+) \succeq 0, \quad \forall\, t \in I.$$

and

$$\|\{z\}(t^+)\| \leqslant \|\{z\}(t^-)\|, \quad \forall\, t \in I.$$

**Proposition 7** *Suppose that $CA^{r-1}B > 0$ and let $z$ be a solution of Problem* **SP**$(z_0;\ I)$. *Let $t \in I$ be given. If $\{\bar{z}\}(t^-) \succeq 0$ then $\{\bar{z}\}(t^+) = \{\bar{z}\}(t^-)$.*

*Proof* Let $1 \leqslant j \leqslant r$ such that $\{z_1\}(t^-) = \cdots = \{z_{j-1}\}(t^-) = 0$ and $\{z_j\}(t^-) > 0$. Then the result follows from part (ii) of Proposition 6. □

*Remark 17* The results of this section can be generalized to the multivariable case $m \geqslant 2$ with vector relative degree $\bar{r}$ provided that one supposes that $CA^{r-1}B$ is a positive definite and symmetric matrix. Note that in this, case, we have for any $t \in I$:

$$\{z_r\}(t^+) = \mathrm{prox}_{(CA^{r-1}B)^{-1}}\left[T_{\Phi}^{r-1}(\{Z_{r-1}\}(t^-)); \{z_r\}(t^-)\right]$$

and the norm

$$\|z\| := \sqrt{\sum_{\alpha=1}^{r-1} \|z_\alpha\|_m^2 + \langle CA^{r-1}Bz_r, z_r\rangle + \|\xi\|_{n-mr}^2}$$

must be utilized to prove these results.

## 4.5 Dissipativity and stability properties

Monotonicity and dissipativity are at the core of the second order sweeping process. They are crucial properties for well-posedness results [4,7] and control/stability [10,17,31,13]. The following results are consequently of interest.

We first investigate the dissipativity properties of the differential operator of the sweeping process.

Let us define the matrices $G \in \mathbb{R}^{r\times r}, \bar{G} \in \mathbb{R}^{n\times r}$ and $H \in \mathbb{R}^{r\times n}$ as follows:

$$\bar{G} = \begin{pmatrix} G \\ 0_{(n-r)\times r} \end{pmatrix}, \tag{80}$$

with

$$G = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \ldots & 0 & 1 & 0 \\ & & & & \\ 0 & \ldots & \ldots & 0 & CA^{r-1}B \end{pmatrix} \tag{81}$$

and

$$H = (I_r \;\; 0_{r\times(n-r)})..$$

Let the triple $(WAW^{-1}, \bar{G}, H)$ be observable, controllable and positive real. The positive realness of $(WAW^{-1}, \bar{G}, H)$ is then equivalent by the Kalman–Yakubovich–Popov Lemma (see Appendix B) to having $J\bar{G} = H^{\mathrm{T}}$ and $JWAW^{-1} + W^{-\mathrm{T}}A^{\mathrm{T}}W^{\mathrm{T}}J$ semi-negative definite for some positive definite and symmetric matrix $J$. Since $J = J^{\mathrm{T}}$ and since $G$ is full rank, the first equality implies that

$$J = \begin{pmatrix} G^{-1} & 0_{r\times(n-r)} \\ & \\ 0_{(n-r)\times r} & J_\xi \end{pmatrix}, \tag{82}$$

and $J_\xi = J_\xi^T \in \mathbb{R}^{(n-r)\times(n-r)}$ is positive definite.

**Proposition 8** (Dissipation inequality) *Let $z$ be a solution of Problem* $\mathbf{SP}(z_0; I)$. *Let the triple $(WAW^{-1}, \bar{G}, H)$ be observable, controllable and positive real. Then $CA^{r-1}B > 0$ and for all $t_1, t_2 \in I$, $t_1 \leqslant t_2$, we have*

$$\{z\}(t_2^+)^{\mathrm{T}} J \{z\}(t_2^+) \leqslant \{z\}(t_1^-)^{\mathrm{T}} J \{z\}(t_1^-), \tag{83}$$

*where $J$ is given in (82).*

*Proof* The first statement of the proposition is a direct consequence of the positive definiteness of the solution $J$ of the equalities in (169).

Let us now set

$$\mathrm{d}\nu := (\mathrm{d}\nu_1, \ldots, \mathrm{d}\nu_r, 0_{1\times(n-r)})^{\mathrm{T}}.$$

We have

$$\mathrm{d}\{z\} = WAW^{-1}\{z\}\mathrm{d}t + \bar{G}\mathrm{d}\nu = WAW^{-1}\{z\}^{+}\mathrm{d}t + \bar{G}\mathrm{d}\nu \tag{84}$$

and therefore

$$J\mathrm{d}\{z\} = JWAW^{-1}\{z\}^{+}\mathrm{d}t + J\bar{G}\mathrm{d}\nu = JWAW^{-1}\{z\}^{+}\mathrm{d}t + H^{\mathrm{T}}\mathrm{d}\nu.$$

It results that

$$(\{z\}^{+})^{\mathrm{T}}J\mathrm{d}\{z\} = (\{z\}^{+})^{\mathrm{T}}(JWAW^{-1}\{z\}^{+})\mathrm{d}t + (\{z\}^{+})^{\mathrm{T}}H^{\mathrm{T}}\mathrm{d}\nu.$$

Here $(\{z\}^{+})^{\mathrm{T}}H^{\mathrm{T}}\mathrm{d}\nu = \sum_{i=1}^{r}\{z_i\}^{+}\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\mathrm{d}\mu_i$ for some nonnegative real-valued Radon measures $\mathrm{d}\mu_i$ $(1 \leqslant i \leqslant r)$ and $(\{z\}^{+})^{T}H^{\mathrm{T}}\mathrm{d}\nu \equiv 0$ on $I$ (see (46) and (47)). Consequently

$$(\{z\}^{+})^{\mathrm{T}}J\mathrm{d}\{z\} = (\{z\}^{+})^{\mathrm{T}}(JWAW^{-1}\{z\}^{+})\mathrm{d}t.$$

We have

$$\mathrm{d}(\{z\}^{\mathrm{T}}J\{z\}) = \sum_{i=1}^{r-1}\mathrm{d}\{z_i\}^2 + (CA^{r-1}B)^{-1}\mathrm{d}\{z_r\}^2 + \mathrm{d}(\xi^{\mathrm{T}}J_{\xi}\xi)$$

and then using (10) and Remark 4 (ii) and recalling that here $CA^{r-1}B > 0$, $J_{\xi} = J_{\xi}^{\mathrm{T}}, J_{\xi}$ is positive definite, we see that

$$\begin{aligned}\mathrm{d}(\{z\}^{\mathrm{T}}J\{z\}) &\leqslant \sum_{i=1}^{r-1}2\{z_i\}^{+}\mathrm{d}\{z_i\} + 2(CA^{r-1}B)^{-1}\{z_r\}^{+}\mathrm{d}\{z_r\} + 2\xi^{\mathrm{T}}J_{\xi}\mathrm{d}\xi \\ &= 2(\{z\}^{+})^{\mathrm{T}}J\mathrm{d}\{z\}\end{aligned}$$

and thus

$$\begin{aligned}d(\{z\}^{\mathrm{T}}J\{z\}) \leqslant 2(\{z\}^{+})^{\mathrm{T}}J\mathrm{d}\{z\} &= 2(\{z\}^{+})^{\mathrm{T}}(JWAW^{-1}\{z\}^{+})\mathrm{d}t \\ &= (\{z\}^{+})^{\mathrm{T}}(JWAW^{-1} + W^{-\mathrm{T}}A^{\mathrm{T}}W^{\mathrm{T}}J)\{z\}^{+}\mathrm{d}t.\end{aligned}$$

Then for $0 \leqslant t_1 \leqslant t_2$, we get

$$\{z\}^{\mathrm{T}}(t_2^+)J\{z\}(t_2^+) - \{z\}(t_1^-)^{\mathrm{T}}J\{z\}(t_1^-) = d(\{z\}^{\mathrm{T}}J\{z\})([t_1,t_2])$$
$$= \int_{t_1}^{t_2} \{z\}^+(s)^{\mathrm{T}}(JWAW^{-1} + W^{-\mathrm{T}}A^{\mathrm{T}}W^{\mathrm{T}}J)\{z\}^+(s)\mathrm{d}s \leqslant 0.$$

□

Proposition 8 and its proof are similar to [4, Proposition 7] and [10, Lemma 3] (see also [17, Theorem 11.2]), which apply to Lagrangian systems or to the case $r = 1$.

Let us now suppose that the matrix $WAW^{-1}$ is Hurwitz (i.e. all the eigenvalues of $WAW^{-1}$ have strictly negative real part), the triple $(WAW^{-1}, \bar{G}, H)$ is observable, controllable and strictly positive real. Then $J\bar{G} = H^{\mathrm{T}}$ and $JWAW^{-1} + W^{-\mathrm{T}}A^{\mathrm{T}}W^{\mathrm{T}}J$ is negative definite for some positive definite and symmetric matrix $J$ (see Appendix B). Let us now denote by $\lambda_{\max} < 0$ the largest eigenvalue of the matrix $JWAW^{-1} + W^{-\mathrm{T}}A^{\mathrm{T}}W^{\mathrm{T}}J$.

**Proposition 9** (Strong dissipation inequality) *Let $z$ be a solution of Problem* $\mathbf{SP}$*($z_0$; $I$). Let $WAW^{-1}$ be Hurwitz, the triple $(WAW^{-1}, \bar{G}, H)$ be observable, controllable and strictly positive real. Then $CA^{r-1}B > 0$ and for all $t_1, t_2 \in I$, $t_1 \leqslant t_2$, we have*

$$\{z\}(t_2^+)^{\mathrm{T}}J\{z\}(t_2^+) \leqslant \{z\}(t_1^-)^{\mathrm{T}}J\{z\}(t_1^-) - |\lambda_{\max}| \int_{t_1}^{t_2} \|\{z\}^+(s)\|^2\mathrm{d}s, \qquad (85)$$

*where $J$ is given in (82).*

*Proof* The proof follows the same steps as in Proposition 8 until we get, for all $0 \leqslant t_1 \leqslant t_2$, that:

$$\{z\}^{\mathrm{T}}(t_2^+)J\{z\}(t_2^+) - \{z\}(t_1^-)^{\mathrm{T}}J\{z\}(t_1^-)$$
$$= \int_{t_1}^{t_2} \{z\}^+(s)^{\mathrm{T}}(JWAW^{-1} + W^{-\mathrm{T}}A^{\mathrm{T}}W^{\mathrm{T}}J)\{z\}^+(s)\mathrm{d}s.$$

Here

$$x^{\mathrm{T}}(JWAW^{-1} + W^{-\mathrm{T}}A^{\mathrm{T}}W^{\mathrm{T}}J)x \leqslant -|\lambda_{\max}|\|x\|^2, \quad \forall\, x \in \mathbb{R}^n,$$

and the result follows. □

For $x \in \mathbb{R}^n$, let us here, for the following result, use the norm:

$$\|x\|_J := \sqrt{x^{\mathrm{T}}Jx}.$$

**Proposition 10** (Stability and Attractivity Result) *i) The trivial solution* 0 *is the unique solution of problem* **SP**$(0; [0,+\infty[)$.

*ii) Let the triple* $(WAW^{-1}, \bar{G}, H)$ *be observable, controllable and positive real. Then the trivial solution* 0 *is stable in the sense that for each* $\varepsilon > 0$*, there exists* $\delta > 0$ *such that for any* $z_0 \in \mathbb{R}^n$, $\|z_0\|_J \leqslant \delta$*, any solution* $z$ *of of problem* **SP**$(z_0; [0,T[)$ $(0 \leqslant T \leqslant +\infty)$ *satisfies* $\|\{z\}(t)\|_J \leqslant \varepsilon$ *for all* $t \in [0,T[$.

*iii) Let* $WAW^{-1}$ *be Hurwitz and let the triple* $(WAW^{-1}, \bar{G}, H)$ *be observable, controllable and strictly positive real. Then the trivial solution* 0 *is stable and globally attractive in the sense that for each* $z_0 \in \mathbb{R}^n$*, any solution* $z$ *of problem* **SP**$(z_0; [0,+\infty[)$ *fulfills*

$$\lim_{t\to+\infty} \|\{z\}(t)\|_J = 0.$$

*Proof* i) It is clear that 0 is a solution of **SP**$(0; [0,+\infty[)$. The uniqueness is a direct consequence of part i) in Theorem 2. iii) It suffices to choose $\delta = \varepsilon$ and to apply Proposition 8. iii) The stability result follows from ii). It remains to check the attractivity result. Let $z$ be a solution problem SP$(z_0; [0,+\infty[)$. From Proposition 8, the mapping $t \mapsto \|\{z\}(t)\|_J$ is non-increasing and the limit $\lim_{t\to+\infty} \|\{z\}(t)\|_J$ exists and is finite. Set

$$a = \lim_{t\to+\infty} \|\{z\}(t)\|_J.$$

It is clear that $a \geqslant 0$. We claim that $a = 0$. Indeed, suppose on the contrary that $a > 0$. Then, for all $t \geqslant 0$, we have $\|\{z\}(t)\|_J \geqslant a$ and thus, from Proposition 9, we obtain that, for all $t \geqslant 0$:

$$0 \leqslant \|\{z\}(t)\|_J^2 \leqslant \|z_0\|_J^2 - |\lambda_{\max}| \frac{a^2}{\||J|\|} t$$

which gives a contradiction for $t > \frac{\||J|\| \|z_0\|_J^2}{|\lambda_{\max}| a^2}$. □

*Remark 18* The results of this Section can easily be generalized to the multivariable case $m \geqslant 2$ with vector relative degree $\bar{r}$.

## 4.6 Existence, uniqueness and regularity results

Let us first recall that for $z \in \mathbb{R}^n$, we use the notation $z^{\mathrm{T}} = (\bar{z}^{\mathrm{T}}, \xi^{\mathrm{T}})$ as in (26) with $\bar{z} \in \mathbb{R}^r$ and $\xi \in \mathbb{R}^{n-r}$.

Let $z_0^{\mathrm{T}} = (\bar{z_0}^{\mathrm{T}}, \xi_0^{\mathrm{T}})$ be given. In this section, we discuss the existence and uniqueness of a solution for problem **SP**$(z_0; I)$.

**Definition 3** *Let* $0 \leqslant a < b \leqslant T \leqslant +\infty$ *be given. We say that a solution* $z \in (\mathcal{T}_\infty([0,T[))^n$ *of problem* **SP**$(z_0; [0,T[)$ *is regular on* $[a,b[$ *if for each* $t \in [a,b[$*, there exists a right neighborhood* $[t, t+\sigma[$ $(\sigma > 0)$ *such that the restriction of* $\{z\}$ *to* $[t, t+\sigma[$ *is analytic.*

**Lemma 4** *Let* $0 < T \leqslant +\infty$ *be given. Suppose that* $CA^{r-1}B > 0$ *and* $\bar{z}_0 \succeq 0, \bar{z}_0 \neq 0$. *If a solution* $z$ *of problem* $\mathbf{SP}(z_0;\ [0,T[)$ *exists, then there exists* $0 < \eta \leqslant T$ *such that* $z \equiv \{z\}$ *is analytic on* $[0,\eta[$ *and* $z_1(t) > 0, \forall t \in ]0,\eta[$.

*Proof* Suppose that Problem $\mathbf{SP}(z_0;\ I)$ has a solution $z^{\mathrm{T}} = (\bar{z}^{\mathrm{T}}, \xi^{\mathrm{T}})$. Using Proposition 7, we first remark that $\{\bar{z}\}(0) = \bar{z}_0$ and since $\bar{z}_0 \succeq 0, \bar{z}_0 \neq 0$, there exists $1 \leqslant \alpha \leqslant r$ such that $\{z_\alpha\}(0) > 0$ and $\{z_k\}(0) = 0$ for all $1 \leqslant k \leqslant \alpha - 1$ (if $\alpha \neq 1$). We claim that there exists $\eta_0 > 0$ such that $\{z_\alpha\}$ is continuous on $[0,\eta_0]$. Indeed, suppose on the contrary that for each $\eta_0 > 0$ there exists a point $\bar{t} \in [0,\eta_0]$ such that $\{z_\alpha\}(\bar{t}^+) \neq \{z_\alpha\}(\bar{t}^-)$. Then we may find a sequence of points $\{t_i\}_{i \in \mathbb{N}}$ such that $t_i \to 0^+$ and $\{z_\alpha\}(t_i^+) \neq \{z_\alpha\}(t_i^-)$. Then using (68) and (69), we deduce that necessarily $\{z_\alpha\}(t_i) = \{z_\alpha\}(t_i^+) \leqslant 0, \forall i \in \mathbb{N}$. The function $\{z_\alpha\}$ is right-continuous and thus $\{z_\alpha\}(0) = \lim_{i \to \infty}\{z_\alpha\}(t_i) \leqslant 0$. This is a contradiction since $\{z_\alpha\}(0) > 0$. Thus $\{z_\alpha\} \equiv z_\alpha$ is continuous on $[0,\eta_0]$. From the chain of integrators in (34) we see also that the functions $\{z_k\} \equiv z_k$ are continuous on $[0,\eta_0]$, for all $1 \leqslant k \leqslant \alpha - 1$ (if $\alpha \neq 1$). Therefore there exists $\eta \in ]0,\eta_0[$ such that

$$z_\alpha(t) > 0, \quad \forall\, t \in [0,\eta[.$$

Moreover, for any $1 \leqslant k \leqslant \alpha - 1$ (if $\alpha \neq 1$), we have also

$$z_k(t) > 0, \quad \forall\, t \in ]0,\eta[$$

since

$$z_k(t) = z_k(0) + \int_0^t z_{k+1}(\tau)\mathrm{d}\tau = \int_0^t z_{k+1}(\tau)\mathrm{d}\tau.$$

In particular, $z_1(t) > 0, \forall\ t \in ]0,\eta[$ and thus $N_{\mathbb{R}^+}(z_1(t)) = \{0\},\ \forall\ t \in ]0,\eta[$ and $T^i_\Phi(Z_i(t)) = \mathbb{R}$ for all $1 \leqslant i \leqslant r-1$ and all $t \in ]0,\eta[$. Recalling that $(CA^{r-1}B)^{-1} \neq 0$, we see that on $[0,\eta[,\ z \equiv \{z\}$ is continuous and is a solution of the ODE:

$$\begin{cases} \dot{z}_1 = z_2 \\ \dot{z}_j = z_{j+1}\ \ (1 \leqslant j \leqslant r-1) \\ \dot{z}_r = CA^r W^{-1} z \\ \dot{\xi} = A_\xi \xi + B_\xi z_1 \\ z(0) = z_0. \end{cases} \tag{86}$$

Thus

$$z(t) = W\mathrm{e}^{At}W^{-1}z_0, \quad \forall\, t \in [0,\eta[$$

and the result follows. □

Let us here recall that we set $\Lambda := |||WAW^{-1}|||$ as defined in (78).

**Theorem 3** *Suppose that $CA^{r-1}B > 0$. If $\bar{z}_0 \succeq 0, \bar{z}_0 \neq 0$, then:*

i) *(Local existence) There exists $T > 0$ such t-hat Problem $\mathbf{SP}(z_0; [0, T[)$ has at least one regular solution $z \equiv \{z\}$ given by*

$$z(t) = W e^{At} W^{-1} z_0, \quad \forall\, t \in [0, T[;$$

ii) $z_1(t) > 0, \quad \forall\, t \in ]0, T[;$

iii) $\|z(t)\| \leqslant e^{\Lambda t}\|z_0\|, \quad \forall\, t \in [0, T[;$

iv) $\mathrm{var}(z, [0, t]) \leqslant \|z_0\|\Lambda e^{\Lambda t} t, \quad \forall\, t \in [0, T[;$

v) *(Local uniqueness) If $z^1$ is a solution of Problem $\mathbf{SP}(z_0;\ [0, T_1[)$ $(0 < T_1 \leqslant +\infty)$ and $z^2$ is a solution of problem $\mathbf{SP}(z_0; [0, T_2[)$ $(0 < T_2 \leqslant +\infty)$ then there exists $T \in ]0, \min\{T^1, T^2\}]$ such that $z^1 \equiv \{z^1\}$ on $[0, T[$, $z^2 \equiv \{z^2\}$ on $[0, T[$ and $\{z^1\}_{|[0,T[} \equiv \{z^2\}_{|[0,T[}$.*

*Proof* Let us denote by $z$ the unique solution of the ODE in (86) on $[0, +\infty[$ given by

$$z(t) = W e^{At} W^{-1} z_0, \quad \forall\, t \geqslant 0.$$

Here $z$ is analytic on $\mathbb{R}^+$. Let $\alpha$ be such that $\bar{z}_{0,\alpha} > 0$ and $\bar{z}_{0,k} = 0$ for all $1 \leqslant k \leqslant \alpha - 1$ (if $\alpha \neq 1$). We claim that there exists $T > 0$ such that

$$z_1(t) > 0, \quad \forall\, t \in ]0, T[. \tag{87}$$

If $\alpha = 1$ then the result is clear. If $\alpha \geqslant 2$ then there exists $T > 0$ such that $z_\alpha(t) > 0, \forall\, t \in ]0, T[$ and by integrating the chain of integrators in the second relations in (86), we finally obtain that (87) holds. Condition (87) entails that $N_{\mathbb{R}^+}(z_1(t)) = \{0\}$ and $T^i_\Phi(Z_i(t)) = \mathbb{R}$ for all $1 \leqslant i \leqslant r - 1$ and for all $t \in ]0, T[$. It results that $z$ satisfies the relations in (49)–(51), (52), (53) on $[0, T[$. Thus parts **i)** and **ii)** are proved.

This solution satisfies the ODE in (86) and thus, for all $t \in [0, T[$, we get

$$\|z(t)\| \leqslant \|e^{WAW^{-1}t} z_0\| \leqslant \|e^{WAW^{-1}t}\|\|z_0\| \leqslant e^{\Lambda t}\|z_0\|.$$

It follows that **iii)** holds. Note that this result can also been obtained as a consequence of Theorem 2.

Let $0 \leqslant t < T$ be given. For all $0 \leqslant s \leqslant t$, we have

$$\|\dot{z}(s)\| = \|WAW^{-1}z(s)\| \leqslant \Lambda e^{\Lambda s}\|z_0\| \leqslant \Lambda e^{\Lambda t}\|z_0\|.$$

Thus

$$\mathrm{var}(z, [0, t]) \leqslant \|z_0\|\Lambda e^{\Lambda t} t$$

and the result in **iv)** holds.

Let us now denote by $z$ a solution of Problem $\mathbf{SP}(z_0;\ [0,\bar{T}[)$ for some $0 < \bar{T} \leqslant +\infty$. From the proof of Lemma 4, we know that there exists $\eta \in ]0,\bar{T}]$ such that $z_1 \equiv \{z_1\}$ is continuous on $[0,\eta[$ and $z_1(t) > 0, \forall\ t \in ]0,\eta[$. So, we have $N_{\mathbb{R}^+}(z_1(t)) = \{0\}$ and $T^i_{\Phi}(Z_i(t)) = \mathbb{R}$ for all $1 \leqslant i \leqslant r-1$ and all $t \in ]0,\eta[$. Thus $z \equiv \{z\}$ is solution of the ODE (86) on $[0,\eta[$ and is therefore uniquely defined on $[0,\eta[$. The local uniqueness result in **v)** follows. □

Let us now discuss the case $\bar{z}_0 = 0$.

**Theorem 4** *Assume that* $CA^{r-1}B > 0$. *If* $\bar{z}_0 = 0$, *then:*

i) *(Local existence) There exists* $T > 0$ *such that Problem* $\mathbf{SP}(z_0;\ [0,T[)$ *has at least one regular solution* $z \equiv \{z\}$*;*
ii) $z_1(t) \geqslant 0, \quad \forall\, t \in [0,T[$;
iii) $\|z(t)\| \leqslant e^{\Lambda t}\|z_0\|, \quad \forall\, t \in [0,T[$*;*
iv) $\mathrm{var}(z,[0,t]) \leqslant \|z_0\|\Lambda e^{\Lambda t}t, \quad \forall\, t \in [0,T[$*;*
v) *(Local uniqueness in the class of regular solutions ) If* $z^1$ *is a regular solution of Problem* $\mathbf{SP}(z_0;\ [0,T_1[)(0 < T_1 \leqslant +\infty)$ *and* $z^2$ *is a regular solution of Problem* $\mathbf{SP}(z_0;[0,T_2[)\ (0 < T_2 \leqslant +\infty)$ *then there exists* $T \in ]0,\min\{T^1,T^2\}]$ *such that* $z^1 \equiv \{z^1\}$ *on* $[0,T[$*,* $z^2 \equiv \{z^2\}$ *on* $[0,T[$ *and* $\{z^1\}_{|[0,T[} \equiv \{z^2\}_{|[0,T[}$.

*Proof* Either **(a)** $d_\xi^{\mathrm{T}} A_\xi^k \xi_0 = 0, \forall\ k \in \mathbb{N}$ or **(b)** there exists $\alpha \in \mathbb{N}$ such that $d_\xi^{\mathrm{T}} A_\xi^\alpha \xi_0 \neq 0$ and (if $\alpha \neq 0$) $d_\xi^{\mathrm{T}} A_\xi^k \xi_0 = 0, \forall\, 0 \leqslant k \leqslant \alpha - 1$.

Case **(a)**. Here we check that $z^{\mathrm{T}} \equiv (\bar{z}^{\mathrm{T}}, \xi^{\mathrm{T}})$ with

$$\bar{z}(t) = 0^r, \quad \xi(t) = e^{A_\xi t}\xi_0$$

is a regular solution of Problem $\mathbf{SP}(z_0;[0,+\infty[)$. Indeed, the mapping $t \mapsto \mathrm{d}_\xi^{\mathrm{T}} e^{A_\xi t}\xi_0$ is analytic and we get here:

$$d_\xi^{\mathrm{T}}\xi(t) = \sum_{k=0}^{\infty} d_\xi^{\mathrm{T}}\xi^{(k)}(0)\frac{t^k}{k!} = \sum_{k=0}^{\infty} d_\xi^{\mathrm{T}} A_\xi^k \xi_0 \frac{t^k}{k!} = 0.$$

We have thus

$$\dot{z}_i(t) - z_{i+1}(t) = 0, \quad \forall\, t \geqslant 0,\ \ 1 \leqslant i \leqslant r-1$$
$$(CA^{r-1}B)^{-1}(\dot{z}_r(t) - d^{\mathrm{T}}\bar{z}(t) - d_\xi^{\mathrm{T}}\xi(t)) = -(CA^{r-1}B)^{-1}d_\xi^{\mathrm{T}}\xi(t) = 0, \quad \forall\, t \geqslant 0,$$

and

$$\dot{\xi}(t) = A_\xi\xi(t)\ (= A_\xi\xi(t) + B_\xi z_1(t)), \quad \forall\, t \geqslant 0.$$

Case **(b)**. Let us first discuss the case **(b-1)** $d_\xi^{\mathrm{T}} A_\xi^\alpha \xi_0 < 0$ and (if $\alpha \neq 0$) $d_\xi^{\mathrm{T}} A_\xi^k \xi_0 = 0, \forall\, 0 \leqslant k \leqslant \alpha - 1$.

We check that $z^{\mathrm{T}} \equiv (\bar{z}^{\mathrm{T}}, \xi^{\mathrm{T}})$ with

$$\bar{z}(t) = 0^r, \quad \xi(t) = \mathrm{e}^{A_\xi t}\xi_0$$

is a local regular solution of our problem. There exists $\sigma > 0$ such that

$$d_\xi^{\mathrm{T}} \xi^{(\alpha)}(s) < 0, \quad \forall\, s \in [0, \sigma[$$

since $d_\xi^{\mathrm{T}} \xi^{(\alpha)}(0) = d_\xi^{\mathrm{T}} A_\xi^\alpha \xi_0 < 0$ and $d_\xi^{\mathrm{T}} \xi(.)$ is continuous.

Let $t \in ]0, \sigma[$ be given. We have (if $\alpha \neq 0$):

$$d_\xi^{\mathrm{T}} \xi^{(\alpha-1)}(t) = d_\xi^{\mathrm{T}} \xi^{(\alpha-1)}(0) + \int_0^t d_\xi^{\mathrm{T}} \xi^{(\alpha)}(s)\mathrm{d}s = \int_0^t d_\xi^{\mathrm{T}} \xi^{(\alpha)}(s)\mathrm{d}s < 0$$

and so on, we get finally that

$$d_\xi^{\mathrm{T}} \xi(t) \leqslant 0, \quad \forall\, t \in [0, \sigma[.$$

We have

$$\begin{aligned} &\dot{z}_i(t) - z_{i+1}(t) = 0, \quad \forall\, t \in [0, \sigma) \;\; (1 \leqslant i \leqslant r-1) \\ &(CA^{r-1}B)^{-1}(\dot{z}_r(t) - d^{\mathrm{T}}\bar{z}(t) - d_\xi^{\mathrm{T}}\xi(t)) = -(CA^{r-1}B)^{-1} d_\xi^{\mathrm{T}}\xi(t) \geqslant 0, \quad \forall\, t \in [0, \sigma] \end{aligned}$$

and thus

$$(CA^{r-1}B)^{-1}(\dot{z}_r(t) - d^{\mathrm{T}}\bar{z}(t) - d_\xi^{\mathrm{T}}\xi(t)) \in -\partial\psi_{T_\Phi^{r-1}(0,\ldots,0)}(0) = \mathbb{R}^+, \; \forall t \in [0, \sigma[.$$

Moreover

$$\dot{\xi}(t) = A_\xi \xi(t) \; (= A_\xi \xi(t) + B_\xi z_1(t)), \quad \forall\, t \in [0, \sigma[.$$

Let us now discuss the case **(b-2)** $d_\xi^{\mathrm{T}} A_\xi^\alpha \xi_0 > 0$ and (if $\alpha \neq 0$) $d_\xi^{\mathrm{T}} A_\xi^k \xi_0 = 0, \forall\, 0 \leqslant k \leqslant \alpha - 1$. We check that $z$ defined by

$$z(t) = W\mathrm{e}^{At}W^{-1}z_0$$

is a local regular solution of our problem. We know that $z$ is the solution of the system in (86) and we have $z_1(0) = \cdots = z_r(0) = 0$. We claim that there exists $\eta > 0$ such that

$$z_r^{(\alpha+1)}(t) > 0, \quad \forall\, t \in [0, \eta[.$$

If $\alpha = 0$ the result is trivial since $\dot{z}_r(0) = d^{\mathrm{T}}\bar{z}(0) + d_\xi^{\mathrm{T}}\xi_0 = d_\xi^{\mathrm{T}}\xi_0 > 0$. If $\alpha \neq 0$ then

$$\dot{z}_r(0) = d^{\mathrm{T}}\bar{z}(0) + d_\xi^{\mathrm{T}}\xi(0) = d_\xi^{\mathrm{T}}\xi_0 = 0.$$

Then

$$\begin{aligned} z_r^{(2)}(0) &= d^{\mathrm{T}}\dot{\bar{z}}(0) + d_\xi^{\mathrm{T}}(A_\xi\xi_0 + B_\xi z_1(0)) \\ &= d_1 z_2(0) + \cdots + d_{r-1}z_r(0) + d_r\dot{z}_r(0) + d_\xi^{\mathrm{T}}(A_\xi\xi_0 + B_\xi z_1(0)) = 0 \end{aligned}$$

and so on, remarking that if $0 \leqslant j \leqslant r-1$ then $z_1^{(j)} = z_{j+1}$ while if $j \geqslant r$ then $z_1^{(j)} = z_r^{(j-r+1)}$, until we get:

$$z_r^{(\alpha+1)}(0) = d^{\mathrm{T}}\bar{z}^{(\alpha)}(0) + d_\xi^{\mathrm{T}}(A_\xi^\alpha\xi_0 + B_\xi z_1^{(\alpha-2)}(0)) = d_\xi^{\mathrm{T}}A_\xi^\alpha\xi_0 > 0.$$

It results that there exists $\eta > 0$ such that

$$z_r^{(\alpha+1)}(t) > 0, \quad \forall\, t \in [0, \eta[.$$

Then we get

$$z_i(t) > 0, \quad \forall\, t \in ]0, \eta[ \ \ (1 \leqslant i \leqslant r).$$

We may now conclude that $z$ is a solution of Problem $\mathbf{SP}(z_0)$ on $[0, \eta[$. Indeed, if $t \in ]0, \eta[$ then $N_{\mathbb{R}^+}(z_1(t)) = \{0\}$ and consequently $-\partial\psi_{T_\Phi^{i-1}(\{Z_{i-1}\}(t^-))}(\{z_i\}(t^+)) = \{0\}$ $(2 \leqslant i \leqslant r)$.

Thus, in each case, there exists a local regular solution of our problem on some interval $[0, T[$ such that $z_1 \geqslant 0$ on $[0, T[$ and the results in **i)** and **ii)** follow.

In cases **(a)** and **(b-1)**, we check that

$$\|z(t)\| \leqslant \mathrm{e}^{|||A_\xi|||t}\|\xi_0\|, \quad \forall\, t \in [0, T[$$

and

$$\operatorname{var}(z, [0, t]) \leqslant \|\xi_0\|\,|||A_\xi|||\mathrm{e}^{|||A_\xi|||t}t, \quad \forall\, t \in [0, T[.$$

In case (b-2), we have

$$\|z(t)\| \leqslant \mathrm{e}^{\Lambda t}\|\xi_0\|, \quad \forall\, t \in [0, T[$$

and

$$\operatorname{var}(z, [0, t]) \leqslant \|\xi_0\|\Lambda\mathrm{e}^{\Lambda t}t, \quad \forall\, t \in [0, T[.$$

Here $\|z_0\| = \|\xi_0\|$ since $\bar{z}_0 = 0$ and $|||A_\xi||| \leqslant \Lambda$ since the matrix $A_\xi$ is a submatrix of $WAW^{-1}$ obtained by deleting a total of $r$ rows and columns from $A$. The results in **iii)** and **iv)** follow.

Let us here first denote by $z_a$ the analytic solution of Problem $\mathbf{SP}(z_0;[0,T[)$ given in part i) of this proof. Let us now denote by $z$ any regular solution of our problem on some interval $[0,\bar{T}[$ $(0<\bar{T}\leqslant+\infty)$. On $[0,\min\{T,\bar{T}\}[$, we may write

$$\mathrm{d}\{z\}=WAW^{-1}\{z\}\mathrm{d}t+N\mathrm{d}\nu$$

and

$$\mathrm{d}\{z_a\}=WAW^{-1}\{z_a\}\mathrm{d}t+N\mathrm{d}\nu_a$$

where the matrix $N$ is the diagonal matrix defined in (79), $\mathrm{d}\nu:=(\mathrm{d}\nu_1,\dots,\mathrm{d}\nu_{r-1},\mathrm{d}\nu_r,0_{n-r})^{\mathrm{T}}$ and $\mathrm{d}\nu_a:=(\mathrm{d}\nu_{a,1},\dots,\mathrm{d}\nu_{a,r-1},\mathrm{d}\nu_{a,r},0_{n-r})^{\mathrm{T}}$. Moreover, as seen above, we have, in each case, on $[0,\min\{T,\bar{T}\}[$:

$$\mathrm{d}\nu_{a,i}\equiv 0,\quad \forall\, 1\leqslant i\leqslant r-1,$$

while

$$\mathrm{d}\nu_{a,r}=\lambda_{a,r}\mathrm{d}t$$

for some nonnegative analytic mapping $\lambda_{a,r}$. Then

$$\mathrm{d}(\{z\}-\{z_a\})=WAW^{-1}(\{z\}-\{z_a\})\mathrm{d}t+N\mathrm{d}\nu-N\mathrm{d}\nu_a$$

and

$$\begin{aligned}(\{z\}^+-\{z_a\}^+)^{\mathrm{T}}d(\{z\}-\{z_a\})=(\{z\}^+-\{z_a\}^+)^{\mathrm{T}}WAW^{-1}(\{z\}-\{z_a\})\mathrm{d}t\\ +(\{z\}^+-\{z_a\}^+)^{\mathrm{T}}N(\mathrm{d}\nu-\mathrm{d}\nu_a).\end{aligned}$$

It is clear that (see (46) and (47)) $(\{z\}^+)^{\mathrm{T}}N\mathrm{d}\nu\equiv 0$ and $(\{z_a\}^+)^{\mathrm{T}}N\mathrm{d}\nu_a\equiv 0$ on $[0,\min\{T,\bar{T}\}[$. Moreover $-(\{z_a\}^+)^{\mathrm{T}}N\mathrm{d}\nu\leqslant 0$ on $[0,\min\{T,\bar{T}\}[$. Indeed, $CA^{r-1}B>0$, $\mathrm{d}\nu\geqslant 0$ on $[0,\min\{T,\bar{T}\}[$ and in cases **(a)** and **(b-1)**, $z_{a,i}\equiv 0$ $(1\leqslant i\leqslant r)$ on $[0,\min\{T,\bar{T}\}[$, while in case **(b-2)**, $z_{a,i}>0$ $(1\leqslant i\leqslant r)$ on $[0,\min\{T,\bar{T}\}[$. Then, recalling that the mappings $z$ and $z_a$ are $\mathbf{RCSLBV}([0,\min\{T,\bar{T}\}[;\mathbb{R}^n))$-functions, we get

$$\mathrm{d}(\|\{z\}-\{z_a\}\|^2)\leqslant 2\big((\{z\}-\{z_a\})^{\mathrm{T}}WAW^{-1}(\{z\}-\{z_a\})\mathrm{d}t-CA^{r-1}B\lambda_{a,r}\{z_r\}\mathrm{d}t\big)$$

and then, for each $0<t<\min\{T,\bar{T}\}$, we get

$$\begin{aligned}&\|\{z\}(t)-\{z_a\}(t)\|^2-\|\{z\}(0^-)-\{z_a\}(0^-)\|^2\\ &\quad=\int_{[0,t]}\mathrm{d}(\|\{z\}-\{z_a\}\|^2)\leqslant 2\Lambda\int_0^t\|\{z\}(s)-\{z_a\}(s)\|^2\mathrm{d}s-2CA^{r-1}B\int_0^t\lambda_{a,r}(s)\{z_r\}(s)\mathrm{d}s.\end{aligned}$$

However, here $\{z\}(0) = \{z_a\}(0) = z_0$, and thus

$$\|\{z\}(t) - \{z_a\}(t)\|^2 \leqslant 2\Lambda \int_0^t \|\{z\}(s) - \{z_a\}(s)\|^2 \mathrm{d}s - 2CA^{r-1}B \int_0^t \lambda_{a,r}(s)\{z_r\}(s)\mathrm{d}s. \tag{88}$$

Let us now check that there exists $0 < \sigma < \bar{T}$ such that

$$\{z_r\}(t) \geqslant 0, \quad \forall\, t \in [0, \sigma[.$$

We know that $\{z_1\}(0) = \cdots = \{z_r\}(0) = 0$ and there exists $\eta \in ]0, \bar{T}]$ such that the mapping $t \mapsto z_r(t)$ is analytic on $[0, \eta[$. Then $\{z_r\} \equiv z_r$ on $[0, \eta[$.

Either **(A)** $z_r^{(k)}(0) = 0, \forall k \in \mathbb{N}$ or **(B)** there exists $\alpha \in \mathbb{N}$ such that $z_r^{(\alpha)}(0) \neq 0$ and (if $\alpha \neq 0$) $z_r^{(k)}(0) = 0, \forall\, 0 \leqslant k \leqslant \alpha - 1$.

In case **(A)**, we get directly that $z_r(t) = 0, \forall t \in [0, \eta[$. In case **(B)**, $z_r^{(\alpha)}(0) \neq 0$. It is clear that $z_r^{(\alpha)}(0) > 0$. Indeed, if we suppose on the contrary that $z_r^{(\alpha)}(0) < 0$ then there exists $0 < \delta < \eta$ such that $z_r^{(\alpha)}(t) < 0, \forall\, t \in ]0, \delta[$. Then integrating the chain of integrators in (49) we obtain finally that $z_1(t) < 0, \forall\, t \in ]0, \delta[$ and a contradiction. Thus $z_r^{(\alpha)}(0) > 0$ and there exists $0 < \sigma < \eta$ such that $z_r^{(\alpha)}(t) > 0, \forall\, t \in [0, \sigma[$. Then we obtain $z_r(t) > 0, \forall\, t \in ]0, \sigma[$ and thus, for all $t \in [0, \min\{T, \sigma\}[$, we get

$$\|\{z\}(t) - \{z_a\}(t)\|^2 \leqslant 2\Lambda \int_0^t \|\{z\}(s) - \{z_a\}(s)\|^2 \mathrm{d}s. \tag{89}$$

Using Lemma 2, we get, for all $0 \leqslant t < \min\{T, \sigma\}$:

$$\|\{z\}(t) - \{z_a\}(t)\|^2 \leqslant 0$$

and the result in **v)** follows. □

If $\bar{z}_0 \not\succeq 0$ then we proceed with an initial state reinitialization in requiring that $z(0^+) = z_0'$ where $z_0'$ is uniquely defined by

$$z_{0,i}' = \mathrm{prox}\left[T_\Phi^{i-1}(Z_{0,i-1}); z_{0,i}\right], \quad \forall\, 1 \leqslant i \leqslant r,$$

and

$$z_{0,l}' = z_{0,l}, \quad (r + 1 \leqslant l \leqslant n).$$

Then $\bar{z}_0' \succeq 0$, $\|z_0'\| \leqslant \|z_0\|$ and we may apply Theorems 3 and 4 to get the following local existence result.

**Corollary 1** (Local existence and uniqueness (in the class of regular solutions)) *Suppose that* $CA^{r-1}B > 0$*. For each* $z_0 \in \mathbb{R}^n$ *there exists* $T > 0$ *such that Problem* $\mathbf{SP}(z_0;\ [0,T[)$ *has at least one regular solution* $z$ *such that:*

**i)** $z \equiv \{z\}$ *on* $]0,T[$*;*
**ii)** $z_1 \equiv \{z_1\} \geqslant 0$ *on* $[0,T[$*;*
**iii)** $z_i = \{z_i\} + \sum_{j=1}^{i-1}(z'_{0,j} - z_{0,j})\delta_0^{(i-j-1)}\quad (2 \leqslant i \leqslant r)$;
**iv)** $\|\{z\}(t)\| \leqslant e^{\Lambda t}\|z_0\|,\ \ \forall\, t \in [0,T[$*;*
**v)** $\mathrm{var}(\{z\},[0,t]) \leqslant \|z_0\|\Lambda e^{\Lambda t}t,\ \ \forall\, t \in [0,T[$.

*Moreover:*

**vi)** *(Local uniqueness in the class of regular solutions) If* $z^1$ *is a regular solution of Problem* $\mathbf{SP}(z_0;\ [0,T_1[)(0 < T_1 \leqslant +\infty)$ *and* $z^2$ *is a regular solution of problem* $\mathbf{SP}(z_0;[0,T_2[)\ (0 < T_2 \leqslant +\infty)$ *then there exists* $T \in ]0,\min\{T^1,T^2\}]$ *such that* $\langle z^1,\varphi\rangle = \langle z^2,\varphi\rangle,\ \ \forall\, \varphi \in C_0^\infty([0,T[;\mathbb{R}^n)$. □

Let us now provide a global existence and uniqueness result.

**Corollary 2** (Global Existence and Uniqueness (in the class of Regular Solutions)) *Suppose that* $CA^{r-1}B > 0$*. For each* $z_0 \in \mathbb{R}^n$*, Problem* $\mathbf{SP}(z_0;\ [0,+\infty[)$ *has at least one regular solution* $z$ *such that:*

**i)** $z_1 \equiv \{z_1\} \geqslant 0$ *on* $[0,+\infty[$*;*
**ii)** $\|\{z\}(t)\| \leqslant e^{\Lambda t}\|z_0\|,\ \ \forall\, t \in [0,+\infty[$.

*Moreover:*

**iii)** *(Uniqueness in the class of regular solutions) If* $z^*$ *denotes a regular solution of problem* $\mathbf{SP}(z_0;\ [0,T^*[)(0 < T^* \leqslant +\infty)$ *then* $\langle z^*,\varphi\rangle = \langle z,\varphi\rangle,\ \ \forall \varphi \in C_0^\infty([0,T[;\mathbb{R}^n)$.

*Proof* Corollary 1 ensures the existence of $0 < T < +\infty$ such that Problem $\mathbf{SP}(z_0;\ [0,T[)$ has a regular solution $z$ such that $z \equiv \{z\}$ on $]0,T[$ and $z_1 \geqslant 0$ on $[0,T[$.

Theorem 2 ensures that the limit

$$\{z\}(T^-) = \lim_{t\to T, t<T}\{z\}(t)$$

exists and is finite. We may then proceed with a state reinitialization by requiring that $\{z\}(T^+) = z_1$, where $z_1$ is uniquely defined by

$$z_{1,i} = \mathrm{prox}\left[T_\Phi^{i-1}(\{Z_{i-1}\}(T^-));\{z_i\}(T^-)\right],\quad (1 \leqslant i \leqslant r),$$
$$z_{1,l} = \{z_l\}(T^-),\quad (r+1 \leqslant l \leqslant n).$$

From Corollary 1, we get a prolongation of $z$ as a regular solution of Problem $\mathbf{SP}(z_0;\ [0,T_1[)$ with $T_1 > T$ such that $z_1 \geqslant 0$ on $[0,T_1[$. Let us now denote by

$H_{\max}$ the supremum of the $H > T$ such that there exists a prolongation of $z$ as a regular solution of Problem $\mathbf{SP}(z_0;\ [0,H[)$. We claim that $H_{\max} = +\infty$. Indeed, suppose on the contrary that $H_{\max} < +\infty$. Then from Theorem 2, we may assert that $z(H_{\max}^-)$ exists and is finite. Then Corollary 1 furnishes a prolongation of $z$ as a regular solution of Problem $\mathbf{SP}(z_0;\ [0,H_{\max}+\varepsilon[)$ for some $\epsilon > 0$, which contradicts the definition of $H_{\max}$. The existence of a global regular solution that satisfies condition **i)** follows. Condition **ii)** is a consequence of Theorem 2. To prove the result in **iii)**, it suffices to check that:

$$\{z\}(t) = \{z^*\}(t), \quad \forall\, t \in [0,T^*[.$$

Suppose by contradiction that there exists $s \in [0,T^*[$ such that $\{z\}(s) \neq \{z^*\}(s)$. It is clear that $s > 0$ since $\{z\}(0) = \{z^*\}(0)$. Let us set

$$\tau = \inf\{s \in ]0,T[ \colon \{z\}(s) \neq \{z^*\}(s)\}.$$

If $\tau > 0$ then, necessarily, $\{z\}(h) = \{z^*\}(h)$ for all $h \in [0,\tau[$ and thus $\{z\}(\tau^-) = \{z^*\}(\tau^-)$ (if $\tau = 0$ then we have also $\{z\}(\tau^-) = \{z^*\}(\tau^-)(= z_0)$). Then $\{z\}(\tau^+) = \{z^*\}(\tau^+)$ and Theorems 3 and 4 ensure the existence of $\delta > 0$ such that $\{z\}(s) = \{z^*\}(s)$ for all $s \in [\tau,\tau+\delta[$ which contradicts the definition of $\tau$. □

*Remark 19* i) If $CA^{r-1}B > 0$ and $r = n$ then any solution of Problem $\mathbf{SP}(z_0;\ [0,T[)$ $(0 < T \leqslant +\infty)$ is regular. Indeed, let $t \in [0,T[$ be given. Here $\{z\}(t) = \{\bar{z}\}(t)$ and from Proposition 6 (see also Remark 16), we deduce that $\{z\}(t) = \{z\}(t^+) \succeq 0$. Then if $\{z\}(t) \neq 0$, the result follows from Lemma 4, while if $\{z\}(t) = 0$ then necessarily $\{z\}(s) = 0,\ \forall s \in [t,T[$ and the result follows.

ii) If $CA^{r-1}B > 0$ and $r = 1$ then any solution of Problem $\mathbf{SP}(z_0;\ [0,T[)$ $(0 < T \leqslant +\infty)$ is regular. Let $t \in [0,T[$ be given. Let us first note that here $\{\bar{z}\}(t) = z_1(t)$. If $z_1(t) > 0$, the result follows from Lemma 4. If $z_1(t) = 0$ then there exists $\varepsilon > 0$ such that for all $s \in [t,t+\varepsilon[$, we have (see (88)):

$$\|\{z\}(s) - \{z_a\}(s)\|^2 \leqslant 2\Lambda \int_t^s \|\{z\}(\tau) - \{z_a\}(\tau)\|^2 \mathrm{d}\tau - 2CA^{r-1}B \int_t^s \lambda_{a,1}(\tau) z_1(\tau)\mathrm{d}\tau.$$

where $z_a$ denotes the regular solution of Problem $\mathbf{SP}(z_0;\ [t,t+\varepsilon[)$ given by Theorem 4. Here we have $\int_t^s \lambda_{a,1}(\tau)z_1(\tau)\mathrm{d}\tau \geqslant 0$ and then from Lemma 2, we see that $\{z\} \equiv \{z_a\}$ on $[t,t+\varepsilon[$ and the result follows. So in both cases uniqueness in the set of regular solutions implies uniqueness of the solution.

*Remark 20* i) Let $0 = T_0 < T_1 < T_2 < \cdots$ be the sequence of non-trivial ($\{z\}(T_i^-) \neq \{z\}(T_i^+)$) state reinitialization points required to get the global solution $z$. Then $z_1 = \{z_1\}$ and

$$z_i = \{z_i\} + \sum_{\alpha} \sum_{j=1}^{i-1} (\{z_j\}(T_\alpha^+) - \{z_j\}(T_\alpha^-))\delta_{T_\alpha}^{(i-j-1)} \quad (2 \leqslant i \leqslant r)$$

ii) Corollary 1 and Corollary 2 can be generalized to the multivariable case $m \geqslant 2$ with relative degree $\bar{r}$ provided that one supposes that $CA^{r-1}B$ is a Stieltjes matrix, i.e. a nonsingular symmetric M-matrix [6]. This assumption secures that the matrix $CA^{r-1}B$ is positive definite and $(CA^{r-1}B)^{-1}$ is nonnegative in the sense that $(CA^{r-1}B)^{-1}_{ij} \geqslant 0$ for all $i,j \in \{1,\ldots,n\}$. In the case $r = 1$ such an assumption is restrictive and other results exist (see e.g. [17,55]) that do not use it. The reason is that our framework encompasses the case $r \geqslant 2$ as well. Note that if $m \geqslant 2$, the state reinitialization of $\{z_r\}$ at time $T$ reads

$$\{z_r\}(T^+) = \mathrm{prox}_{(CA^{r-1}B)^{-1}}\left[T_\Phi^{r-1}(\{Z_{r-1}\}(T^-)); \{z_r\}(T^-)\right]$$

and it would be convenient to use the norm

$$\|z\| := \sqrt{\sum_{\alpha=1}^{r-1} \|z_\alpha\|_m^2 + \langle CA^{r-1}Bz_r, z_r\rangle + \|\xi\|_{n-mr}^2}$$

in order to generalize the results.
iii) Applying Corollary 2 and Proposition 8 it follows that autonomous “dissipative” higher order sweeping processes are well-posed systems. An example is detailed in Sect. 6.

*Remark 21* An LCS consists of (24) with a complementarity relation $0 \leqslant w(t) \perp \lambda(t) \geqslant 0$ [26,28]. It is clear that the formalism proposed in this paper involves complementarity conditions between some variables, see e.g. (47), (72) or (76). However the work in [26,28] and the work in this paper differ in several fundamental ways. The model (or formalism) that is considered in [28] and the one in this paper completely differ, rendering any comparison between both works hazardous. Let us however cite a few ones. The class of distributions that is considered in [28] as potential solutions, is strictly included in $\mathcal{T}_\infty$, see Examples 2 and 4. A local well-posedness result has been obtained in [28, Theorem 6.3] for $m \geqslant 1$, and a global well-posedness result has been obtained in [26, Theorem 3.6.10] for the case $m = 1$. Nevertheless the major discrepancy is that the approach in [26,28] relies on an event-driven point of view on hybrid systems, as is illustrated in their definition of a solution [28, Definition 4.10]. In other words, the system is initialized in a so-called mode (a DAE), and then integration progresses until the system has to switch to another mode. Especially, this way of thinking, leads to event-driven time-integration schemes. No numerical algorithm is proposed in [28] but only a rough guideline for constructing such a scheme [28, Sect. 7]. It is well-known that event-driven schemes possess strong drawbacks such as the impossibility to pass through finite accumulations of events. It seems also quite difficult to obtain convergence proofs with event-driven schemes [15]. The point of view that is adopted here, is dramatically different. It is the point of view of (unbounded) differential inclusions and is in the continuity of the works on the sweeping process and measure differential inclusions by Moreau, Schatzman, Paoli, Stewart, Monteiro-Marques, Mabrouk, to cite a few. It naturally yields time-stepping numerical schemes and paves the way towards convergence proofs, as illustrated in Sect. 5.

## 5 A solution method for the numerical time integration: the EMTS scheme

Following the work of Moreau [45,46,48,51,53] and co-workers [30,35,42], we aim at designing the so-called "Extended Moreau's Time-Stepping" (EMTS) scheme. This time-stepping scheme i.e., without an explicit procedure for handling the times of events, is based on the approximation on a time interval of the Measure Differential Formalism of the Higher Order Moreau's Sweeping Process (57)–(59), (52).

This section is organized as follows. In Sect. 5.1, we recall basic facts about the numerical time-integration of the Moreau's sweeping process with a relative degree less or equal to 2. In Sect. 5.2, the principle of the construction of an approximated solution is stated. In the spirit of the work in [42], several major properties of the proposed scheme are outlined in Sect. 5.3 that pave the way to a convergence result. In Sect. 5.4, we give an overview of a possible implementation of the numerical scheme. Finally, we conclude by giving some numerical applications in Sect. 5.5. Particularly, the numerical scheme is compared to a direct use of a Backward Euler scheme as in [16] (Sect. 5.5.1). Finally, the influence of the zero-dynamics is highlighted (Sect. 5.5.2) and the empirical order of the scheme is evaluated (Sect. 5.5.3) on a particular example.

The following notation is used throughout this section. We denote by $0 = t_0 < t_1 < \cdots < t_k < \cdots < t_N = T$ a finite partition (or a subdivision) of the time interval $[0, T]$ $(T > 0)$. The integer $N$ stands for the number of time intervals in the subdivision. The length of a time step is denoted by $h_k = t_{k+1} - t_k$. For simplicity sake, we consider only in the sequel a constant time length $h = h_k$ $(0 \leqslant k \leqslant N-1)$. Then $T = N h$. The approximation of $f(t_k)$, the value of a real function $f$ at the time $t_k$, is denoted by $f_k$ .

### 5.1 Background on the numerical time integration of the Moreau's sweeping process

In this section, we give some details about the seminal work of Jean Jacques Moreau on the numerical time integration of the sweeping process.

#### *5.1.1 First order sweeping process*

Let us consider the first order sweeping process, or more precisely, the sweeping process of relative degree 1 introduced in (1),

$$\begin{cases} -\mathrm{d}z \in N_{K(t)}(z(t)) \quad (t \geqslant 0), \\ z(0) = z_0. \end{cases} \tag{90}$$

Under suitable hypothesis on the multivalued function $t \mapsto K(t)$ (Lipschitz, bounded variations, finite retraction in the sense of the Hausdorff distance, ...), numerous convergence results [35,42,48] have been given together with well-posedness results using the so-called "Catching-up algorithm" defined in [48] as

$$-(z(t_{k+1}^{+}) - z(t_k^{+})) \in \partial\psi_{K(t_{k+1}^{+})}(z(t_{k+1}^{+})). \tag{91}$$

By elementary convex analysis, and using the convention that $z_{k+1} = z(t_{k+1}^{+})$, the inclusion (91) is equivalent to

$$z_{k+1} = \mathrm{prox}(K(t_{k+1}^{+}), z_k). \tag{92}$$

Contrary to the standard backward Euler scheme with which it might be confused, the catching-up algorithm is based on the evaluation of the measure $\mathrm{d}z$ on the interval $]t_k, t_{k+1}]$, i.e. $\mathrm{d}z(]t_k, t_{k+1}]) = z(t_{k+1}^{+}) - z(t_k^{+})$. Indeed, the backward Euler scheme is based on the approximation of $\dot{z}(t)$ which is not defined in a classical sense for our case. When the time step vanishes, the approximation of the measure $dz$ tends to a finite value corresponding to the jump of $z$. This remark is crucial for the consistency of the scheme. Particularly, this fact ensures that we handle only finite values.

In the same way, using higher order numerical schemes is at best useless, more often it is dangerous. Basically, a general way to obtain a finite difference-type scheme of order $n$ is to write a Taylor expansion of order $n$ or higher. Such a scheme is meant to approximate the $n$-th derivative of the discretized function. If the solution we are dealing with is obviously not differentiable, what is the meaning of using a scheme with order $n \geqslant 2$? Such a scheme will try to approximate derivatives which do not exist. At the times of non-differentiability, it may introduce in the solution some artificial unbounded terms creating spurious oscillations. In summary, higher-order numerical schemes are inadequate for time-stepping discretization of complementarity systems.

### *5.1.2 Second order sweeping process: overview of the contact dynamics method*

The "Non Smooth Contact Dynamics (NSCD)" method [30,53,54] is the numerical discretization of the second order Moreau Sweeping process introduced by Moreau in [50,51] in the context of Lagrangian mechanical systems subject to a generalized position constraint $q \in \Phi$, $\Phi \subset \mathbb{R}^n$, reformulated as a measure differential inclusion [50,51,64],

$$M(q)\,\mathrm{d}v + F(q, v, t)\,\mathrm{d}t = \mathrm{d}p \tag{93}$$

where $q$ is the vector of absolutely continuous generalized coordinates, $\mathrm{d}v$ is a differential measure associated with the velocity $v(t) = \dot{q}(t^{+})$ considered as a RCLBV function, $\mathrm{d}t$ is the Lebesgue measure, $M(q)$ is the mass matrix and $F(q, v, t)$ is the set of forces acting upon the system. The unilateral constraint $q \in \Phi$ is enforced by the multiplier measure, $\mathrm{d}p$. To complete this measure differential equation, Moreau proposed a compact formulation of an inelastic impact law as a measure inclusion,

$$\mathrm{d}p \in \partial\psi_{V(q(t))}(v(t^+) + ev(t^-)) \tag{94}$$

where $V(q)$ is the tangent cone to $\Phi$ at $q$, $e$ is the coefficient of restitution. Finally, we obtain a MDI, the so-called sweeping process,

$$M(q(t))\,\mathrm{d}v + F(q(t), v(t), t)\,\mathrm{d}t \in -\partial\psi_{V(q(t))}(v(t^+) + ev(t^-)). \tag{95}$$

The NSCD method performs the numerical time integration of the MDI (95) on an interval $]t_k, t_{k+1}]$. Using the notation

$$v_{k+1} \approx v(t^+_{k+1}); \quad p_{k+1} \approx \mathrm{d}p(]t_k, t_{k+1}]), \tag{96}$$

it may be written down as follows ($\theta \in [0, 1]$):

$$\begin{cases} -M(v_{k+1} - v_k) - hF((1-\theta)v_{k+1} + \theta v_k, (1-\theta)q_{k+1} + \theta q_k, t_{k+1}) = p_{k+1}, \\ p_{k+1} \in \partial\psi_{T_\Phi(\tilde{q}_{k+1})}(v_{k+1}), \\ q_{k+1} = q_k + h((1-\theta)v_{k+1} + \theta v_k), \\ \tilde{q}_{k+1} = q_k + hv_k. \end{cases}$$

The value $\tilde{q}_{k+1}$ is a prediction of the position which allows the computation of the tangent cone $T_\Phi(\tilde{q}_{k+1})$. A $\theta$-method is used for the integration of the position assuming that $q$ is absolutely continuous. The same approximation is made with the term $F(v(t^+), q(t), t)$.

If $\Phi$ is finitely represented as,

$$\Phi = \{h_\alpha(q) \geqslant 0, \quad \alpha = 1, \ldots, \nu\}, \tag{97}$$

the inclusion can be stated equivalently under some constraints qualification conditions as a conditional complementarity problem,

$$\text{if} \quad h(\tilde{q}_{k+1}) \leqslant 0 \quad \text{then} \quad 0 \leqslant \nabla h^{\mathrm{T}}_\alpha(\tilde{q}_{k+1})v_{k+1} \perp \mu_{k+1} \geqslant 0 \quad , \tag{98}$$

where $\mathrm{d}p = \nabla h_\alpha(q)\lambda$ and $\mu_{k+1} \approx \lambda(]t_k, t_{k+1}])$.

**Comments** From a numerical point of view, two major lessons can be learned form this work: First, the various terms manipulated by the numerical algorithm are of finite value. The use of differential measures of the time interval $]t_k, t_{k+1}]$, i.e., $\mathrm{d}v(]t_k, t_{k+1}]) = v(t_{k+1}^+) - v(t_k^+)$ and $p_{k+1} = \mathrm{d}p(]t_k, t_{k+1}])$, is fundamental and allows a rigorous treatment of the non smooth evolutions. When the time-step $h$ vanishes, it enables to deal with finite jumps. When the evolution is smooth, the scheme is equivalent to a backward Euler scheme. We can remark that nowhere an approximation of the acceleration is used. Secondly, the inclusion in terms of velocity allows us to treat the displacement as a secondary variable. A viability Lemma ensures that the constraints on $q$ will be respected at convergence. We will see further that this formulation gives more stability to the scheme.

These remarks on the contact dynamics method might be viewed only as some numerical tricks. In fact, the mathematical study of the second order MDI by Moreau provides a sound mathematical ground to this numerical scheme. It is noteworthy that some convergence results have been proved for such time-stepping schemes [40,42,69].

**Example of the bouncing ball** The NSCD method provides a numerical scheme with very nice properties. The reader may convince his/herself of this by studying the simple bouncing ball on a rigid plane subject to gravity and with elastic restitution. The proposed time-discretization of the motion of this ball is

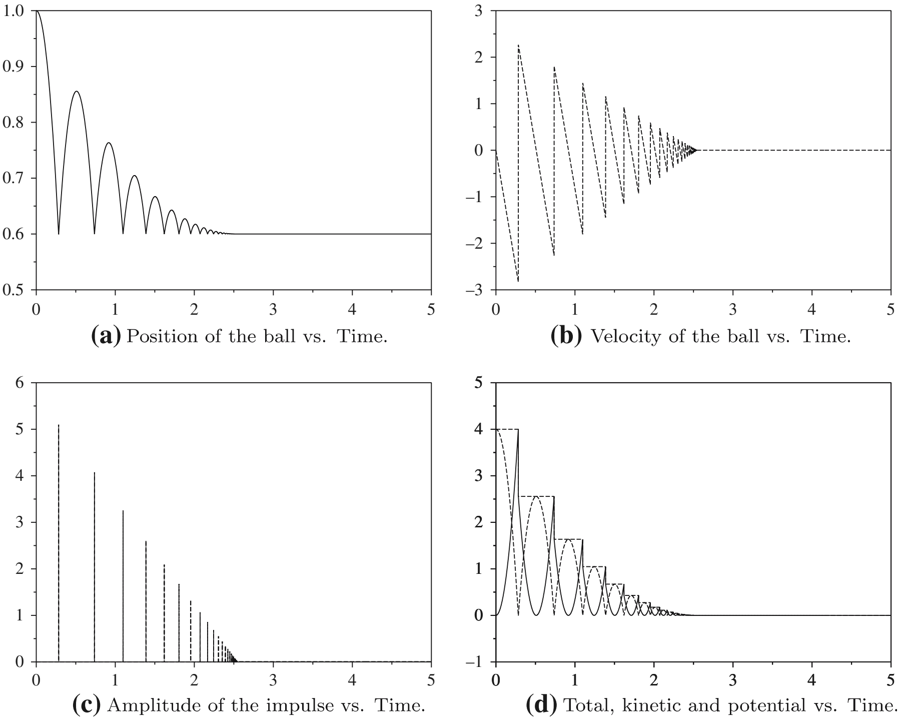
$$-m(v(t_{k+1}) + v(t_k)) - hmg \in \partial\psi_{V(\tilde{q}_{k+1})}(v(t_{k+1}) + ev(t_k)) \tag{99}$$

where $m$ is the mass of the ball, $e$ the coefficient of restitution and $g$ the gravity. One notes that the dissipativity property shows through the power of $h$ in the term $hg$ which has the dimension of an impulse (there is no $h$ pre-multiplying the right-hand-side since this is a cone).

If $q_0 > 0$ then the ball falls down until penetration is detected at step $k^*$ (i.e. $q_{k^*-1} > 0$ while $q_{k^*} < 0$). Then the velocity is reversed, i.e. $v_{k^*+1} = -ev_{k^*}$ while $q_{k^*+1} = q_{k^*}$. When $e = 1$ the system is re-initialized at each impact, with the same velocity and at the same position. There are no errors introduced by the numerical scheme and one can simulate several billions of such cycles. Clearly this is not possible with an event-driven scheme, even if a very accurate detection procedure is used. The unavoidable penetration is not a major issue, since anyway the discretized system cannot be exactly at $q = 0$. What is crucial is that the penetration goes to zero when $h \to 0$. In the case $e \in ]0, 1[$, an infinity of rebounds in finite time occurs in the continuous time model. This Zeno behavior is correctly integrated as depicted on Fig. 1.

## 5.2 Principle

Let us start with a generic equation of the measure differential formalism for the extended sweeping process (57) for $1 \leqslant i \leqslant r - 1$,



**(a)** Position of the ball vs. Time.

**(b)** Velocity of the ball vs. Time.

**(c)** Amplitude of the impulse vs. Time.

**(d)** Total, kinetic and potential vs. Time.

**Fig. 1** Bouncing ball on a rigid plane. $e = 0.8, g = 10\,\mathrm{m.s}^{-2}, m = 1\,\mathrm{kg}, h = 5 \times 10^{-3}\mathrm{s}$

$$\begin{aligned} &\mathrm{d}z_i - z_{i+1}(t)\mathrm{d}t = \mathrm{d}\nu_i, \\ &\quad \mathrm{d}\nu_i \in -\partial\psi_{T_\Phi^{i-1}(Z_{i-1}(t^-))}(z_i(t^+)). \end{aligned} \tag{100}$$

As in Sect. 2, it results from Proposition 1 that an evaluation of this MDI on the time interval $(t_k, t_{k+1}]$ yields

$$\begin{aligned} &\mathrm{d}z_i(]t_k, t_{k+1}]) - \int\limits_{]t_k, t_{k+1}]} z_{i+1}(\tau)\mathrm{d}\tau = \mathrm{d}\nu_i(]t_k, t_{k+1}]) \\ &\mathrm{d}\nu_i(]t_k, t_{k+1}]) \in \overline{\mathrm{co}}\big(\cup_{\tau \in ]t_k, t_{k+1}]} -\partial\psi_{T_\Phi^{i-1}(Z_{i-1}(\tau^-))}(z_i(\tau^+))\big). \end{aligned} \tag{101}$$

The values of the measures $\mathrm{d}z_i(]t_k, t_{k+1}])$ and $\mu_{i,k+1} \stackrel{\Delta}{=} \mathrm{d}\nu_i(]t_k, t_{k+1}])$ are kept as primary variables and this fact is crucial for the consistency of the method for the nonsmooth evolutions. The integral term is approximated thanks to

$$\int\limits_{]t_k, t_{k+1}]} z_{i+1}(\tau)\mathrm{d}\tau \approx h z_{i+1}(t_{k+1}^+) = h z_{i+1,k+1} \tag{102}$$

and then we obtain

$$z_{i,k+1} - z_{i,k} - hz_{i+1,k+1} = \mu_{i,k+1}. \tag{103}$$

For the approximation of the inclusion, the union of convex cones is approximated in the following way:

$$\overline{\mathrm{co}}\big(\cup_{\tau\in]t_k,t_{k+1}]} -\partial\psi_{T_\Phi^{i-1}(Z_{i-1}(\tau^-))}(z_i(\tau^+))\big) \approx -\partial\psi_{T_\Phi^{i-1}(Z_{i-1}(t_k^-))}(z_i(t_{k+1}^+)). \tag{104}$$

Assuming, as in (102), that the approximation of $z_i$ is constant on each interval $]t_k, t_{k+1}]$, we get

$$\mu_{i,k+1} \in -\partial\psi_{T_\Phi^{i-1}(Z_{i-1,k})}(z_{i,k+1}). \tag{105}$$

Finally, the time integration of a generic equation of the MDI in (57) is given by

$$z_{i,k+1} - z_{i,k} - hz_{i+1,k+1} = \mu_{i,k+1} \in -\partial\psi_{T_\Phi^{i-1}(Z_{i-1,k})}(z_{i,k+1}) \quad (1 \leqslant i \leqslant r-1). \tag{106}$$

The last equation (58) is discretized in the same way as

$$\begin{aligned} &z_{r,k+1} - z_{r,k} - hCA^rW^{-1}z_{k+1} = CA^{r-1}B\ \mu_{r,k+1},\\ &\mu_{r,k+1} \in -\partial\psi_{T_\Phi^{r-1}(Z_{r-1,k})}(z_{r,k+1}). \end{aligned} \tag{107}$$

For the zero dynamics defined in (59), we use for the sake of simplicity[1] an Euler Backward scheme,

$$\xi_{k+1} - \xi_k - hA_\xi\xi_{k+1} - hB_\xi z_{1,k+1} = 0. \tag{108}$$

*Remark 22* (Extended Moreau's Time stepping (EMTS) scheme) The inclusions in (106), (107) and (108) define a numerical time integration of the Higher Order Sweeping Process $\mathbf{SP}(z_0; [0, T])$ that we call the Extended Moreau's Time Stepping (EMTS) scheme.

The following notation is used for the discretized variables. Let us denote the discretized state vector by

$$z_{k+1} = [z_{1,k+1}, \ldots, z_{r,k+1}, \xi_{k+1}^{\mathrm{T}}]^{\mathrm{T}} = [\bar{z}_{k+1}^{\mathrm{T}}, \xi_{k+1}^{\mathrm{T}}]^{\mathrm{T}},$$

the vector of discretized multipliers by $\mu_{k+1}$, i.e.

$$\mu_{k+1} = [\mu_{1,k+1}, \ldots, \mu_{r,k+1}]^{\mathrm{T}}.$$

[1] Depending on the regularity of $z_1$, a higher order scheme might be used for the time-integration of the zero dynamics.

Then the discrete-time system in (106), (107) and (108) can be rewritten compactly as (see (80) and (81))

$$z_{k+1} - z_k = hWAW^{-1}z_{k+1} + \bar{G}\mu_{k+1} \tag{109}$$

which is the discrete-time counterpart to (84).

**Comments** As we have seen earlier, the measures of the time interval $]t_k, t_{k+1}]$, i.e. $\mathrm{d}z(]t_k, t_{k+1}])$ and $\mu_{i,k+1} \overset{\Delta}{=} \mathrm{d}\nu_i(]t_k, t_{k+1}])$ are kept as primary variables. This fact ensures that the various terms handled by the numerical algorithm are of finite values. The use of differential measures of the time interval $]t_k, t_{k+1}]$ enables a rigorous treatment of the nonsmooth evolutions. When the time-step $h$ vanishes, it allows to deal with finite jumps. When the evolution is smooth, the scheme is equivalent to a backward Euler scheme. We can remark that nowhere a direct approximation of the density $z'_t$ with respect to the Lebesgue measure is made. The use of a first order algorithm is not chosen as usual through the approximation of the integral term (102) but required by the evaluation of the differential measure. As it has been observed in Remark 15, it is possible to incorporate a different jump mapping. This is also the case for the numerical time integration in approximating the jump mapping by

$$\mu_{i,k+1} \in -\partial\psi_{T_\Phi^{i-1}(Z_{i-1,k})}\left(\frac{z_{i,k+1} - e_{i+1}z_{i,k}}{1 + e_{i+1}}\right). \tag{110}$$

In the case of Lagrangian systems, the function $z_1$ is assumed to be absolutely continuous and the measure $\mathrm{d}\nu_1$ is chosen identically equal to zero. So in this case we have to take care that $\mu_1$ vanishes when the time step vanishes also. Another way is to choose $\mu_1 \equiv 0$. In this case, a slight numerical violation of the constraint, increasing with $r$ is expected.

## 5.3 Properties of the discrete-time extended sweeping process

We therefore consider here the discrete-time system in (106), (107) and (108). In this section, some important properties are shown which are thought to pave the way towards a convergence proof of the discrete-time solutions towards a solution of the continuous-time sweeping process. In the case $r = 2$ and $z_1(0) \geqslant 0$, such a convergence has been established in [42, Sect. 3.2]. The hardest part of the proof is not convergence towards some limit itself, but showing that the limit *is* a solution of the sweeping process. One discrepancy with respect to the works in [42, Sect. 3.2] [40,69] is that the well-posedness of the higher order sweeping process has already been proved in Corollary 2.

- In this section, we assume, as in Proposition 8, that:

$(H)$ The triple $(WAW^{-1}, \bar{G}, H)$ is observable, controllable and positive real.

### *5.3.1 A dissipation inequality*

Let us consider (106), (107) and (108), and the matrix $J$ in (82). We have:

**Proposition 11** *Suppose that condition* $(H)$ *holds. Then:*

$$\frac{1}{2}z_{k+1}^{\mathrm{T}}Jz_{k+1} - \frac{1}{2}z_k^{\mathrm{T}}Jz_k \leqslant -\frac{1}{2}(z_{k+1}-z_k)^{\mathrm{T}}J(z_{k+1}-z_k) + hz_{k+1}^{\mathrm{T}}JWAW^{-1}z_{k+1} \tag{111}$$

*for all* $0 \leqslant k \leqslant N-1$.

*Proof* Let us pre-multiply both sides of (109) by $z_{k+1}^{\mathrm{T}}J$. This gives

$$z_{k+1}^{\mathrm{T}}Jz_{k+1} - z_{k+1}^{\mathrm{T}}Jz_k = hz_{k+1}^{\mathrm{T}}JWAW^{-1}z_{k+1} + z_{k+1}^{\mathrm{T}}J\bar{G}\mu_{k+1}.$$

We have

$$z_{k+1}^{\mathrm{T}}Jz_{k+1} - z_{k+1}^{\mathrm{T}}Jz_k = \frac{1}{2}z_{k+1}^{\mathrm{T}}Jz_{k+1} - \frac{1}{2}z_k^{\mathrm{T}}Jz_k + \frac{1}{2}(z_{k+1}-z_k)^{\mathrm{T}}J(z_{k+1}-z_k) \tag{112}$$

and

$$hz_{k+1}^{\mathrm{T}}JWAW^{-1}z_{k+1} + z_{k+1}^{\mathrm{T}}J\bar{G}\mu_{k+1} \leqslant hz_{k+1}^{\mathrm{T}}JWAW^{-1}z_{k+1} \tag{113}$$

where we used the fact that $z_{k+1}^{\mathrm{T}}J\bar{G}\mu_{k+1} \leqslant 0$ as a consequence of the relations in (105) and (107) (recall that $(J\bar{G})^{\mathrm{T}} = (I_r \; 0_{(n-r)\times r})$). From (112) and (113) the result follows. □

### *5.3.2 Boundedness properties*

**Proposition 12** *Suppose that condition* $(H)$ *holds. There exists a constant* $\alpha > 0$ *such that for all* $h > 0$ *and all* $0 \leqslant k \leqslant N-1$, $||z_k|| \leqslant \alpha$. *Moreover, for any given* $h^* > 0$, *there exists a constant* $M \equiv M(h^*) > 0$ *such that* $||\bar{G}\mu_k|| \leqslant M, \forall h \in ]0, h^*[$.

*Proof* Notice that from (111), we get

$$\begin{aligned} z_{k+1}^{\mathrm{T}}Jz_{k+1} &\leqslant z_k^{\mathrm{T}}Jz_k + hz_{k+1}^{\mathrm{T}}(JWAW^{-1} + W^{-\mathrm{T}}A^{\mathrm{T}}W^{\mathrm{T}}J)z_{k+1} \\ &\quad -\frac{1}{2}(z_{k+1}-z_k)^{\mathrm{T}}J(z_{k+1}-z_k) \leqslant z_k^{\mathrm{T}}Jz_k, \end{aligned}$$

where we have used the facts that the matrix $JWAW^{-1} + W^{-\mathrm{T}}A^{\mathrm{T}}W^{\mathrm{T}}J$ is negative semi-definite and the matrix $J$ is positive definite. Proceeding by induction one concludes that

$$z_{k+1}^{\mathrm{T}}Jz_{k+1} \leqslant z_0^{\mathrm{T}}Jz_0 \quad (\forall\, 0 \leqslant k \leqslant N-1).$$

Let us denote by $\lambda_J > 0$ the smallest eigenvalue of the matrix $J$. We get

$$\lambda_J \|z_{k+1}\|^2 \leqslant z_0^{\mathrm{T}} J z_0 \quad (\forall\, 0 \leqslant k \leqslant N-1)$$

and the result is proved with $\alpha = \sqrt{\frac{z_0^{\mathrm{T}} J z_0}{\lambda_J}}$. From (109) one can choose $M = 2\alpha + h^* |||WAW^{-1}|||\alpha$. □

*5.3.3 Local bounded variation*

What follows is strongly inspired from [42, Lemma 2.5]. We first notice that since all the cones $T_\Phi^i(\cdot)$ are either $\mathbb{R}$ or $\mathbb{R}^+$, it follows that the closed ball $\bar{B}(a,R) = \{z \in \mathbb{R} :\ ||z-a|| \leqslant R\} \subset T_\Phi^i(\cdot)$ for any $a > 0$ and $R < \frac{a}{2}$. We define $z_i^N : [0,T[\rightarrow \mathbb{R}; t \mapsto z_i^N(t)$ as the step function given by $z_i^N(t) = z_{i,k}$ for all $t \in [t_k, t_{k+1}[$, $0 \leqslant k \leqslant N-1$ and $z_i^N(t_N) = z_{i,N}$, $1 \leqslant i \leqslant r$.

**Proposition 13** *Suppose that condition* $(H)$ *holds. The total variation of* $z_i^N$*,* $1 \leqslant n$*, in* $[0,T]$ *is bounded above according to:*

$$\begin{aligned}
&var(z_i^N,[0,T]) \leqslant \tfrac{1}{2R}|z_{i,0}-a|^2 + \tfrac{\alpha^2}{2R}T^2 + \alpha T(1+\tfrac{1}{R}|z_{i,0}-a|) \quad (1 \leqslant i \leqslant r-1)\\
&var(z_r^N,[0,T]) \leqslant \tfrac{1}{2R}|z_{r,0}-a|^2 + \tfrac{\beta^2\alpha^2}{2R}T^2 + \beta\alpha T(1+\tfrac{1}{R}|z_{1,0}-a|) \qquad (114)\\
&var(\xi^N,[0,T]) \leqslant (\gamma+\delta)\alpha T
\end{aligned}$$

*where* $|||CA^rW^{-1}||| \leqslant \beta$*,* $|||A_\xi||| \leqslant \gamma$ *and* $|||B_\xi||| \leqslant \delta$*, whereas* $\alpha$ *is as in Proposition 12. Moreover there exists a constant* $K > 0$ *such that for all* $N \in \mathbb{N}, N \geqslant 1$*:*

$$var(z^N,[0,T]) \leqslant K. \qquad (115)$$

*Proof* In this proof we suppress the $N$ in $z_i^N$ and $\xi^N$ to simplify the notation. Recall that

$$z_{i,k+1} - z_{i,k} - hz_{i+1,k+1} \in -\partial\psi_{T_\Phi^{i-1}(Z_{i-1,k})}(z_{i,k+1}) \quad (1 \leqslant i \leqslant r-1)$$

and

$$\begin{aligned}
z_{r,k+1} - z_{r,k} - hCA^rW^{-1}z_{k+1} &\in -CA^{r-1}B\partial\psi_{T_\Phi^{r-1}(Z_{r-1,k})}(z_{r,k+1})\\
&= -\partial\psi_{T_\Phi^{r-1}(Z_{r-1,k})}(z_{r,k+1})
\end{aligned}$$

where we have used the facts that $CA^{r-1}B > 0$ and that $\partial\psi_{T_\Phi^{r-1}(Z_{r-1,k})}(z_{r,k+1})$ is a closed cone. Then we may write

$$z_{i,k+1} = \mathrm{prox}[T_\Phi^{i-1}(Z_{i-1,k}); z_{i,k} + hz_{i+1,k+1}] \quad (1 \leqslant i \leqslant r-1)$$

and

$$z_{r,k+1} = \mathrm{prox}[T_{\Phi}^{r-1}(Z_{r-1,k}); z_{r,k} + hCA^{r}W^{-1}z_{k+1}].$$

It is then easy to see, that for $1 \leqslant i \leqslant r-1, 0 \leqslant k \leqslant N-1$, we have:

$$|z_{i,k+1} - a| \leqslant |z_{i,0} - a| + h(k+1)\alpha,$$

and

$$|z_{r,k+1} - a| \leqslant |z_{r,0} - a| + h(k+1)\beta\alpha.$$

Let us set $w_i = z_{i,0} + hz_{i+1,1}$ and $w_r = z_{r,0} + hCA^{r}W^{-1}z_1$. One has $|z_{i,1} - z_{i,0}| \leqslant |z_{i,1} - w_i| + |w_i - z_{i,0}|$ and using [42, Lemma 0.4.3] (see Appendix A), we get, for $1 \leqslant i \leqslant r-1$:

$$|z_{i,1} - z_{i,0}| \leqslant \frac{1}{2R}(|w_i - a|^2 - |z_{i,1} - a|^2) + h\alpha,$$

and

$$|z_{r,1} - z_{r,0}| \leqslant \frac{1}{2R}(|w_r - a|^2 - |z_{r,1} - a|^2) + h\beta\alpha.$$

More generally, for $0 \leqslant k \leqslant N-1$ and $1 \leqslant i \leqslant r-1$, we get the inequalities:

$$\begin{aligned}|z_{i,k+1} - z_{i,k}| &\leqslant \frac{1}{2R}(|z_{i,k} + hz_{i+1,k+1} - a|^2 - |z_{i,k+1} - a|^2) + h\alpha \\ &\leqslant \frac{1}{2R}(|z_{i,k} - a|^2 + h^2\alpha^2 + 2h\alpha\ |z_{i,k} - a| - |z_{i,k+1} - a|^2) + h\alpha.\end{aligned}$$

Moreover, for $0 \leqslant k \leqslant N-1$, we have also:

$$\begin{aligned}|z_{r,k+1} - z_{r,k}| &\leqslant \frac{1}{2R}(|z_{r,k} + hCA^{r}W^{-1}z_{k+1} - a|^2 - |z_{r,k+1} - a|^2) + h\alpha\beta. \\ &\leqslant \frac{1}{2R}(|z_{r,k} - a|^2 + h^2\beta^2\alpha^2 + 2h\beta\alpha\ |z_{r,k} - a| - |z_{r,k+1} - a|^2) + h\beta\alpha.\end{aligned}$$

We know that $|z_{i,k} - a| \leqslant |z_{i,0} - a| + hk\alpha$ for all $1 \leqslant i \leqslant r-1$, and $|z_{r,k} - a| \leqslant |z_{r,0} - a| + hk\beta\alpha$, and setting $T_k = kh$ (so that $T = T_N = Nh$), for $1 \leqslant i \leqslant r-1$ we get:

$$\begin{aligned}|z_{i,k+1} - z_{i,k}| \leqslant \frac{1}{2R}(|z_{i,k} - a|^2 - |z_{i,k+1} - a|^2) + \alpha h\left(1 + \frac{1}{R}|z_{i,0} - a|\right) \\ + \frac{\alpha^2}{2R}(h^2 + 2hT_k).\end{aligned} \tag{116}$$

We have also:

$$|z_{r,k+1}-z_{r,k}| \leqslant \frac{1}{2R}(|z_{r,k}-a|^2-|z_{r,k+1}-a|^2) \\ +\alpha\beta h\left(1+\frac{1}{R}|z_{r,0}-a|\right)+\frac{\alpha^2\beta^2}{2R}(h^2+2hT_k). \qquad (117)$$

We claim that

$$\sum_{k=0}^{j-1}|z_{i,k+1}-z_{i,k}| \leqslant \frac{1}{2R}(|z_{i,0}-a|^2-|z_{i,j}-a|^2)+\alpha T_j\left(1+\frac{1}{R}|z_{i,0}-a|\right)+\frac{\alpha^2}{2R}T_j^2 \qquad (118)$$

for $1 \leqslant i \leqslant r-1$, and

$$\sum_{k=0}^{j-1}|z_{r,k+1}-z_{r,k}| \leqslant \frac{1}{2R}(|z_{r,0}-a|^2-|z_{r,j}-a|^2)+\alpha\beta T_j\left(1+\frac{1}{R}|z_{r,0}-a|\right) \\ +\frac{\alpha^2\beta^2}{2R}T_j^2. \qquad (119)$$

We prove the result by induction. The result holds for $j=1$ as a direct consequence of (116) and (117) with $k=0$. Suppose now that (118) holds at step $j$. We claim that (118) holds at step $j+1$. Indeed, using again (116), we get for step $j+1$:

$$\begin{aligned}\sum_{k=0}^{j}|z_{i,k+1}-z_{i,k}| &= \left(\sum_{k=0}^{j-1}|z_{i,k+1}-z_{i,k}|\right)+|z_{i,j+1}-z_{i,j}| \\ &\leqslant \frac{1}{2R}(|z_{i,0}-a|^2-|z_{i,j}-a|^2)+\alpha T_j\left(1+\frac{1}{R}|z_{i,0}-a|\right) \\ &\quad +\frac{\alpha^2}{2R}T_j^2+\frac{1}{2R}(|z_{i,j}-a|^2-|z_{i,j+1}-a|^2) \\ &\quad +\alpha h\left(1+\frac{1}{R}|z_{i,0}-a|\right)+\frac{\alpha^2}{2R}(h^2+2hT_j) \\ &= \frac{1}{2R}(|z_{i,0}-a|^2-|z_{i,j+1}-a|^2) \\ &\quad +\alpha T_{j+1}\left(1+\frac{1}{R}|z_{i,0}-a|\right)+\frac{\alpha^2}{2R}T_{j+1}^2\end{aligned}$$

where we have used the fact that $T_{j+1}=T_j+h$ and $T_{j+1}^2=T_j^2+h^2+2hT_j$. The same approach can be used to prove (119). The first two inequalities in (114) are then deduced. The third inequality follows from (108) from which one deduces that

$$||\xi_{k+1}-\xi_k|| \leqslant (\gamma+\delta)\alpha h \qquad (120)$$

The result in (115) is a consequence of the results obtained in (114). □

Consider the step function $\mu^N : [0, T] \to \mathbb{R}^r; t \mapsto \mu^N(t)$ such that $\mu^N(t) = \mu_{k+1}$ for all $t \in [t_k, t_{k+1}[$ $(0 \leqslant k \leqslant N-1)$ and $\mu^N(t_N) = \mu_N$.

**Proposition 14** *Suppose that condition* $(H)$ *holds. For any given* $h^* > 0$*, there exists a constant* $K' \equiv K'(h^*) > 0$ *such that:*

$$var(\mu^N, [0, T]) \leqslant K', \quad \forall\, h \in ]0, h^*[. \tag{121}$$

*Proof* Let $h^* > 0$ be given. From (109) we deduce that

$$\begin{aligned}\mu_{k+1} - \mu_k = G^{-1}(I_r\ \ 0_{r\times(n-r)})(I_n - hWAW^{-1})(z_{k+1} - z_k)\\ +G^{-1}(I_r\ \ 0_{r\times(n-r)})(z_k - z_{k-1})\end{aligned} \tag{122}$$

It follows that

$$\begin{aligned}\sum_{k=1}^{N-1} |||\mu_{k+1} - \mu_k||| &\leqslant |||G^{-1}(I_r\ 0_{r\times(n-r)})(I_n - hWAW^{-1})||| \sum_{k=1}^{N-1} ||z_{k+1} - z_k||\\ &+|||G^{-1}(I_r\ 0_{r\times(n-r)})||| \sum_{k=0}^{N-2} ||z_{k+1} - z_k||\\ &\leqslant |||G^{-1}(I_r\ 0_{r\times(n-r)})(I_n - hWAW^{-1})||| \sum_{k=0}^{N-1} ||z_{k+1} - z_k||\\ &+|||G^{-1}(I_r\ 0_{r\times(n-r)})||| \sum_{k=0}^{N-1} ||z_{k+1} - z_k|| \qquad (123)\\ &\leqslant |||G^{-1}(I_r\ 0_{r\times(n-r)})|||(2 + h^*|||WAW^{-1}|||) \sum_{k=0}^{N-1} ||z_{k+1} - z_k||.\end{aligned} \tag{124}$$

It follows from (115) that $K'$ exists that depends on $K$ and the system parameters, such that (121) is satisfied. □

*Remark 23* i) Let us study the behavior of $z_{1,k}$. One peculiarity of the inclusion (106) is that $z_{1,k}$ =prox$[\mathbb{R}^+; z_{1,k-1}+hz_{2,k}] \geqslant 0$ for all $k \geqslant 1$. In other words $z_1^N(\cdot)$ may be negative only on $t \in [0, t_1[$. On the contrary, the other variables $z_i^N(\cdot)$ may become negative at any time $t \geqslant 0$. Consider for instance $T_\Phi^{i-1}(z_{1,k}, \ldots, z_{i-1,k}) = \mathbb{R}$ for all $2 \leqslant i \leqslant r$, so that $z_{2,k} = z_{2,k-1} + h\sum_{i=3}^r z_{i,k} + hCA^rW^{-1}z_k$. Nothing hampers that $\xi_k$ takes a value such that $z_{2,k} < 0$ even if $z_{i,k-1} > 0$ for all $2 \leqslant i \leqslant r$.

ii) It follows from (106), (107) that

$$0 \leqslant z_{1,k+1} \perp \mu_{1,k+1} \geqslant 0 \quad \text{for all } \; k \geqslant 0 \tag{125}$$

and

$$0 \leqslant z_{1,k+1} \perp \mu_{r,k+2} \geqslant 0 \quad \text{for all } \; k \geqslant 0 \tag{126}$$

which is the discrete-time complementarity condition corresponding to (72) and (76).

### *5.3.4 Convergence of the discretized solution*

We now denote $\{z^N\}$ the sequence of functions constructed from the functions $z^N(\cdot)$, and similarly for $\mu^N$.

**Proposition 15** *Suppose that condition* (*H*) *holds. There exists a subsequence* $\{z^{N_k}\}$ *of* $\{z^N\}$ *which converges pointwise to some function* $z : [0,T] \to \mathbb{R}^n$*, such that var*$([0,T]) \leqslant K$*, and a subsequence* $\{\mu^{N_k}\}$ *of* $\{\mu^N\}$ *which converges pointwise to some function* $\mu(\cdot) : [0,T] \to \mathbb{R}^r$ *such that var*$([0,T]) \leqslant K'$.

*Proof* The function $z^N(\cdot)$ is uniformly bounded on $[0,T]$, and it has bounded variation on $[0,T]$, see Propositions 12 and 13. From [42, Theorem 0.2.1 (i), (6)] the result follows. The proof is the same for $\mu^N(\cdot)$, using Propositions 12 and 14. □

*Remark 24* The convergence of $\mu^N$ towards a **LBV** *function* reflects the fact that, as said in the introduction of Sect. 5.2, the primary variables are $\mu_{i,k+1} \overset{\Delta}{=} \mathrm{d}v_i(]t_k, t_{k+1}])$. Hence the Dirac measures do not appear in the limit $\mu(\cdot)$ which is by construction a (bounded) function.

**Proposition 16** *Suppose that condition* (*H*) *holds. If* $z$ *is right-continuous, then for every continuous function of bounded variation* $\varphi : [0,T] \to \mathbb{R}$ *we have:*

$$\int\limits_{]s,t]} \varphi \, \mathrm{d}z_i^{N_k} \to \int\limits_{]s,t]} \varphi \, \mathrm{d}z_i \;\; (s < t) \quad \textit{as } N_k \to +\infty \;\; (1 \leqslant i \leqslant r). \tag{127}$$

*Proof* This is a consequence of [42, Theorem 0.2.1 (iii)] and the fact that the $z_i^N(\cdot)$ are right continuous functions. □

The differential measures $\mathrm{d}z_i^{N_k}$ are of the form $\mathrm{d}z_i^{N_k} = \{\dot{z}_i^{N_k}\}(t)\mathrm{d}t + \sum_{k=1}^{N-1}(z_{i,k+1} - z_{i,k})\delta_{t_{N,k}}$(the singular part being zero since $z_i^{N_k}(\cdot)$ is a step function). Since the $z_i$'s are **LBV**, $\mathrm{d}z_i$ admits a similar decomposition (see Sect. 2), with atoms at times $\tau_k$.

The proof that the limit functions are solutions of the time-continuous sweeping process (57)–(58)–(59) is left as a future work. Examples 8 and 9 in Sect. 5.5.1 demonstrates however the fact that the solutions of (106), (107) and (108) do converge to the higher order sweeping process solutions in simple cases.

## 5.4 Overview of the implementation

In this section, we provide a short overview of the implementation of the EMTS. Our goal is to present the algorithm in a pedagogical way through a straightforward implementation. Obviously, for efficiency reasons, the code may slightly differ from what is below.

**Matrix formulation of the ZD form in view of numerical integration** Given $W \in \mathbb{R}^{n\times n}$, the linear transformation of the state space, we introduce the following matrix notation,

$$[I - h\bar{A}]z_{k+1} = z_k + \bar{G}\mu_{k+1}, \tag{128}$$

where the matrices $\bar{A} \in \mathbb{R}^{n\times n}$ and $\bar{G} \in \mathbb{R}^{n\times r}$ are defined as follows:

$$\bar{A} = \left[\begin{array}{ccccc|c}
0 & 1 & 0 & \dots & 0 & \\
\vdots & 0 & \ddots & \ddots & \vdots & 0 \\
\vdots & \vdots & \ddots & \ddots & 0 & \\
0 & 0 & \dots & 0 & 1 & \\ \hline
\multicolumn{6}{c}{CA^{r}W^{-1}} \\ \hline
\multicolumn{1}{c|}{B_\xi} & \multicolumn{4}{c|}{0} & A_\xi
\end{array}\right], \quad
\bar{G} = \left[\begin{array}{ccccc}
1 & 0 & \dots & 0 & 0 \\
0 & 1 & \ddots & \vdots & \vdots \\
\vdots & \ddots & \ddots & 0 & \vdots \\
0 & \dots & 0 & 1 & 0 \\
0 & \dots & \dots & 0 & CA^{r-1}B \\ \hline
\multicolumn{5}{c}{0_{(n-r)\times r}}
\end{array}\right]. \tag{129}$$

Notice that $\bar{A} = WAW^{-1}$ is defined in (25) and that $\bar{G}$ is defined by (80).

**Expression of the inclusions (105) in terms of nested complementarity problems** Let us consider the following inclusion,

$$\mu_{1,k+1} \in -\partial\psi_{\Phi}(z_{1,k+1}).$$

Let us set $\Phi = \mathbb{R}^{+}$. This inclusion may be stated equivalently as a complementarity problem,[2]

$$0 \leqslant z_{1,k+1} \perp \mu_{1,k+1} \geqslant 0.$$

If $r > 1$ then we must handle the second inclusion,

$$\mu_{2,k+1} \in -\partial\psi_{T^{1}_{\Phi}(z_{1,k})}(z_{2,k+1})$$

[2] In a more general setting, a cone complementarity problem has to be written $\Phi \ni z_{1,k+1} \perp -\mu_{1,k+1} \in \Phi^{\star}$, with $\Phi^{\star}$ the dual cone of $\Phi$.

which can be reformulated in terms of a complementarity problem,

$$\text{If } z_{1,k} \leqslant 0, \quad \text{then } 0 \leqslant z_{2,k+1} \perp \mu_{2,k+1} \geqslant 0.$$

In this way, for $r > 2$, we get the following complementarity problem,

$$\text{If } z_{1,k} \leqslant 0 \text{ and } z_{2,k} \leqslant 0, \quad \text{then } 0 \leqslant z_{3,k+1} \perp \mu_{3,k+1} \geqslant 0$$

In the general case, we search the integer $1 \leqslant r^\star \leqslant r$ satisfying the condition,

$$r^\star = \begin{cases} 1, & \text{if } z_{1,k} > 0 \\ 1 + \max\{j \leqslant r-1 : \; z_{i,k} \leqslant 0, \forall\, i \leqslant j\}, \end{cases}$$

yielding the following set of nested complementarity problems,

$$0 \leqslant z_{i,k+1} \perp \mu_{i,k+1} \geqslant 0, \quad 1 \leqslant i \leqslant r^\star. \tag{130}$$

We define the vectors collecting the state and the multiplier for the "active" constraints by

$$\begin{aligned} z^\star_{k+1} &= [z_{1,k+1}, \ldots, z_{r^\star,k+1}]^{\mathrm{T}}, \\ \mu^\star_{k+1} &= [\mu_{1,k+1}, \ldots, \mu_{r^\star,k+1}]^{\mathrm{T}}. \end{aligned} \tag{131}$$

and we introduce the matrix $R \in \mathbb{R}^{r^\star \times r}$ describing the relation between $z^\star_{k+1}$ and $\bar{z}_{k+1}$ :

$$z^\star_{k+1} = R\bar{z}_{k+1}. \tag{132}$$

Assuming that $\mu_{i,k+1} = 0, i > r^\star$, we get the relation between $\mu_{i,k+1}$ and $\mu^\star_{k+1}$ :

$$\mu_{k+1} = R^{\mathrm{T}} \mu^\star_{k+1}. \tag{133}$$

**Formulation of the one-step LCP problem** To be more explicit in the computation of the state vector, we introduce the matrix $P \in \mathbb{R}^{r \times n}$ such that

$$\bar{z}_{k+1} = Pz_{k+1}. \tag{134}$$

Assuming that $r^\star$ is computed at each step and that $z_k$ is known, the following set of discretized equations has to be solved to advance from step $k$ to step $k+1$:

$$\begin{cases} [I-h\bar{A}]z_{k+1} = z_k + \bar{G}\mu_{k+1} \\ \bar{z}_{k+1} = Pz_{k+1} \\ z^\star_{k+1} = R\bar{z}_{k+1} \\ \mu_{k+1} = R^{\mathrm{T}}\mu^\star_{k+1} \\ 0 \leqslant \mu^\star_{k+1} \perp z^\star_{k+1} \geqslant 0. \end{cases} \tag{135}$$

This yields the following closed form for the one-step LCP problem :

$$\begin{cases} z^\star_{k+1} = RP[I-h\bar{A}]^{-1}z_k + RP[I-h\bar{A}]^{-1}\bar{G}R^{\mathrm{T}}\mu^\star_{k+1} \\ 0 \leqslant \mu^\star_{k+1} \perp z^\star_{k+1} \geqslant 0. \end{cases} \tag{136}$$

**Pseudo-algorithm for the EMTS** A pseudo-algorithm for the EMTS is given in order to clearly outline the major features of the numerical resolution of the EMTS (see Algorithm 1).

*Remark 25* The interest of the ZD canonical form from the numerical point of view partly lies in the consistency of the resulting scheme. This consistency property can be illustrated with the limit value of the LCP matrix in (136) when the time-step $h$ vanishes. It is clear that if $h$ vanishes, then the LCP matrix is close to the matrix

$$RP\bar{G}R^{\mathrm{T}} = RGR^{\mathrm{T}} = \begin{cases} I_{r^\star} & \text{if } 1 \leqslant r^\star < r \\ G & \text{if } r^\star = r \end{cases} \tag{137}$$

which is the LCP matrix of the time-continuous ZD form given for instance by the inclusions (68) and (69). By continuity argument on the time-step $h$, the LCP matrix in (136) inherits from $G$ and $I_r$ the properties of definiteness and positiveness.

*Remark 26* (The multivariable case) If we consider the multivariable case with a vector relative degree $\bar{r}$, the discrete approximation of the dynamics (128) continues to hold with the dimensions defined in Remark 3. In the same way, the matrix $P$ defined in (134) can be obviously extended to the multivariable case. The case of the operator $R$ is a little bit more tricky. For each constraint, we need to compute the rank, $r^\star_l, 1 \leqslant l \leqslant m$ which selects the variables concerned by the LCP:

$$r^\star_l = \begin{cases} 1, \quad \text{if } z_{1,l,k} > 0 \\ 1 + \max\{j \leqslant r-1 : \; z_{i,l,k} \leqslant 0, \forall\, i \leqslant j\}, \;\; 1 \leqslant l \leqslant m \end{cases}$$

where $z_{1,l,k}$ denotes the $l$-component of the m-dimensional vector $z_{i,k}$.

We define the vectors collecting the state and the multiplier for the "active" constraints by:[3]

$$
\begin{aligned}
z^{\star}_{k+1} &= [z_{1,1,k+1},\dots,z_{r^{\star}_1,1,k+1},\dots,z_{1,l,k+1},\dots,z_{r^{\star}_l,l,k+1},\\
&\qquad \dots z_{1,m,k+1},\dots,z_{r^{\star}_m,m,k+1}]^{\mathrm{T}}\\
\mu^{\star}_{k+1} &= [\mu_{1,1,k+1},\dots,\mu_{r^{\star}_1,1,k+1},\dots,\mu_{1,l,k+1},\dots,\mu_{r^{\star}_l,l,k+1},\\
&\qquad \dots \mu_{1,m,k+1},\dots,\mu_{r^{\star}_m,m,k+1}]^{\mathrm{T}}.
\end{aligned}
\tag{138}
$$

and we introduce the matrix $R = [I_{r^{\star}} \;\; 0_{r-r^{\star}}] \in \mathbb{R}^{r^{\star}\times r}$ describing the relation between $z^{\star}_{k+1}$ and $\bar{z}_{k+1}$ :

$$
z^{\star}_{k+1} = R\bar{z}_{k+1}, \quad \mu_{k+1} = R^{\mathrm{T}}\mu^{\star}_{k+1}.. \tag{139}
$$

With this convention, the EMTS defined by (135) and (139) can be extended to the multivariable case with a vector relative degree $\bar{r}$. It is noteworthy that the limit LCP matrix when the time step $h$ vanishes will be the identity matrix if $r^{\star}_l < r, 1 \leqslant l \leqslant m$ and shares the properties (P-matrix, positiveness) of the matrix $CA^{r-1}B$ in other cases due to the structure of the selection matrix $R$.

## 5.5 Numerical applications

### *5.5.1 Comparison with a backward Euler scheme*

Several attempts have been already made to solve numerically dynamical systems with arbitrary relative degree. In [28], an algorithm for constructing solutions rather than a pure numerical scheme, is given based on an event-driven strategy. This type of strategy cannot encompass general evolutions with accumulation of events, and is not well suited for convergence proofs. In [16], the direct use of a backward Euler scheme with an implicit evaluation of the complementarity condition yields a time-stepping scheme, which is very similar to the "catching up algorithm" of Moreau. This scheme works well with systems of relative degree less or equal to 1, but exhibits characteristic examples of inconsistency for relative degree $\geqslant 2$. We have collected in this section some remarks on the behavior of the numerical scheme presented above which lead us to believe that the EMTS scheme solves this problem.
For the first and the second order sweeping process, the time integration method is often confused with a standard backward Euler scheme. To highlight the difference with the numerical time integration of the Moreau's sweeping process, we consider several examples of inconsistencies, some of which are introduced in [16]. A naive way of integrating an LCS is to apply directly a backward Euler scheme:

[3] The reordering is only introduced for sake of simplicity in notation.

$$\begin{cases} \dfrac{x_{k+1}-x_k}{h} = Ax_{k+1} + B\lambda_{k+1} \\ w_{k+1} = Cx_{k+1} + D\lambda_{k+1} \\ 0 \leqslant \lambda_{k+1} \perp w_{k+1} \geqslant 0 \end{cases} \tag{140}$$

which can be reduced to an LCP by a straightforward substitution:

$$0 \leqslant \lambda_{k+1} \perp C(I-hA)^{-1}x_k + (hC(I-hA)^{-1}B + D)\lambda_{k+1} \geqslant 0. \tag{141}$$

In the sequel, such an LCP will be denoted as $(w_{k+1}, \lambda_{k+1}) = LCP(M, b_{k+1})$ where

$$M = hC(I-hA)^{-1}B + D \tag{142}$$

$$b_{k+1} = C(I-hA)^{-1}x_k. \tag{143}$$

In [16], some consistency and convergence results are proved. Shortly, under the assumption that $D$ is nonnegative definite or that the triplet $(A, B, C)$ is a observable and controllable and $(A, B, C, D)$ is positive real, they exhibit that some subsequences of $\{w_k\}, \{\lambda_k\}, \{x_k\}$ converge weakly to a solution $w, \lambda, x$ of the LCS. Such assumptions imply that the relative degree $r$ is less or equal to 1. In the case of the relative degree 0, the LCS is equivalent to a standard system of ordinary differential equations with a Lipschitz-continuous vector field (see [23], Remark 10). The result of convergence is then the standard result of convergence for the Euler backward scheme. In the case of a relative degree equal to 1, these results corroborate the results of [42,45,48].

*Remark 27* As in Remark 25, the consistency of the LCP $(w_{k+1}, \lambda_{k+1}) = LCP(M, b_{k+1})$ can be analyzed with respect to the time-continuous LCP. Clearly, if $r \geqslant 1$, i.e. $D = 0$, the limit value of the LCP matrix $M$ in (142) is equal to zero when the time-step $h$ vanishes. Contrary to the EMTS scheme, the consistency of the time-stepping scheme can not be retrieved for the case $r \geqslant 1$.

In the case $r = 1$ ($D = 0$, $CB \neq 0$), the state may jump at the initial instant and the multiplier $\lambda$ possesses an atom. Knowing this fact, we can choose as a primary variable of the LCP the value $\Lambda_{k+1} = h\lambda_{k+1}$ which approximates the amplitude of the distribution rather than the value of a function. The resulting LCP is then $(w_{k+1}, \Lambda_{k+1}) = LCP(\frac{1}{h}M, b_{k+1})$. With this change of variable, we recover the EMTS scheme with its consistency. The multiplier $\Lambda_{k+1}$ is finite when $h$ vanishes and the new LCP matrix tends towards $CB$, which is consistent with the time-continuous formulation.

Unfortunately, this trick can no longer apply for the case $r = 2$ ($D = 0, CB = 0, CAB \neq 0$) (see the example below). If we apply the time discretization given by (140) with the new variable $\Lambda_{k+1}$, we can remark that

$$\lim_{h \longrightarrow 0} \frac{1}{h}M = \lim_{h \longrightarrow 0} C(I-hA)^{-1}B = 0. \tag{144}$$

There is no way to obtain a consistent time-stepping scheme in this way, compromising the chance to prove convergence results. In many practical cases, it is clear that if $h$ is taken very small, the LCP matrix in (141) has little chance to be well conditioned due to the fact that $CB = 0$.

As we said earlier, several examples for which the backward Euler scheme does not work at all are also detailed in [16]. These systems are of higher relative degree. We will consider below two similar examples and comment on the difference between the backward Euler scheme and our approach.

*Example 8* Let us consider an LCS with the following matrix definition:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 \; 0 \; 0 \end{pmatrix}, \quad D = 0. \tag{145}$$

The relative degree $r$ of this LCS is equal to 2 ($D = 0, CB = 0, CAB \neq 0$). If we consider the initial data $x_0 = (0, -1, 0)^{\mathrm{T}}$, we obtain by a straightforward application of the scheme (140) the following solution:

$$x_k = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \forall\, k \geqslant 1, \tag{146}$$

$$\lambda_1 = \frac{1}{h}, \quad \lambda_k = 0, \ \forall\, k \geqslant 2. \tag{147}$$

We can remark that the multiplier $\lambda_1$ which is the solution of the LCP at the first step, tends towards $+\infty$ when $h$ vanishes. In this example, the state $x$ seems to be well approximated but both the LCP matrix and the multiplier tend to inconsistent values when $h$ vanishes. This inconsistency is just the result of an attempt to approximate the point value of a distribution, which is nonsense.

If we consider now the initial data $x_0 = (-1, -1, 0)^{\mathrm{T}}$, we obtain the following numerical solution from (140) :

$$x_k = \begin{pmatrix} k \\ \frac{1}{h} \\ 0 \end{pmatrix}, \quad \forall\, k \geqslant 1, \tag{148}$$

$$\lambda_1 = \frac{1}{h^2}, \quad \lambda_k = 0, \quad \forall\, k \geqslant 2. \tag{149}$$

With such an initial data, the exact solution should be $x_k = 0, \forall\, k \geqslant 1$. We can see that there is an inconsistency in the result because the first component of the approximate state does not depend on the time-step. We cannot expect that this approximation converges to the exact solution. If we apply the EMTS scheme, we obtain the following solution:

$$x_k = \begin{pmatrix}0\\0\\0\end{pmatrix}, \quad \forall\, k \geqslant 1 \tag{150}$$

$$\mu_{1,1} = 1, \quad \mu_{2,1} = 1, \tag{151}$$

$$\mu_{1,k} = 0, \quad \mu_{2,k} = 0, \quad \forall\, k \geqslant 2, \tag{152}$$

which converges to the time-continuous solution of the higher order Moreau's sweeping process, i.e. $x(0) = x_0, x(t) = (0,0,0)^{\mathrm{T}}, \forall\, t > 0$.

*Example 9* Let us consider another simple example:

$$A = \begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix}, \quad B = \begin{pmatrix}0\\0\\1\end{pmatrix}, \quad C = \begin{pmatrix}1&0&0\end{pmatrix}, \quad D = 0. \tag{153}$$

In this case, the relative degree $r$ is equal to 3. The direct discretization of the system leads to the same problem as in the previous example even in the case where the initial data satisfies the constraints. Let us consider $x_0 = (0,-1,0)^{\mathrm{T}}$. From (140), we obtain the following numerical solution:

$$x_k = \begin{pmatrix}\dfrac{k(k+1)}{2h}\\ k\\ \dfrac{1}{h}\end{pmatrix}, \quad \forall\, k \geqslant 1, \tag{154}$$

$$\lambda_1 = \frac{1}{h^2}; \quad \lambda_k = 0, \quad \forall k \geqslant 2. \tag{155}$$

This solution can not converge to an analytical solution. The solution given by the EMTS scheme is

$$x_k = \begin{pmatrix}0\\0\\0\end{pmatrix}, \quad \forall\, k \geqslant 1 \tag{156}$$

$$\mu_{1,1} = 1; \quad \mu_{2,1} = 1, \quad \mu_{3,1} = 0, \tag{157}$$

$$\mu_{i,k} = 0, \quad \forall\, k \geqslant 2, \; i = 1, \ldots 3, \tag{158}$$

which is the time-continuous solution of the higher order Moreau's sweeping process, i.e. $x(0) = x_0, x(t) = (0,0,0)^{\mathrm{T}}, \forall\, t > 0$.

### *5.5.2 Influence of the zero-dynamics on the solution*

In this example, we illustrate how the zero dynamics may influence the behavior of the following system

$$\begin{cases} \dot{z}_1(t) = z_2(t) \\ \dot{z}_2(t) = z_3(t) \\ \dot{z}_3(t) = -z_1(t) - z_2(t) - z_3(t) - d_\xi^{\mathrm{T}} \xi(t) + \lambda(t) \\ \dot{\xi}_1(t) = \alpha \xi_2(t) \\ \dot{\xi}_2(t) = -\omega \xi_1(t) + z_1(t) \\ w(t) = z_1(t) \geqslant 0 \end{cases} \tag{159}$$

with the initial condition $z(0) = (1,0,0,0,0)^{\mathrm{T}}$. The system (159) of relative degree $r = 3$, is embedded in the MDI formalism (57)–(59). All the simulations are performed with Scilab©. The LCP is numerically solved with either the Lemke's algorithm or an iterative splitting method (a Gauss–Seidel like algorithm). The time interval is $[0,10]$ and the time step is equal to $h = 10^{-2}$.

In first experiment depicted on Fig. 2, we choose $d_\xi^{\mathrm{T}} = (0,1)$ but with a trivial zero dynamics $\alpha = 0$ and $\omega = 0$. The function $\xi_2(t)$ is just a time integration of the function $z_1(t)$, which has to remain non negative. The $\bar{z}$-dynamics is identically equal to zero after the first jump and remains at zero due to the action of the zero-dynamics which pushes $z_3(t)$ on the constraint. In this case, the multiplier $d\nu_3$ possess a function part $\chi_3(t)$ which is not equal to zero on a non trivial interval.

We choose now a zero dynamics, which is a pure harmonic oscillator $\alpha = 1$ and $\omega = 1$ in order to understand what its influence on the $\bar{z}$-dynamics may be. The first case is given by $d_\xi^{\mathrm{T}} = (0,0)$ where the zero-dynamics is decoupled from the $\bar{z}$-dynamics, i.e., the zero dynamics does not play any role in the $\bar{z}$-dynamics. The results are depicted on Fig. 3. The system seems to be stable and the state vanishes. Choosing $d_\xi^{\mathrm{T}} = (0,1)$, the same simulation is made with a coupled zero dynamics. The influence of the zero dynamics is shown on Fig. 4. Due to the fact that $z_{2,k}$ and $z_{3,k}$ are negative at each state jumps, the whole vector $\bar{z}$

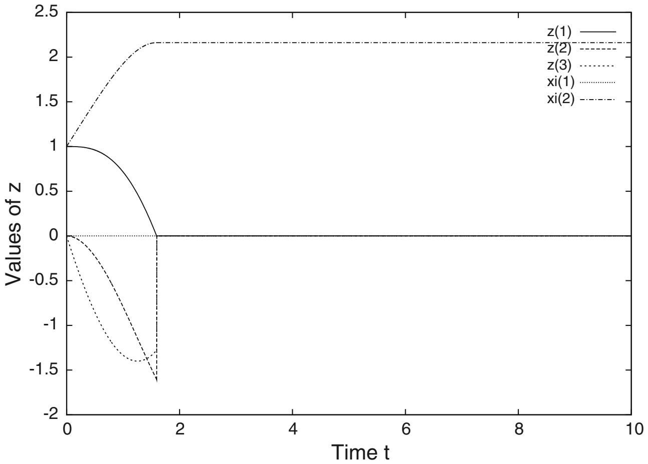


**Fig. 2** EMTS scheme $d_\xi^{\mathrm{T}} = (0,1)$, $\alpha = 0$ and $\omega = 0$

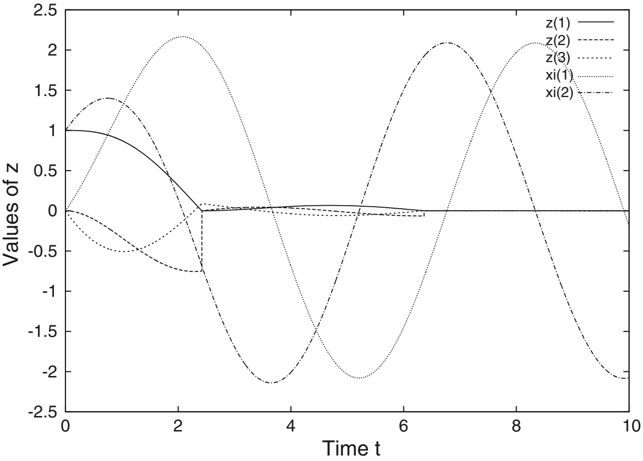


**Fig. 3** EMTS scheme $d_\xi^{\mathrm{T}} = (0,0)$, $\alpha = 1$ and $\omega = 1$. Decoupled zero-dynamics

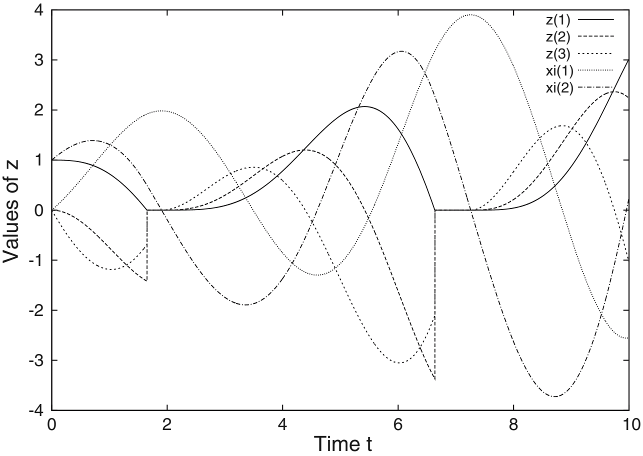


**Fig. 4** EMTS scheme $d_\xi^{\mathrm{T}} = (0,1)$, $\alpha = 1$ and $\omega = 1$

is reinitialized at zero and remains stuck at zero up to the change of sign of the function $\xi_2(t)$. As in the first experiment, there are non-zero intervals on which the constraints remain active and then the function $\chi_3(t)$ is not equal to zero on such intervals. Finally, if we choose $d_\xi = (0,-1)$, we observe on Fig. 5 the same kind of behavior.

It is therefore clear that the zero-dynamics matrix $A_\xi$ and the connection vector $d_\xi$, have a strong influence on the system's dynamics. In particular, one sees that despite the state reinitialization mapping sets the post-impact states $\bar{z}$ to zero (plastic impacts), the zero dynamics may force the system to detach from the constraint and undergo subsequent jumps after a boundary trajectory has occurred.

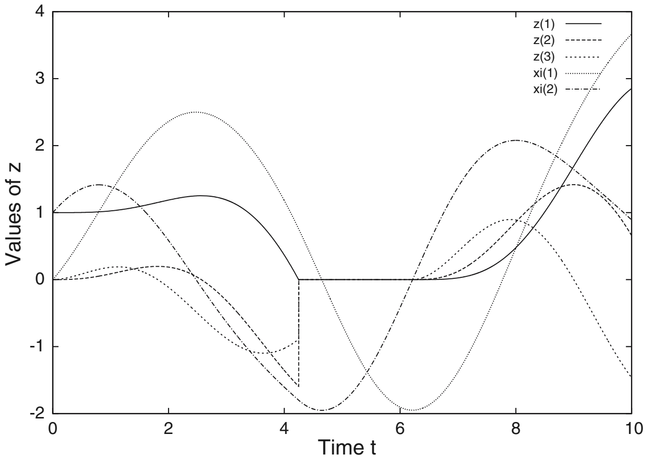


**Fig. 5** EMTS scheme $d_\xi^{\mathrm{T}} = (0, -1)$, $\alpha = 1$ and $\omega = 1$

### *5.5.3 Empirical order of the scheme*

This section is devoted to provide an empirical estimation of the order of the scheme. Let us consider the previous example with $d_\xi = (0, -1), \alpha = 1, \omega = 1$ and with the initial condition $z(0) = (1, 0, 0, 0, 0)^{\mathrm{T}}$.

In order to evaluate the order of accuracy of the scheme on this simple example, we need to use a norm which is consistent with the set of **RCLBV** functions and to introduce a notion of convergence providing a reasonable substitute to the uniform convergence of the continuous functions. To overcome this difficulty, the convergence in the sense of filled-in graph has been introduced by Moreau [49]. Shortly, for a **RCLBV** function $f : [0, T] \mapsto \mathbb{R}^n$, we define the filled-in graph, $gr^\star(f)$ by adding some line segments to the graph of $f$ in such a way that all the gaps are filled:

$$gr^\star(f) = \{(t, x) \in [0, T] \times \mathbb{R}^n, 0 \leqslant t \leqslant T \text{ and } x \in [f(t^-), f(t^+)])\}. \tag{160}$$

Such graphs are closed bounded subsets of $[0, T] \times \mathbb{R}^n$, hence, we can use the Hausdorff distance between two such sets with a suitable metric:

$$\mathrm{d}((t, x), (s, y)) = \max\{|t - s|, \|x - y\|\}. \tag{161}$$

Defining the excess of separation between two graphs by

$$e(gr^\star(f), gr^\star(g)) = \sup_{(t,x) \in gr^\star(f)} \inf_{(s,y) \in gr^\star(g)} \mathrm{d}((t, x), (s, y)), \tag{162}$$

the Hausdorff distance between two filled-in graphs $h^\star$ is defined by

$$h^\star(gr^\star(f), gr^\star(g)) = \max\{e(gr^\star(f), gr^\star(g)), e(gr^\star(g), gr^\star(f))\}. \tag{163}$$

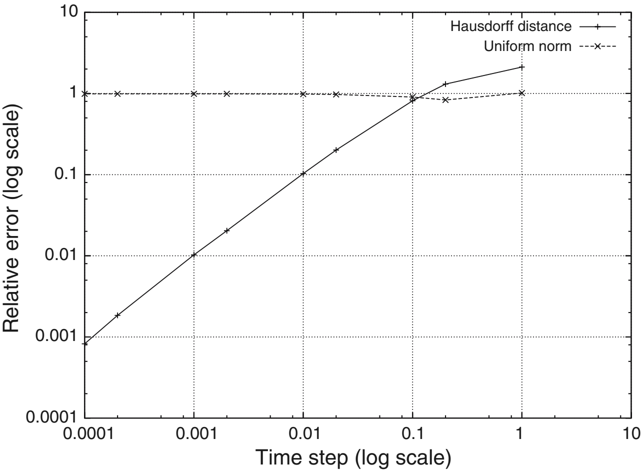


**Fig. 6** Empirical order of the scheme with the uniform norm and the Hausdorff distance

To compute a reference solution, the number of time-steps is chosen as $N = 10^6$, i.e., for a time step $h = 10^{-5}$. The error with the norm of the uniform convergence $\|.\|_\infty$ is displayed in log scale on the Fig. 6. We can see that there is no way to measure the rate of convergence with the $\|.\|_\infty$ norm. The result of the distance in the sense of filled-in graph is also displayed in log scale on the Fig. 6. On this example, the order of accuracy of the EMTS scheme is close to 1, as expected.

## 6 Applications: electrical circuits with ideal diodes

The well-posedness of dissipative circuits with ideal diodes ($r = 1$) has been investigated in [10,17] and their time-discretization with time-stepping Euler implicit schemes is studied in [16]. In this section we briefly illustrate how the application of feedback signals in simple electrical circuits may lead to higher relative degree complementarity systems. In other words we show how the material of the foregoing sections may help in understanding the closed-loop dynamics of some complementarity systems.

**A single input/single output case** Let us consider the following dynamics that corresponds to the circuit depicted in Fig. 7:

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = -\frac{R}{L}x_2(t) + \frac{u(t)}{L} - \frac{1}{LC}x_1(t) - \frac{1}{L}\lambda(t) \\ \\ 0 \leqslant \lambda(t) \perp w(t) = -x_2(t) \geqslant 0 \end{cases} \tag{164}$$

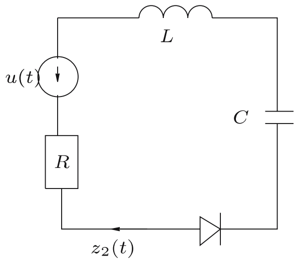


**Fig. 7** A simple electrical circuit with ideal diodes

with inductance $L > 0$, resistance $R > 0$, capacitance $C > 0$ and where $x_2$ is the current across the circuit, $-\lambda$ is the voltage of the diode and $u$ is a voltage control. If $u = 0$ then one sees that the transfer function of the operator $\lambda \mapsto w$ is given by $\frac{Cs}{LCs^2+RCs+1}$ and is positive real. Let us apply the following feedback

$$\begin{cases} u(t) = \lambda(t) + Lx_3(t) \\ \dot{x}_3(t) = x_4(t) \\ \dot{x}_4(t) = \lambda(t). \end{cases} \tag{165}$$

Inserting (165) into (164) one obtains

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = -\frac{R}{L}x_2(t) - \frac{1}{LC}x_1(t) + x_3(t) \\ \dot{x}_3(t) = x_4(t) \\ \dot{x}_4(t) = \lambda(t) \\ \\ 0 \leqslant w(t) = -x_2(t) \end{cases} \iff \begin{cases} \dot{z}_1(t) = z_2(t) \\ \dot{z}_2(t) = z_3(t) \\ \dot{z}_3(t) = \frac{R}{L}z_3(t) + \frac{1}{LC}z_2(t) + \lambda(t) \\ \dot{\xi}(t) = z_1(t) \\ \\ 0 \leqslant w(t) = z_1(t). \end{cases} \tag{166}$$

We embed the ZD dynamics in (166) in the higher order sweeping process formalism. The relative degree is $r = 1$ in (164), but applying a dynamic feedback changes it since $r = 3$ in (166), and $CA^2B = 1$. It is immediate that $r - 1$ is equal to the number of integrations in the control input in (165). It is noteworthy that one can apply Proposition 8. In other words, the feedback law in (165) augments the relative degree, but does not destroy the dissipativity of the closed loop system, since the transfer function of the operator $\lambda \mapsto z_3$ in (166) is positive real. This puts electrical circuits in a perspective that perfectly fits with the higher order sweeping process. Here $CA^2B = 1 > 0$ and for each $(\bar{z}_0, \xi_0) \in \mathbb{R}^n$, the system in (166) has a unique regular solution.

*Remark 28* i) We do not wish to discuss here the physical applicability of the feedback law in (165) (in practice measuring the voltage of the diode may introduce further dynamics), nor of an ideal diode model. At this stage it is however fundamental to keep in mind that studying such models allows the designer to point out and understand some phenomena which may be only limits (in the mathematical sense) of the real phenomena, but which would have been hidden by any other sort of modeling approach like penalization. It is also worth recalling that our model has some intrinsic flexibility, see e.g. Remark 15, and may therefore be adapted to better fit with the physical observations.

ii) Optimal control under state constraints also yields higher order complementarity systems [2,25,73]. Consequently it may benefit from the work in this paper.

## 7 Conclusion

In this paper we present an extension of Moreau's sweeping process, a widely studied differential inclusion in the field of unilateral Mechanics. This provides a new formalism for higher relative degree complementarity systems. It allows us to (1) obtain a clear understanding of the dynamical mechanism which permits the integration of such systems where higher degree distributions naturally appear, (2) to derive a numerical time-stepping scheme for Initial Value Problems. This paper focuses on the formalism and time-discretization aspects. The dynamical framework that is presented possesses several interesting features:

- The formalism and the functional framework are expected to extend to nonlinear, time varying vector fields, time varying and state dependent sets $\Phi(t,.)$, and the state re-initialization law can be modified (consequently it encompasses Lagrangian systems),
- the sweeping process differential inclusion is a suitable formalism to design time-stepping numerical schemes, which paves the way towards convergence analysis thanks to its compact formulation as a discretized differential inclusion,
- further extensions and their analysis may benefit a lot from the numerous studies on the first and second order sweeping processes,
- qualitative, dissipativity, well-posedness results show that the framework is sound,
- important potential application fields like electrical circuits, optimal control with state constraints (which is itself quite related to the dual problem of the so-called Continuous-Time Linear Programming problem [1,56] which inherently involves distributional solutions), may benefit from the approach.

This paper brings some elements of answer to a questioning in [55], about the interpretation of the existence theory for differential variational inequalities with index $\geqslant 3$ (i.e. $r \geqslant 2$ in the language of this paper).

**Acknowledgements** This work was partially supported by the European project SICONOS IST2001-37172. We thank Professor Anstreicher for pointing out the references [1] and [56]. Thanks are also due to the Associate Editor and the referees for helpful suggestions.

## Appendix A: Lemma 0.4.3 in [42]

**Lemma 5** *In a Hilbert space $H$, we consider a closed convex set $C$ which contains a closed ball $\bar{B}(a,R)$, $R > 0$. Let $x \in H$ be given. Then*

$$||x - \text{prox}\,[C;x]|| \leqslant \frac{1}{2R}(||x-a||^2 - ||\text{prox}\,[C;x]-a||^2). \tag{167}$$

## Appendix B: Kalman–Yakucovich–Popov lemma

Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{m\times n}$. One says that the representation $(A,B,C)$ is observable and controllable provided that $(A,B)$ is controllable and that $(A,C)$ is observable, i.e. the matrices $(B\ AB\ A^2B\ \ldots\ A^{n-1}B)$ and $(C\ CA\ CA^2\ \ldots\ CA^{n-1})^{\mathrm{T}}$ have full rank.

Let us now consider the real, rational matrix-valued transfer function $H : \mathbb{C} \to \mathbb{C}^{m\times m}$ given by

$$H(s) = C(sI_n - A)^{-1}B. \tag{168}$$

**Definition 4** *One says that H is positive real if*

- *$H$ is analytic in $\mathbb{C}^+ := \{s \in \mathbb{C} : \mathbf{Re}[s] > 0\}$,*
- *$H(s) + H^{\mathrm{T}}(\bar{s})$ is positive semidefinite for all $s \in \mathbb{C}^+$,*

*where $\bar{s}$ is the conjugate of s.*

**Definition 5** *One says that H is strictly positive real if*

- *$H$ is analytic in $\mathbb{C}^+ := \{s \in \mathbb{C} : \mathbf{Re}[s] > 0\}$,*
- *$H(s) + H^{\mathrm{T}}(\bar{s})$ is positive definite for all $s \in \overline{\mathbb{C}^+} = \{s \in \mathbb{C} : \mathbf{Re}[s] \geqslant 0\}$.*

One says that the triple $(A,B,C)$ is positive real (resp. strictly positive real) provided that the transfer function $H$ defined in (168) is positive real (resp. strictly positive real)

The following results are called Kalman–Yakubovich–Popov lemmas [13].

**Lemma 6** *Let $(A,B,C)$ be an observable and controllable realization and let H be defined in (168). The transfer function matrix H is positive real if and only if there exist a symmetric and positive definite matrix $P \in \mathbb{R}^{n\times n}$ and a matrix $L \in \mathbb{R}^{n\times m}$, such that*

$$\begin{aligned} PA + A^{\mathrm{T}}P &= -LL^{\mathrm{T}} \\ PB &= C^{\mathrm{T}}. \end{aligned} \tag{169}$$

**Lemma 7** *Let $(A,B,C)$ be an observable and controllable realization and suppose that A is Hurwitz. Let H be defined in (168). The transfer function matrix H is strictly positive real if and only if there exist a symmetric and positive definite matrix $P \in \mathbb{R}^{n\times n}$, a matrix $L \in \mathbb{R}^{n\times m}$ and $\varepsilon > 0$ such that*

$$\begin{aligned} PA + A^{\mathrm{T}}P &= -LL^{\mathrm{T}} - \varepsilon P \\ PB &= C^{\mathrm{T}}. \end{aligned} \tag{170}$$

**Algorithm 1** Sketch of the extended Moreau's time stepping (EMTS) scheme

```
Require: Classical form of the system : $A, B, C, x_0$
Require: Zero-Dynamic form : $r, W, A_\xi, B_\xi$
Require: Numerical parameters $h, T$
Ensure: $(\{x_n\}, \{z_n\}, \{\mu_n\}) = Approx(A, B, C, D, x_0, h, T)$
  // Computation of the operator associated with the ZD form
  $\bar{A}, \bar{G}, P$
  // Time discretization $N := [\frac{T}{h}]$
  // Computation of the time invariant numerical operators:
  $\bar{M} := (I - h\bar{A})^{-1}$
  $M_{lcptmp} := P\bar{M}\bar{G}$
  $b_{tmp} := P\bar{M}$
  $z_0 := Wx_0$
  // Loop on time.
  for $k = 0$ to N do
    //Computation of the rank $r^\star$
    $r^\star = 1, i = 1;$
    while $z_k(i) \leqslant 0$ do
      $r^\star = r^\star + 1$
      $i = i + 1;$
    end while
    // Computation of $R$
    // Solve the one-step LCP problem
    $M_{lcp} := RM_{lcptmp}R^{\mathrm{T}}$
    $b := b_{tmp}z_k$
    $(\mu_{k+1}, z^\star_{k+1}) :=$ SolveLCP$(\mathsf{M_{lcp}}, \mathsf{b})$
    // State update
    $z_{k+1} := \bar{M}\left[z_k + \bar{G}\mu_{k+1}\right]$
    $x_{k+1} := W^{-1}z_{k+1}$
  end for
```

## References

1. Anstreicher, K.M.: Generation of feasible descent directions in continuous time linear programming. Technical Report SOL 83-18, Systems Optimization Laboratory, Dept. of Operations Research, Stanford University (1983)
2. Arutyunov, A.V., Aseev, S.M.: State constraints in optimal control. The degeneracy phenomenon. Systems Control Lett. **26**, 267–273 (1995)
3. Aubin, J.P., Cellina, A.: Differential Inclusions. Springer, Berlin Heidelberg New York (1994)
4. Ballard, P.: The dynamics of discrete mechanical systems with perfect unilateral constraints. Arch. Rational Mech. Anal. **154**, 199–274 (2000)
5. Benabdellah, H., Castaing, C., Salvadori, A., Syam, A.: Nonconvex sweeping process. J. Appl. Anal. **2**(2), 217–240 (1996)
6. Berman, A., Plemmons: Nonnegative Matrices in The Mathematical Sciences. Computer Science and Applied Mathematics. Academic, New York (1979)
7. Brezis, H.: Opérateurs maximaux monotones et semi-groupes de contraction dans les espaces de Hilbert. North Holland, Amsterdam (1973)
8. Brogliato, B.: Nonsmooth Mechanics, 2nd edn. Springer, London (1999)
9. Brogliato, B.: Some perspectives on the analysis and control of complementarity systems. IEEE Trans. Autom. Control **48**(6), 918–935 (2003)
10. Brogliato, B.: The absolute stability problem and the Lagrange–Dirichlet theorem with monotone multivalued mappings. Systems Control Lett. **51**(5), 343–353 (2004)
11. Brogliato, B.: Some results on the controllability of planar variational inequalities. Systems Control Lett. **54**: 65–71 (2005)

12. Brogliato, B., Daniilidis, A., Lemaréchal, C., Acary, V.: On the equivalence between complementarity systems, projected systems and differential inclusions. Systems Control Lett. **55**(1), 45–51 (2006)
13. Brogliato, B., Lozano, R., Maschke, B., Egeland, O.: Dissipative Systems Analysis and Control. Theory and Applications. 2nd edn.. Springer, London (2006)
14. Brogliato, B., Niculescu, S., Orhant, P.: On the control of finite-dimensional mechanical systems with unilateral constraints. IEEE Trans. Autom. Control **42**(2), 200–215 (1997)
15. Brogliato, B., ten Dam, A.A., Paoli, L., Genot, F., Abadie, M.: Numerical simulation of finite dimensional multibody nonsmooth mechanical systems. ASME Appl. Mech. Rev. **55**(2), 107–150 (2002)
16. Camlibel, K., Heemels, W.P.M.H., Schumacher, J.M.: Consistency of a time-stepping method for a class of piecewise-linear networks. IEEE Trans. Circuits Systems I **49**, 349–357 (2002)
17. Camlibel, K., Heemels, W.P.M.H., Schumacher, J.M.: On linear passive complementarity systems. Eur. J. Control **8**(3), 220–237 (2002)
18. Castaing, C., Duc Ha, T.X., Valadier, M.: Evolution equations governed by the sweeping process. Set-Valued Anal. **1**, 109–139 (1993)
19. Castaing, C., Monteiro-Marques, M.D.P.: Evolution problems associated with nonconvex closed moving sets with bounded variation. Portugaliae Math. 1, 73–87 (1996)
20. Cobb, D.: On the solution of linear differential equations with singular coefficients. J. Differ. Equ. **46**, 310–323 (1982)
21. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. Studies in Advanced Mathematics. CRC Press, Boca Raton (1992)
22. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, vol I & II of Springer Series in Operations Research. Springer, Berlin Heidelberg New York (2003)
23. Goeleven, D., Brogliato, B.: Stability and instability matrices for linear evolution variational inequalities. IEEE Trans. Autom. Control **49**(4), 521–534 (2004)
24. Goeleven, D., Motreanu, D., Dumont, Y., Rochdi, M.: Variational and Hemivariational Inequalities: Theory, Methods and Applications, vol. I: Unilateral Analysis and Unilateral Mechanics. Nonconvex Optimization and its Applications. Kluwer (2003)
25. Hartl, R.F., Sethi, S.P., Vickson, R.G.: A survey of the maximum principles for optimal control problems with state constraints. SIAM Rev. **37**, 181–218 (1995)
26. Heemels, W.P.M.H.: Linear Complementarity Systems. A Study in Hybrid Dynamics. PhD thesis, Technical University of Eindhoven (1999). ISBN 90-386-1690-2
27. Heemels, W.P.M.H., Brogliato, B.: The complementarity class of hybrid dynamical systems. Eur. J. Control **9**, 311–349 (2003)
28. Heemels, W.P.M.H., Schumacher, J.M., Weiland, S.: Linear complementarity problems. SIAM J. Appl. Math. **60**(4), 1234–1269 (2000)
29. Imura, J.I.: Well-posedness analysis of switch-driven piecewise affine systems. IEEE Trans. Autom. Control **48**(11), 1926–1935 (2003)
30. Jean, M.: The non smooth contact dynamics method. Comput. Methods Appl. Mech. Eng. **177**, 235–257 (1999). Special issue on computational modeling of contact and friction, Martins, J.A.C., Klarbring, A. (eds.)
31. Juloski, A., Heemels, W.P.M.H., Brogliato, B.: Observer design for Lur'e systems with multivalued mappings. In: IFAC World Congress Prague, 4–8 July 2005
32. Krejci, P., Vladimorov, A.: Polyhedral sweeping processes with oblique reflection in the space of regulated functions. Set-Valued Anal **11**, 91–110 (2003)
33. Kunze, M., Monteiro Marques, M.D.P.: Existence of solutions for degenerate sweeping processes. J. Convex Anal. **4**, 165–176 (1997)
34. Kunze, M., Monteiro Marques, M.D.P.: On parabolic quasi-variational inequalities and state-dependent sweeping processes. Topol Methods Nonlinear Anal. **12**, 179–191 (1998)
35. Kunze, M., Monteiro Marques, M.D.P.: An introduction to Moreau's sweeping process. In: Brogliato, B. (ed.) Impact in Mechanical systems: Analysis and Modelling, pp. 1–60 Lecture Notes in Physics vol. 551. Springer, Berlin Heidelberg New York (2000)
36. Kunze, M., Monteiro MarquFs, M.D.P.: Yosida-Moreau regularisation of sweeping processes with unbounded variation. J. Differ. Equ. **130**, 292–306 (1996)
37. Liu, W.Q., Lan, W.Y., Teo, K.L.: On initial instantaneous jumps of singular systems. IEEE Trans. Autom. Control **40**(9), 1650–1655 (1995)

38. Lootsma, Y.J., van der Schaft, A.J., Camlibel, K.: Uniqueness of solutions of relay systems. Automatica **35**(3), 467–478 (1999)
39. Lotstedt, P.: Mechanical systems of rigid bodies subject to unilateral constraints. SIAM J. Appl. Math. **42**(2), 281–296 (1982)
40. Mabrouk, M.: A unified variational for the dynamics of perfect unilateral constraints. Eur. J. Mech. A/Solids **17**, 819–842 (1998)
41. Machanda, P., Siddiqi, A.H.: Rate dependent evolution quasivariational inequalities and state-dependent sweeping processes. Adv. in Nonlinear Var. Inequal. **5**(1):1–16 (2002)
42. Monteiro Marques, M.P.D.: Differential Inclusions in NonSmooth Mechanical Problems: Shocks and Dry Friction. Birkhauser (1993)
43. Moreau, J.J.: Les liaisons unilatérales et le principe de Gauss. Comptes Rendus de l' des Sci **256**, 871–874 (1963)
44. Moreau, J.J.: Quadratic programming in mechanics: dynamics of one sided constraints. SIAM J. Control **4**(1), 153–158 (1966)
45. Moreau, J.J.: Rafle par un convexe variable (première partie), exposé no 15. Séminaire d'analyse convexe, University of Montpellier, 43 pp (1971)
46. Moreau, J.J.: Rafle par un convexe variable (deuxième partie) exposé no 3. Séminaire d'analyse convexe, University of Montpellier, 36 pp (1972)
47. Moreau, J.J.: Problme d'évolution associć a un convexe mobile d'un espace hilbertien. Comptes Rendus de l'AcadTmie des Sci, Séries A-B, **t.276**, 791–794 (1973)
48. Moreau, J.J.: Evolution problem associated with a moving convex set in a Hilbert space. J. Differ. Equ. **26**, 347–374 (1977)
49. Moreau, J.J.: Approximation en graphe d'une évolution discontinue. RAIRO Analyse numérique/ Numer. Anal. **12**, 75–84 (1978)
50. Moreau, J.J.: Liaisons unilatérales sans frottement et chocs inélastiques. Comptes Rendus de l'AcadTmie des Sci. **296** serie II, 1473–1476 (1983)
51. Moreau, J.J.: Unilateral contact and dry friction in finite freedom dynamics. In: Moreau, J.J., Panagiotopoulos, P.D. (eds.) Nonsmooth Mechanics and Applications, pp. 1–82. CISM, Courses and Lectures, vol. 302. Springer, Berlin Heidelberg New York (1988)
52. Moreau, J.J.: Some numerical methods in multibody dynamics: Application to granular materials. Eur. J. Mech. A/Solids **supp.(4)**, 93–114 (1994)
53. Moreau, J.J.: Numerical aspects of the sweeping process. Comput. Methods Appl. Mech. Eng. **177**, 329–349 (1999). Special issue on computational modeling of contact and friction, Martins, J.A.C., Klarbring, A. (eds.)
54. Moreau, J.J.: An introduction to unilateral dynamics. In: FrTmond, M., Maceri, F. (eds.) Novel Approaches in Civil Engineering, Series: Lecture Notes in Applied and Computational Mechanics. Springer, Berlin Heidelberg New York (2003)
55. Pang, J.S., Stewart, D.: Differential variational inequalities. Preprint (2004)
56. Perold, A.: Fundamentals of a continuous time simplex method. Technical Report SOL 78-26, Systems Optimization Laboratory, Dept. of Operations Research, Stanford University (1978)
57. Pogromsky, A.Y., Heemels, W.P.M.H., Nijmeijer, H.: On solution concepts and well-posedness of linear relay systems. Automatica **39**(12), 2139–2147 (2003)
58. Saberi, A., Sannuti, P.: Cheap and singular controls for linear quadratic regulators. IEEE Trans. Autom. Control **32**(3), 208–219 (1987)
59. Sain, M.K., Massey, J.L.: Invertibility of linear time-invariant dynamical systems. IEEE Trans. Autom. Control **14**(2), 141–149 (1969)
60. Sannuti, P.: Direct singular pertubation analysis of high gain and cheap control problem. Automatica **19**(1), 424–440 (1983)
61. Sannuti, P., Saberi, A.: Special coordinate basis for multivariable linear systems—finite and infinite zero structure, squaring down and decoupling. Int. J. Control **45**(5), 1655–1704 (1987)
62. Sannuti, P., Wason, H.: Multiple time-scale decomposition in cheap control problemes—singular control. IEEE Trans. Autom. Control **30**(7), 633–644 (1985)
63. Schatzman, M.: Sur une classe de problémes hyperboliques non linéaires. Comptes Rendus de l'Académie des Sci Série A **277**, 671–674 (1973)
64. Schatzman, M.: A class of nonlinear differential equations of second order in time. Nonlinear Anal. Theory Methods Appl. **2**(3), 355–373 (1978)
65. Schwartz, L.: Analyse III, Calcul Intégral. Hermann (1993)

66. Shilov, G.E., Gurevich, B.L.: Integral Measure and Derivative. A Unified Approach. Prentice-Hall, Englewood Cliffs, (1966). Hermann, Paris (1993)
67. Siddiqi, A.H., Machanda, P.: Variants of Moreau's sweeping process. Adv. Nonlinear Var. Inequal. **5**(1), 17–28 (2002)
68. Siddiqi, A.H., Machanda, P., Brokate, M.: On some recent developments concerning Moreau's sweeping process. In: Siddiqi, A.H., Kocvara, M. (eds.) International Conference on Emerging Areas in Industrial and applied mathematics. GNDU Amritsar, India. Kluwer Boston-London-Dordrecht (2001)
69. Stewart, D.: Convergence of a time-stepping scheme for rigid-body dynamics and resolution of PainlevT's problem. Archi. Rational Mech. Anal. **145**, 215–260 (1998)
70. Stewart, D.: Rigid body dynamics with friction and impact. SIAM Rev. **42**(1), 3–39 (2000)
71. ten Dam, A.A., Dwarshuis, E., Willems, J.C.: The contact problem for linear continuous-time dynamical systems: a geometric approach. IEEE Trans. Autom. Control **42**(4), 458–472 (1997)
72. Thibault, L.: Sweeping process with regular and nonregular sets. J. Differ. Equ. **193**(1), 1–26 (2003)
73. van der Schaft, A.J., Schumacher, J.M.: An Introduction to Hybrid Dynamical Systems. Springer, London (2000)
74. van Der Schaft, A.J., Schumacher, J.M.: The complementary-slackness class of hybrid systems. Math Control Signals Systems **9**(3), 266–301 (1996)
75. van der Schaft, A.J., Schumacher, J.M.: Complementarity modeling of hybrid systems. IEEE Trans. Autom. Control **43**(4), 483–490 (1998)