Preservation of the dissipativity properties of a class of nonsmooth dynamical systems with the (θ,γ)-algorithm

Scott Greenhalgh, Vincent Acary, Bernard Brogliato

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Preservation of the dissipativity properties of a class of nonsmooth dynamical systems with the (θ, γ)-algorithm

Scott Greenhalgh — Vincent Acary — Bernard Brogliato

N° 7632

May 30, 2011

Thème NUM

ISSN 0249-6399 ISRN INRIA/RR--7632--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
GRENOBLE - RHÔNE-ALPES

Preservation of the dissipativity properties of a class of nonsmooth dynamical systems with the (θ, γ)-algorithm

Scott Greenhalgh*†, Vincent Acary* , Bernard Brogliato*

Thème NUM — Systèmes numériques
Équipe-Projet Bipop

Rapport de recherche n° 7632 — May 30, 2011 — 48 pages

Abstract: In this work we study the following problem: given a numerical method (an extended $\theta-$method named the (θ, γ)-method), find the class of dissipative linear complementarity systems such that their discrete-time counterpart is still dissipative, with the same storage (energy) function, supply rate (reciprocal variables), and dissipation function. Systems with continuous solutions, and with state jumps are studied. The notion of numerical dissipation is given a rigorous meaning.

Key-words: dissipative systems, (θ, γ)-method, energy function, state jumps, complementarity systems, complementarity problems; numerical dissipation.

* INRIA, Bipop team-project, ZIRST Montbonnot, 655 avenue de l'Europe, 38334 Saint Ismier cedex, France. firstname.lastname@inrialpes.fr

† Work performed while at INRIA Grenoble Rhône-Alpes, Bipop project-team, ZIRST Montbonnot, 655 avenue de l'Europe, 38334 Saint Ismier cedex, France; supported by French National Research Agency (ANR) through COSINUS program (project SALADYN ANR-08-COSI-014). On leave from the University of Guelph, Department of Mathematics and Statistics, Canada. greenhas@uoguelph.ca.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Préservation des propriétés de dissipativité pour une classe de systèmes dynamiques non réguliers grâce au (θ, γ) schéma

Résumé : Dans ce travail, on se préoccupe du problème suivant: étant donnée une méthode numérique (une extension du $\theta-$schéma dénommé (θ, γ) schéma), trouver la classe de systèmes de complémentarité linéaire pour laquelle la discrétisation en temps est aussi dissipative, avec la même fonction d'énergie, le même taux d'apport d'énergie (en variables réciproques) et la même fonction de dissipation. Les systèmes avec des solutions continues et avec des sauts dans l'état sont étudiés. La notion de dissipation numérique est donnée avec une interprétation rigoureuse.

Mots-clés : systèmes dissipatifs, (θ, γ) schéma, fonction d'énergie, sauts dans létat, systèmes de complémentarié, problémes de complémentarité, dissipation numérique

1 Introduction

The works that deal with numerical schemes that preserve energy, momentum, or other types of linear or nonlinear constraints are numerous in the numerical mechanics literature (symplectic methods, conservation of first integrals, RK and multistep methods), see *e.g.* [1, 2, 3, 4, 5, 48]. Most of them deal with (conservative) Hamiltonian systems. In parallel to these results and due to the importance of the notion of dissipativity in Control applications [27] (dissipativity being the positive real property of transfer functions in the case of linear, time-invariant systems), the preservation of dissipativity properties (or of the positive realness) after time-discretization has been studied for a long time, see *e.g.* [16, 20, 24, 25, 30, 32, 33, 47]. In the above works the question usually answered is: given a positive real system, perform a time-discretization (Euler, or zero order hold) and examine whether the obtained discrete-time system is still positive real, possibly with different storage function and dissipation function. For instance four different types of discretizations are studied in [20]. Whether or not the continuous-time and the discrete-time possess the same storage function or the same dissipation function, is tackled in [16, 24, 25, 31]. Recently the interest has focused on dissipativity of nonsmooth dynamical systems like dynamical complementarity systems [9, 11, 12, 13, 14, 18, 19, 22, 28], hybrid systems [7], and multivalued Lur'e systems [8, 10].

In this paper we deal with linear complementarity dynamical systems, possibly with state jumps. We deal with preservation of passivity (in Willems' sense [27]) after discretization by an extended $\theta-$method called the (θ, γ)-algorithm. In view of the state of the art on time discretization of such nonsmooth systems, higher order methods are not yet available and only first order methods (implicit or explicit Euler, Paoli-Schatzman's scheme [40]) have been shown to converge. Extensions towards higher-order methods is an open issue, not tackled in this paper (see [49] for some preliminary work in the field of nonsmooth mechanical systems). This means that RK, multistep methods are outside the scope of this study. The problem that is tackled in this paper is as follows:

Given a discretization method, find the class of linear complementarity dissipative systems such that their discretized counterpart is still dissipative with the same storage function set, supply rate and dissipation function.

In addition the method, when applied to complementarity dynamical system, should guarantee that the so-called one-step-nonsmooth-problem to be solved at each time step, possesses a unique solution, and, in case the solution jumps, that the energetic properties of the jump rule are preserved. Usually, all this yields quite stringent conditions and narrow classes of continuous-time systems, and may be seen as the counterpart of the problem tackled in [16, 24, 25] which is: find a discretization method such that any dissipative system is transformed into a dissipative discrete-time system. Finally we do not want to stick to the conservative (or lossless) case as in most of the above cited works on mechanical systems, since it is desirable to deal with systems that possess a non-zero dissipation function and to seek conditions under which the dissipation function is also preserved.

The case of linear complementarity systems without state jumps is dealt with first, and then we focus on state jumps. In this paper we are not interested in convergence results as the time-step goes to zero, but on the algorithm properties when $h > 0$. It is however pointed out at some places that preserving dissipativity (which is characterized by three ingredients: the energy function, the dissipation function and the supply rate) may not yield "good" numerical results and that using schemes that do not preserve one of the three ingredients may be preferred. The paper is organized as follows: in section 2 the continuous-time and the discrete-time systems are presented, the definitions of dissipativity are recalled, and a definition of numerical dissipation is given. Section 3 is dedicated to the study of the conditions such that dissipativity is preserved after the discretization. In section 4 we examine whether the numerical method consistently approximates state jumps. Conclusions are given in section 5 and some technical details are provided in the Appendix. Many academic and physical examples (electrical circuits with multivalued nonsmooth components) are used throughout the paper to illustrate the theoretical developments. All the numerical results have been obtained with the SICONOS platform of the INRIA, see [39, 40, 41].

Notation The right and left limits of a function f at t are denoted as $f(^+)$ and $f(t^-)$ respectively. The normal cone to a convex non-empty set $K \subseteq \mathbb{R}^n$ at $x \in K$ is $N_K(x) = \{v \in \mathbb{R}^n \mid \langle v, z-x\rangle \leqslant 0 \text{ for all } z \in K\}$. A matrix is a P-matrix if all its principal minors are positive. A linear complementarity problem (LCP) with unknown $\lambda \in \mathbb{R}^m$ is a problem of the form $\lambda \geqslant 0$, $M\lambda+q \geqslant 0$, $\lambda^T(M\lambda+q)=0$, written compactly as $0 \leqslant \lambda \perp M\lambda+q \geqslant 0$. This is denoted LCP$(q,M)$, and the set of solutions is SOL(q,M). Let $K \subseteq \mathbb{R}^m$ be a convex non-empty closed cone, its dual cone is the set $K^* = \{v \in \mathbb{R}^m \mid v^T z \geqslant 0 \text{ for all } z \in K\}$. A linear cone CP (LCCP) is a problem of the form $K \ni \lambda \perp M\lambda+q \in K^*$. Ker$(A)$ is the kernel of the matrix A. A positive semi definite (PSD) matrix M, possibly nonsymmetric, is such that for all $x \in \mathbb{R}^n$ one has $x^T Mx \geqslant 0$. It is positive definite if $x^T Mx > 0$ for all $x \neq 0$. I is the identity matrix with appropriate dimension.

2 The dynamical system and its discretization

In this section and in section in section 3 we deal with the case without state jumps.

2.1 Continuous-time systems: the dynamics and dissipativity LMIs

We consider the following linear complementarity systems (LCS):

$$\begin{cases} \dot{x}(t) = Ax(t) + B\lambda(t) + Eu(t) \\ w(t) = Cx(t) + D\lambda(t) + Fv(t) \\ 0 \leqslant \lambda(t) \perp w(t) \geqslant 0 \\ x(0^-) = x_{in} \end{cases} \tag{1}$$

with $x(t) \in \mathbb{R}^n$, $\lambda(t) \in \mathbb{R}^m$, $w(t) \in \mathbb{R}^m$. The well-posedness (existence and uniqueness of solutions) of such systems has been studied. Depending on the data solutions may be continuous, of class C^1, discontinuous functions, measures, or distributions, see *e.g.* [6, 8, 9, 13, 15, 28, 46]. A general assumption althrough the paper is that $v(\cdot)$ and $u(\cdot)$ are bounded functions of time. In the first part of the paper we suppose that the solutions are absolutely continuous so that the first equality in (1) is satisfied almost everywhere, and we also suppose that $E=0$ and $F=0$.

Let us assume that the quadruple (A,B,C,D) is dissipative with supply rate $S(w,\lambda) = \langle \lambda, w\rangle = \lambda^T w$, *i.e.* it satisfies a linear matrix inequality (LMI) of the form: there exists $P \in \mathbb{R}^{n\times n}$ such that:

$$\begin{pmatrix} A^TP + PA & PB - C^T \\ B^TP - C & -D - D^T \end{pmatrix} \leqslant 0$$
$$P = P^T \geqslant 0, \tag{2}$$

or equivalently, there exists $L \in \mathbb{R}^{n\times m}$, $W \in \mathbb{R}^{m\times m}$ and $P \in \mathbb{R}^{n\times n}$ such that:

$$\begin{cases} A^TP + PA = -LL^T & (3) \\ B^TP - C = -W^TL^T & (4) \\ -D - D^T = -W^TW & (5) \\ P = P^T \geqslant 0. & (6) \end{cases}$$

There are three ingredients in the definition of a dissipative system: a storage function (given by $V(x) = \frac{1}{2}x^TPx$), a supply rate (*i.e.* $S(\lambda,w)$), and a dissipation function that is quadratic in (x,λ)

(with the matrix $\mathcal{Q} \triangleq \begin{pmatrix} LL^T & W^TL^T \\ LW & W^TW \end{pmatrix}$). When the supply rate is $S = \langle w, \lambda \rangle$, the system is said to be *passive*. Systems with $L = 0$ are *state-lossless*, with $W = 0$ are *input-lossless*, and with $\mathcal{Q} = 0$ are *lossless*. When $Q \geqslant 0$ the system is said to be *strictly passive*.

Remark 1 *A particular feature of LCS is that due to the complementarity constraints $S(\lambda(t), w(t)) = 0$ for all $t \geqslant 0$.*

These three ingredients transform into three equalities and characterize the dissipativity, as in (3-6). The LMI in (2) means that the system is dissipative with the supply rate defined from the "reciprocal" variables $w(t)$ and $\lambda(t)$: $\langle w(t), \lambda(t) \rangle$. The solution set of the LMI (2) is denoted $\mathcal{P}$. When the LMI (2) holds then the dissipation equality in (7) holds also, and *vice versa*:

$$V(x(T)) - V(x(0)) = -\frac{1}{2}\int_0^T (x^T(t), \lambda^T(t))\mathcal{Q}\begin{pmatrix} x(t) \\ \lambda(t) \end{pmatrix} dt, \quad \forall\ T \geqslant 0 \tag{7}$$

The *infinitesimal* dissipation inequality writes as:

$$\dot{V}(x(t)) = -\frac{1}{2}(x^T(t), \lambda^T(t))\mathcal{Q}\begin{pmatrix} x \\ \lambda \end{pmatrix} \tag{8}$$

which is equivalent to (7) as long as $x(\cdot)$ is differentiable or absolutely continuous (hence with a derivative almost everywhere). We define the (continuous) cumulative dissipation function as:

$$\int_0^t \frac{1}{2}(x^T(t), \lambda^T(t))\mathcal{Q}\begin{pmatrix} x(t) \\ \lambda(t) \end{pmatrix} dt \tag{9}$$

As the next example shows, allowing for $P \geqslant 0$ in (2) is important because if the pair (A, C) is not observable the LMI (2) may possess positive semi definite solutions only.

Example 1 *Consider (A, B, C, D) defined as:*

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 1 \\ 0 & \cdots & \cdots & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad C = (0 \cdots 0\ \ 1), \quad D = 0. \tag{10}$$

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$, $C \in \mathbb{R}^{1\times n}$. Suppose we want to characterize the lossless property, i.e. *$\leqslant 0$ is replaced by $= 0$ in (2). Enforcing $A^TP + PA = 0$ and $PB = C^T$, implies $p_{i,j} = \begin{cases} 1 & \text{for } (i,j) = (n,n) \\ 0 & o.w. \end{cases}$, and hence $P = \begin{pmatrix} 0_{n-1,n-1} & 0_{n-1} \\ 0_{n-1}^T & 1 \end{pmatrix}$ where the notation $0_{n,n}$ and 0_n is used to represent zero matrices and zero vectors of dimensions $n \times n$ and n respectively. Consequently the LMI in (2) for the lossless case does not possess any positive definite solution, but only positive semi definite solutions. The system is not observable as $CA = 0$. Observability implies $P > 0$ in (2), because the kernel of P satisfying (2) is a subset of the unobservability space of (C, A) [28, Lemma 2].*

When $D = 0$ in (2) then by [27, Proposition A.63] one has $PB = C^T$. Physically speaking, the terms $A^TP + PA$ and $-D - D^T$ are responsible for energy decay (dissipation), whereas $PB - C^T$ corresponds to an "input-output" constraint that says that the slack variables w and λ are reciprocal variables (like velocity and force in mechanics, or voltage and current in electricity).

2.2 The time-discretization

In this section we present the discretization of (1), its LMI characterization, and a result in [24, 25] is recalled.

2.2.1 The $(\theta,\gamma)-$method and the discrete-time LMIs

Let us propose the following $(\theta,\gamma)-$method for (1):

$$\left\{\begin{array}{l} \frac{x_{k+1}-x_k}{h} = Ax_{k+\theta} + B\lambda_{k+\gamma} \\ \\ 0 \leqslant \lambda_{k+\gamma} \perp w_{k+\gamma} = Cx_{k+\gamma} + D\lambda_{k+\gamma} \geqslant 0 \\ \\ x_0 = x_{in}, \end{array}\right. \tag{11}$$

where θ and $\gamma \in [0,1]$, and the notation $k+\theta$ implies $x_{k+\theta} = \theta x_{k+1} + (1-\theta)x_k$, $\lambda_{k+\gamma} = \gamma\lambda_{k+1} + (1-\gamma)\lambda_k$. By specifying (θ,γ) we completely characterize the form of the discretization: $(\theta,\gamma) = (1,1)$ is a fully implicit scheme, $(\theta,\gamma) = (0,1)$ is a semi-implicit scheme, and $(\theta,\gamma) = (0,0)$ is a fully explicit scheme. Assuming that the inverse $(I_n - h\theta A)^{-1}$ is well defined (a sufficent condition is $h < \frac{1}{\theta||A||}$ where $||\cdot||$ is a norm for which $||I_n|| = 1$ [23, Theorem 1, Chapter 11], but in many cases $I_n - h\theta A$ may be full rank for $h > 0$ not necessarily small), we define:

$$\left\{\begin{array}{l} \tilde{A} = (I_n - h\theta A)^{-1}(I_n + h(1-\theta)A) \\ \\ \tilde{B} = h(I_n - h\theta A)^{-1}B \\ \\ \tilde{C} = \gamma C\tilde{A} + (1-\gamma)C \\ \\ \tilde{D} = \gamma C\tilde{B} + D. \end{array}\right. \tag{12}$$

The $(\theta,\gamma)-$ discretization of the LCS is compactly written as:

$$\left\{\begin{array}{l} x_{k+1} = \tilde{A}x_k + \tilde{B}\lambda_{k+\gamma} \\ \\ w_{k+\gamma} = \tilde{C}x_k + \tilde{D}\lambda_{k+\gamma} \\ \\ 0 \leqslant \lambda_{k+\gamma} \perp w_{k+\gamma} \geqslant 0 \\ \\ x_0 = x_{in} \end{array}\right. \tag{13}$$

One can infer directly from (12) that it is necessary that $\gamma > 0$ when $D = 0$. Indeed a discrete-time system that is passive has a non-zero feedthrough matrix [27]. We further note that $\tilde{A} = \tilde{A}(\theta)$ and $\tilde{B} = \tilde{B}(\theta)$, $\tilde{C} = \tilde{C}(\theta,\gamma)$, $\tilde{D} = \tilde{D}(\theta,\gamma)$. From the discrete-time version of the passivity LMI (see *e.g.* [27, §3.12]), we deduce that the system in (13) is passive if and only if the following discrete-time LMI is satisfied:

$$\left(\begin{array}{cc} \tilde{A}^T R\tilde{A} - R & \tilde{A}^T R\tilde{B} - \tilde{C}^T \\ \\ \tilde{B}^T R\tilde{A} - \tilde{C} & -\tilde{D} - \tilde{D}^T + \tilde{B}^T R\tilde{B} \end{array}\right) \leqslant 0 \tag{14}$$

$$R = R^T \geqslant 0,$$

or equivalently, that there exists $R \in I\!R^{n\times n}$, $\tilde{L} \in \mathbb{R}^{n\times m}$ and $\tilde{W} \in \mathbb{R}^{m\times m}$ such that:

$$\left\{\begin{array}{ll} \tilde{A}^T R\tilde{A} - R = -\tilde{L}\tilde{L}^T & (15) \\ \\ \tilde{B}^T R\tilde{A} - \tilde{C} = -\tilde{W}^T\tilde{L}^T & (16) \\ \\ \tilde{B}^T R\tilde{B} - \tilde{D} - \tilde{D}^T = -\tilde{W}^T\tilde{W} & (17) \\ \\ R = R^T \geqslant 0 & (18) \end{array}\right.$$

Remark 2 *Like the continuous-time case, we notice that $\lambda_{k+\gamma}^T w_{k+\gamma} = 0$ for all $k \geqslant 0$, which is guaranteed by solving a linear complementarity problem at each step, see section 2.2.4.*

One defines the positive semi definite storage function $V(x_k) = \frac{1}{2}x_k^T R x_k$ which satisfies for all $k \geqslant 0$:

$$V(x_{k+1}) - V(x_k) = -\frac{1}{2}(x_k^T \tilde{L}\tilde{L}^T x_k + x_k^T \tilde{L}\tilde{W}\lambda_{k+\gamma} + \lambda_{k+\gamma}^T \tilde{W}^T \tilde{L}^T x_k^T + \lambda_{k+\gamma}^T \tilde{W}^T \tilde{W}\lambda_{k+\gamma}^T) \quad (19)$$

Then when the discrete-time LMI (14) holds, the discrete dissipation equality (with the supply rate $\langle w_{k+\gamma}, \lambda_{k+\gamma}\rangle$) holds and *vice versa*:

$$V(x_{k+1}) - V(x_k) = -\frac{1}{2}(x_k^T, \lambda_{k+\gamma}^T)\tilde{Q}\begin{pmatrix} x_k \\ \lambda_{k+\gamma} \end{pmatrix} \quad (20)$$

or equivalently:

$$V(x_{k+1}) - V(x_0) = -\frac{1}{2}\sum_{i=0}^{k}(x_i^T, \lambda_{i+\gamma}^T)\tilde{Q}\begin{pmatrix} x_i \\ \lambda_{i+\gamma} \end{pmatrix}, \quad (21)$$

with $\tilde{\mathcal{Q}} = \begin{pmatrix} \tilde{L}\tilde{L}^T & \tilde{W}^T\tilde{L}^T \\ \tilde{L}\tilde{W} & \tilde{W}^T\tilde{W} \end{pmatrix}$, and for all $k \geqslant 0$. We denote the set of solutions of (14) as $\mathcal{R}_{\theta,\gamma}$. We define the (discrete) cumulative dissipation function as:

$$\sum_i \frac{h}{2}(x_i^T, \lambda_{i+\gamma}^T)\tilde{\mathcal{Q}}\begin{pmatrix} x_i \\ \lambda_{i+\gamma} \end{pmatrix}. \quad (22)$$

Let $\tilde{R} = (I - h\theta A)^{-T} R (I - h\theta A)^{-1} \in I\!R^{n\times n}$. Suppose that $I_n - h\theta A$ has full rank n. Then $R \geqslant 0 \Leftrightarrow \tilde{R} \geqslant 0$. Using (12) we may equivalently rewrite (15-18) as follows:

$$\begin{cases} h(A^T\tilde{R} + \tilde{R}A) + h^2(1-2\theta)A^T\tilde{R}A = -\tilde{L}\tilde{L}^T & (23) \\ hB^T\tilde{R}(I + h(1-\theta)A) - \tilde{C} = -\tilde{W}^T\tilde{L}^T & (24) \\ h^2B^T\tilde{R}B - \tilde{D}^T - \tilde{D} = -\tilde{W}^T\tilde{W} & (25) \\ \tilde{R} = \tilde{R}^T \geqslant 0. & (26) \end{cases}$$

In the next Lemma are stated useful expressions for the sequel of the paper.

Lemma 1 *(i) The equality (15) is equivalent to:*

$$h(A^TR + RA) + h^2(1-2\theta)A^TRA = -(I - h\theta A)^T\tilde{L}\tilde{L}^T(I - h\theta A). \quad (27)$$

(ii) The equality (16) is equivalent to:

$$\begin{aligned} &hB^TR\big(h(1-\theta-\gamma)A - h^2\theta(\gamma-\theta)A^2\big) - \theta B^T\tilde{L}\tilde{L}^T(I-h\theta A)(I+h(1-\theta)A) \\ &= W^TL^T(I+h(\gamma-\theta)A)(I+h\theta A) - \tilde{W}^T\tilde{L}^T(I-h\theta A)(I+h\theta A). \end{aligned} \quad (28)$$

(iii) The equality (17) is equivalent to:

$$\begin{aligned} &h^2B^T(I-h\theta A)^{-T}((1-2\gamma)R - \theta\gamma LL^T)(I-h\theta A)^{-1}B \\ &= W^TW - \tilde{W}^T\tilde{W} + h\gamma W^TL^T(I-h\theta A)^{-1}B + h\gamma B^T(I-h\theta A)^{-T}LW. \end{aligned} \quad (29)$$

The proof is given in section B in the appendix. It appears straightforwardly from (25) that a necessary condition for this problem to possess a solution is that

$$h^2 B^T \tilde{R} B \leqslant \tilde{D}^T + \tilde{D}, \tag{30}$$

i.e. $h^2 B^T \tilde{R} B \leqslant \tilde{D}^T + \tilde{D}$ is in the set of PSD matrices. When $D = 0$ this implies that $\gamma > 0$, excluding fully explicit methods. Imposing that $\mathcal{P} = \mathcal{R}_{\theta,\gamma}$ the three equalities in (23) –(25) (equivalently in (15)–(17)) govern the storage function and the input-output constraint of the discrete-time system. The dissipation function preservation will be dealt with later. Notice that due to the fact that $\lambda_{k+\gamma}^T w_{k+\gamma} = 0$ from (19) and since the reciprocal variables belong to a cone, the supply rate may be scaled by any positive constant without changing the system's dissipativity properties.

Proposition 1 *Examining (23), (27) and (3), two natural ansatzes are:*

$$\begin{cases} R = \frac{1}{h} P, \ (\textit{resp. } R = P) \\ \tilde{L}\tilde{L}^T = (I - h\theta A)^{-T} L L^T (I - h\theta A)^{-1}, \ (\textit{resp. } \tilde{L}\tilde{L}^T = h(I - h\theta A)^{-T} L L^T (I - h\theta A)^{-1}) \\ \textit{Either } \theta = \frac{1}{2} \textit{ or } A^T R A = 0 \end{cases} \tag{31}$$

or

$$\begin{cases} \tilde{R} = \frac{1}{h} P, \ (\textit{resp. } \tilde{R} = P) \\ \tilde{L}\tilde{L}^T = L L^T, \ (\textit{resp. } \tilde{L}\tilde{L}^T = h L L^T) \\ \textit{Either } \theta = \frac{1}{2} \textit{ or } A^T \tilde{R} A = 0 \end{cases} \tag{32}$$

Proof: The form of ansatz (31) reduces (27) to exactly the first continuous dissipative condition (3). Hence if the first continuous dissipative condition (3) is satisfied, under ansatz (31), the first discrete dissipative condition (15) is satisfied. Similarly ansatz (32) reduces (15) to (3). However, it remains to be verified that $\tilde{R}$ satisfies the remaining conditions (16,17). ■

One sees that in (31) the constraint is put on the dissipation, whereas in (32) the constraint is imposed on the energy function. In the ideal case one would like to have $P = R$ and $\mathcal{Q} = \tilde{\mathcal{Q}}$. However Proposition 1 shows this is not possible in general. From (20) it is equivalent to consider $R = P$ and $\tilde{\mathcal{Q}} = h\mathcal{Q}$, or $R = \frac{1}{h} P$ and $\tilde{\mathcal{Q}} = \mathcal{Q}$. In the first case, recalling that $\lambda_{k+\gamma}^T w_{k+\gamma} = 0$, one writes :

$$x_{k+1} P x_{k+1} - x_k P x_k = -\frac{h}{2} (x_k^T \ \lambda_{k+1}^T) \mathcal{Q} \begin{pmatrix} x_k \\ \lambda_{k+1} \end{pmatrix},$$

and in the second case

$$\frac{x_{k+1} P x_{k+1} - x_k P x_k}{h} = -\frac{1}{2} (x_k^T \ \lambda_{k+1}^T) \mathcal{Q} \begin{pmatrix} x_k \\ \lambda_{k+1} \end{pmatrix}.$$

Both options are equivalent and we choose arbitrarily the second one in the sequel. One sees that the second option has the form of the approximation of the continuous-time storage function derivative, with the instantaneous dissipation (hence it approximates the infinitesimal dissipation inequality (8)), whereas the first option rather approximates the integral form (7) of the passivity equality.

2.2.2 Midpoint discretization and its relation to the (θ,γ)-method

In this section a result from [24, 25] is recalled. Consider the continuous and discrete linear invariant systems (as defined in [24]):

$$\begin{cases} \dot{x}(t) = A_c x(t) + B_c \lambda(t) \\ y(t) = C_c x(t) + D_c \lambda(t), \end{cases} \tag{33}$$

and

$$\begin{cases} x_{k+1} = A_d x_k + B_d \lambda_k \\ w_{k+1} = C_d x_k + D_d \lambda_k, \end{cases} \tag{34}$$

where $x \in \mathbb{R}^n, u \in \mathbb{R}^m$ and the constant matrices (A_c, B_c, C_c, D_c) and (A_d, B_d, C_d, D_d) are of appropriate dimensions. Then the (Cayley) transformation defined by:

$$\begin{cases} A_d = (I - A_c)^{-1}(I + A_c) \\ B_d = \sqrt{2}(I - A_c)^{-1} B_c \\ C_d = \sqrt{2} C_c (I - A_c)^{-1} \\ D_d = D_c + C_c (I - A_d)^{-1} B_c + B_c^T (I - A_c^T)^{-1} C_c^T, \end{cases} \tag{35}$$

associates the discrete system (A_d, B_d, C_d, D_d) with the continuous system (A_c, B_c, C_c, D_c) and respectively the inverse transformation:

$$\begin{cases} A_c = (A_d + I)^{-1}(A_d - I) \\ B_c = \sqrt{2}(A_d + I)^{-1} B_d \\ C_c = \sqrt{2} C_d (A_d + I)^{-1} \\ D_c = D_d - C_d (A_d + I)^{-1} B_d - B_d^T (A_d^T + I)^{-1} C_d^T, \end{cases} \tag{36}$$

associates the continuous system (A_c, B_c, C_c, D_c) with the discrete system (A_d, B_d, C_d, D_d).

Theorem 1 *[24, 25] Consider an observable, asymptotically stable linear system defined by (33) (respectively (34)) and the transformations (35) (respectively (36)). Suppose that the pairs (A_c, L) and $(A_d, \tilde{L})$ are controllable. Then one system is passive if and only if the other system is passive, and the energy functions are the same (*i.e. $\mathcal{P} = \mathcal{R}_{\frac{1}{2},\frac{1}{2}}$*).*

Note that Theorem 1 corresponds to the particular case of the $(\theta,\gamma)-$method (referred to as midpoint discretization) with:

$$A_c = \tfrac{h}{2} A, \quad B_c = \tfrac{h}{\sqrt{2}} B, \quad C_c = \tfrac{1}{\sqrt{2}} C, \quad D_c = D + D^T, \tag{37}$$

and

$$A_d = \tilde{A}, \quad B_d = \tilde{B}, \quad C_d = \tilde{C}, \quad D_d = \tilde{D} + \tilde{D}^T, \tag{38}$$

for $\theta = \gamma = \frac{1}{2}$. Further details on the midpoint discretization and how it pertains to dissipativity can be found in [26].

Remark 3 *This approach is to find a particular discretization of* any *passive quadruplet (A_c, B_c, C_c, D_c) that produces a passive quadruplet (A_d, B_d, C_d, D_d), such that both systems share the same storage function set $\mathcal{P}$. The approach we follow next is, given a discretization method, find the classes of*

passive systems (A, B, C, D) which are transformed to a passive discrete-time system $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$, with the same storage function set, reciprocal variables and same dissipation function. It is also worth mentioning the results of [16] that concern the zero-order-hold discretization (ZOH). The storage functions are preserved after ZOH discretization at the price of modifying the output function, hence the supply rate. This is an important point because due to the complementarity constraints, the freedom in the choice of the reciprocal variables for LCS may be restricted. The least requirement is that the method assures that the LCP constructed at each step possesses a unique solution, see section 2.2.4.

In view of Theorem 1 the midpoint method is of interest. An example of the application of this method is given in Example 3, where its limitations are pointed out. In the sequel the midpoint method will be systematically studied from the point of view of the preservation of the dissipation function and the consistency of state jumps approximations, which are not stated in Theorem 1. Also the observability, controllability and asymptotic stability conditions of Theorem 1 are relaxed.

2.2.3 The numerical dissipation

Let us focus on the dissipation functions, *i.e.* the quadratic forms with PSD matrices $\mathcal{Q}$ and $\tilde{\mathcal{Q}}$.

Definition 1 (Numerical dissipation) *The numerical algorithm is said to produce:*

- *Numerical over-dissipation (NOD) if $\mathcal{Q} < \tilde{\mathcal{Q}}$.*
- *Numerical under-dissipation (NUD) if $\mathcal{Q} > \tilde{\mathcal{Q}}$.*
- *Numerical equal-dissipation (NED) if $\mathcal{Q} = \tilde{\mathcal{Q}}$.*
- *Numerical indefinite-dissipation (NID) if $\mathcal{Q} - \tilde{\mathcal{Q}}$ is not a definite matrix.*

Usually one says that a scheme does not dissipate energy when it is of the NED type: if the continuous-time system is lossless, the discrete-time system is lossless as well. We may refine Definition 1 by treating separately the state dissipation (governed by LL^T and $\tilde{L}\tilde{L}^T$) and the input dissipation (governed by W^TW and $\tilde{W}^T\tilde{W}$). We may then define the notion of numerical under state dissipation (NUSD), numerical indefinite state dissipation (NISD), numerical under input dissipation (NUID), numerical equal input dissipation (NEID), numerical over input dissipation (NOID), *etc.* The study in [31] aims at characterizing such properties for the zero order hold discretization method. From Proposition 1 one infers that state dissipation is preserved if and only if:

$$\tilde{L}\tilde{L}^T = (I - h\theta A)^{-T} LL^T (I - h\theta A)^{-1} \text{ when } R = \frac{1}{h}P.$$

Proposition 2 *Under ansatz (31) of Proposition 1 numerical state dissipation is charaterized by the eigenvalues of $-(A^TLL^T + LL^TA)$ for sufficiently small h.*

Proof: Taylor expanding $LL^T - (I - h\theta A)^{-T} LL^T (I - h\theta A)^{-1}$ about $h = 0$, for h sufficiently small yields the desired result. ■

Proposition 3 *Suppose that $\tilde{L}$ is as in (31) and that $L^TA = 0$. Then $\tilde{L}\tilde{L}^T = LL^T$.*

Proof: One has:

$$\begin{aligned}(I_n - h\theta A)^T LL^T (I_n - h\theta A) &= LL^T(I_n - h\theta A) - h\theta A^T LL^T + h^2\theta^2 A^T LL^T A \\ &= LL^T\end{aligned} \tag{39}$$

so that $LL^T = (I_n - h\theta A)^{-T} LL^T (I_n - h\theta A)^{-1} = \tilde{L}\tilde{L}^T$. ■

Clearly the state dissipation is also preserved if $\theta = 0$. Notice that if the constraint is put on the energy function as in ansatz (32) then the dissipation will be preserved but the energy function will be preserved under the same conditions as that of the dissipation in (31). There is therefore a kind of "constraint exchange" between the energy and the storage functions. In general one has the choice between preserving the storage function (and not the dissipation), or perserving the dissipation function (and not the energy function). This depends on the application (numerical simulation or feedback control). In this paper we make the choice of energy function preservation, *i.e.* ansatz (31) with equality (27), but for some control application the ansatz (32) with equality (23) may be preferred.

2.2.4 The One-Step Nonsmooth Problem

At each step advancing the algorithm (13) (equivalently (11)) boils down to solving the following linear complementarity problem (LCP):

$$0 \leqslant \lambda_{k+\gamma} \perp w_{k+1} = \tilde{C}x_k + \tilde{D}\lambda_{k+\gamma} \geqslant 0. \tag{40}$$

From a classical result [17] this LCP has a unique solution for any $\tilde{C}x_k$, if and only if $\tilde{D}$ is a P-matrix. Passivity of the continuous-time system is known to be a crucial property for the discrete-time LCP well-posedness (see [12, Lemma 24] when $\theta = \gamma = 1$, see also [28, 45] for $\theta = 0$, $\gamma = 1$). It is obvious from (12) that when D is a P-matrix, so is $\tilde{D}$ for small enough h or γ. When $D = 0$ then it is necessary that $\gamma > 0$ for otherwise $\tilde{D} = 0$ and the only solution of the LCP (40) is $\lambda_k = 0$ whatever x_k may be.

3 Preservation of passivity properties after discretization

In this section we present conditions for the preservation of the passivity properties after discretization with the (θ,γ)-method. Four cases are analyzed in detail, depending on L and W being zero or not.

3.1 Losslessness preservation $(L = 0, W = 0)$

Let us start with the lossless case. It is noteworthy that usually what is referred to as a conservative system in the literature corresponds to having $L = 0$ solely (the state energy is constant along trajectories). Here the losslessness applies to both the state (the LL^T term) and the "input" (the W^TW term).

3.1.1 The general case

In the lossless case the LMI conditions in (2) are satisfied with equality (*i.e.* with $= 0$ instead of $\leqslant 0$ in the first matrix inequality, or (3-5) with $L = 0, W = 0$). Therefore equivalently:

$$D + D^T = 0, \quad PB = C^T, \quad A^TP + PA = 0. \tag{41}$$

Considering the skew-symmetric feedthrough matrix D is important in applications of nonsmooth circuits [39]. Imposing the continuous conditions (41) for a lossless system on the discrete conditions (15-18) with $L = \tilde{L} = 0$, $W = \tilde{W} = 0$ (*i.e.* $\tilde{Q} = Q = 0$) and taking $P = hR$, we are left with the following sufficient and necessary conditions for the preservation of losslessness under discretization:

$$\begin{cases} A^TRA = 0 \text{ or } \theta = \frac{1}{2} \\ (1-\theta-\gamma)B^TRA = 0 \\ (2\gamma-1)B^T(I-h\theta A)^{-T}R(I-h\theta A)^{-1}B = 0. \end{cases} \tag{42}$$

The conditions (42) are obtained by imposing ansatz (31), $L = \tilde{L} = 0$, and $W = \tilde{W} = 0$ on Lemma 1 in conjuction with noting that if $A^T RA = 0 \Rightarrow RA^2 = 0$ and that when $\theta = \frac{1}{2}$, one can factor out $(I + \frac{h}{2}A)$ from the second condition of the lemma. The equalities in (42) impose not only that the energy function is preserved, but also the dissipation function since $\tilde{\mathcal{Q}} = \mathcal{Q} = 0$. Notice that neither observability nor controllability nor asymptotic stability conditions are required contrarily to Theorem 1. Let us now state a result which consists of finding a class of quadruples (A, B, C, D) such that lossless passivity is preserved.

Proposition 4 *For the following choices of θ and γ, the conditions listed preserve lossless passivity upon discretization:*

- *(i) For $\theta, \gamma \in [0, 1]$:*

$$A^T RA = 0, \quad (B^T RA = 0 \text{ or } 1 - \theta - \gamma = 0), \quad B^T RB = 0. \tag{43}$$

- *(ii) For $\theta = \frac{1}{2}$, $\gamma \in [0, 1]$:*

$$B^T RA = 0, \quad B^T RB = 0. \tag{44}$$

- *(iii) For $\gamma = \frac{1}{2}$, $\theta \in [0, 1]$:*

$$A^T RA = 0, \quad B^T RA = 0. \tag{45}$$

- *(iv) For $\theta = \gamma = \frac{1}{2}$ (midpoint method): the equalities (42) are satisfied for any (A, B, C, D).*

Proof: For the various cases we assume that $P = hR$ and that the continuous conditions (3-5) are satisfied, see (41).

(i) The condition $A^T RA = 0$ together with (41) guarantees that (27) is satisfied with $\tilde{L} = 0$. The condition $A^T RA = 0$ implies that $RA^2 = 0$ since $A^T R = -RA$, and thus $B^T RA^2 = 0$. This plus $B^T RA = 0$ or $1-\theta-\gamma = 0$ guarantees that the second line of (42) is satisfied. Since $A^T RA = 0$ and $A^T R + RA = 0$ we have that $(I - h\theta A)^T R(I - h\theta A) = R$ so that $(I - h\theta A)^{-T} R(I - h\theta A)^{-1} = R$. Thus the condition $B^T RB = 0$ guarantees that the third line of (42) is satisfied.
(ii) With the choice of $\theta = \frac{1}{2}$, automatically we have that the first condition (23) is satisfied, since

$$\begin{array}{c} A^T P + PA = 0 \\ \Downarrow \\ h(A^T R + RA) = 0. \end{array} \tag{46}$$

The second condition of (42) reduces to:

$$(\frac{1}{2} - \gamma) B^T RA = 0 \Rightarrow B^T RA = 0. \tag{47}$$

Finally, using $(I - \frac{h}{2}A)^T R(I - \frac{h}{2}A) = R + \frac{h^2}{4}A^T RA$ on the third condition of (42) yields:

$$\begin{array}{ll} B^T(I - \frac{h}{2}A)^{-T} R(I - \frac{h}{2}A)^{-1} B & = B^T(R - \frac{h^2}{4}A^T(I - \frac{h}{2}A)^{-T} R(I - \frac{h}{2}A)^{-1} A)B \\ & = B^T(R - \frac{h^2}{4}A^T R(I + \frac{h}{2}A)^{-1}(I - \frac{h}{2}A)^{-1} A)B \\ & = B^T RB \\ & = 0. \end{array} \tag{48}$$

(iii) Under the same assumptions as in case (i) for A and the definition of P, we have that the first and third conditions of (42) are satisfied for general $\theta \in [0, 1]$ and $\gamma = \frac{1}{2}$. Similarly, we also have that the second condition of (42) simplifies to

$$B^T RA = 0. \tag{49}$$

(iv) The $\theta=\gamma=\frac{1}{2}$ case, is satisfied from direct inspection of (42). ■

Notice that the conditions for (i) are quite stringent. Indeed $B^TP = C \Rightarrow B^TPB = 0$, so in case $B \neq 0$ necessarily P is low rank. Thus the pair (A, C) cannot be observable, since observability implies that the solutions P of the LMI (2) are positive definite. In particular if $m = 1$ these conditions imply $P = D = 0$ so $C = 0$ and the system cannot be passive since it has a relative degree larger than 1 [27]. We infer that (i) applies only to non-observable multi-input multi-output systems. Here (iv) shows that Theorem 1 can be extended in the sense that the midpoint method preserves also the dissipation function in the lossless case, with relaxed assumptions.

Example 2 *Consider the continuous-time dissipative LMI with (A, B, C, D) and P defined as:*

$$A=\begin{pmatrix} 0 & a_{1,2} & -a_{1,2} & 0\\ -a_{1,2} & 0 & a_{1,2} & 0\\ a_{1,2} & -a_{1,2} & 0 & 0\\ 0 & 0 & 0 & 0 \end{pmatrix},\quad B=\begin{pmatrix} b_1\\ b_2\\ b_3\\ \frac{-p_2\pm\sqrt{p_2^2-p_3p_1}(b_1+b_2+b_3)}{p_1} \end{pmatrix},\quad C=B^TP,\quad D+D^T=0, \tag{50}$$

$$P=\begin{pmatrix} p_1 & p_1 & p_1 & p_2\\ p_1 & p_1 & p_1 & p_2\\ p_1 & p_1 & p_1 & p_2\\ p_2 & p_2 & p_2 & p_3 \end{pmatrix}. \tag{51}$$

Then for $R=\frac{1}{h}P$ with the added conditions $p_2^2-p_3p_1 \geqslant 0$ and $p_1 \neq 0$ (which are required conditions for the entries of B and C to be real), passivity is preserved under any (θ,γ)-discretization.

Remark 4 *A "stiff" passive LCS using the midpoint approximation may exhibit several unwanted characteristics (due to the fact that the midpoint method is not L-Stable [29]). If the 'stiff' system was described by Proposition 4 (i), then one may use the implicit Euler discretization (which is known to be L-Stable), and the discrete system would still be passive.*

Example 3 *Let us consider the system:*

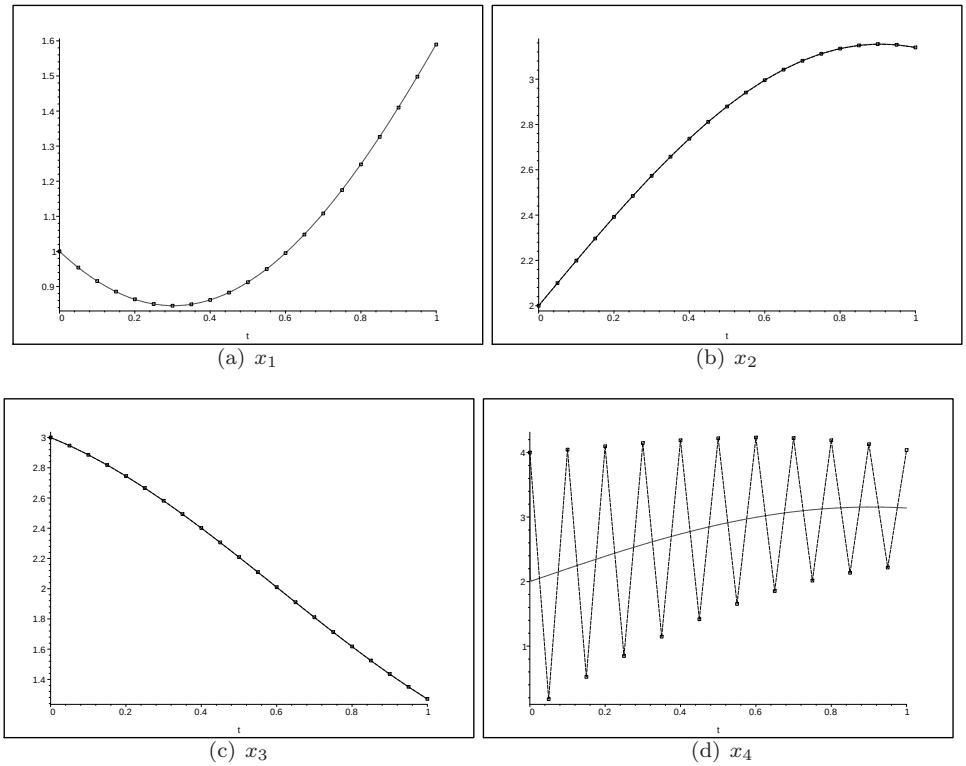
$$A=\begin{pmatrix} 0 & 1 & -1 & 0 & 0\\ -1 & 0 & 1 & 0 & 0\\ 1 & -1 & 0 & 0 & 0\\ 0 & 2000 & 1 & -2000 & 0\\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},\ \textit{with } B^TP-C=0,\ \textit{and } D+D^T=0. \tag{52}$$

Noting that the kernel of A^T is $\{(1\,1\,1\,0\,0)^T,\ (0\,0\,0\,0\,1)^T,\ (0\,0\,0\,0\,0)^T\}$, one can construct P as $P=\sum_{\forall v_i,v_j\in KerA^T} p_{i,j}\ v_i\cdot v_j^T$, where $p_{i,j}=p_{j,i}\quad \forall i,j$. Finally, taking either $\gamma=\frac{1}{2}$ or $CB=0$ results in all required conditions being satisfied. Since the eigenvalues of A (which are 0, -2000 and $\pm\sqrt{-3}$) are of different orders of magnitude using the midpoint method to approximate $\dot{x}=Ax$ results in an undesirable 'oscillatory' behaviour as seen in Figure 1(d). For this particular example we may choose a value of $\theta\neq\frac{1}{2}$ (i.e. $\theta=1$), which yields a much better approximation of x_4 (as seen in Figure 2(d)), but the approximations of x_1, x_2 and x_3 are not as accurate.

3.1.2 The lossless case with $D = 0$

The following case is considered in this section:

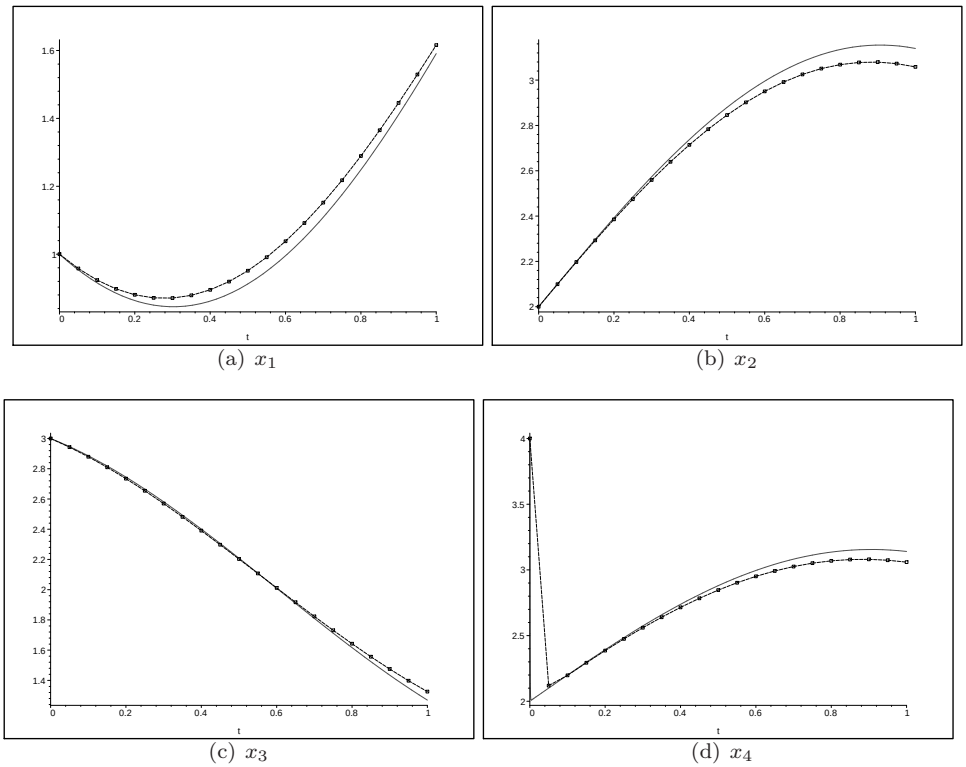
$$D=0,\quad PB=C^T,\quad A^TP+PA=0 \tag{53}$$



(a) x_1 (b) x_2

(c) x_3 (d) x_4

Figure 1: Exact solution (solid line), midpoint method (line with boxes). Time step $h = 0.05$.



(a) x_1 (b) x_2

(c) x_3 (d) x_4

Figure 2: Exact solution (solid line), implicit Euler method (line with boxes). Time step $h = 0.05$.

Let $P = P^T > 0$ (which is guaranteed by the observability of the pair (A, C)) with a square root $\Psi = \Psi^T > 0$ (*i.e.* $\Psi^2 = P$). Without loss of generality we may suppose that A is skew symmetric. Indeed let us make the state vector change $z = \Psi x$. Then:

$$\left\{\begin{array}{l} \dot{z}(t) = \Psi A\Psi^{-1} z(t) + \Psi B\lambda(t) \\ \\ 0 \leqslant \lambda(t) \perp w(t) = C\Psi^{-1} z(t) \geqslant 0, \end{array}\right. \tag{54}$$

which is

$$\left\{\begin{array}{l} \dot{z}(t) = \Psi A\Psi^{-1} z(t) + \Psi^{-1} C^T\lambda(t) \\ \\ 0 \leqslant \lambda(t) \perp w(t) = C\Psi^{-1} z(t) \geqslant 0. \end{array}\right. \tag{55}$$

This system is of the gradient type since by symmetry $\Psi^{-1}C^T = (C\Psi^{-1})^T$. Losslessness then implies that $\Psi A\Psi^{-1} + \Psi^{-1}A^T\Psi = 0$, *i.e.* the matrix $\bar{A} \stackrel{\Delta}{=} \Psi A\Psi^{-1}$ is skew symmetric. Therefore in the lossless case with positive definite symmetric P and $D = 0$ one can work without loss of generality with systems of the gradient form:

$$\left\{\begin{array}{l} \dot{z}(t) = \bar{A} z(t) + \bar{C}^T\lambda(t) \\ \\ 0 \leqslant \lambda(t) \perp w(t) = \bar{C} z(t) \geqslant 0 \end{array}\right. \tag{56}$$

with $\bar{A} + \bar{A}^T = 0$. In other words $P = I$ is a solution of the system's LMI. Note that $\tilde{D} = h\gamma\bar{C}(I - h\theta\bar{A})^{-1}\bar{C}^T$ *i.e.* the discretized version of (56) is not a gradient system, except if $\gamma = \theta = 1$ (a fully implicit scheme).

Proposition 5 *Losslessness of the LCS (56) is conserved under discretization, (*i.e. $\tilde{Q} = Q$*) with $R = \frac{1}{h} I_n$ if one of the following conditions is satisfied:*

- *(i) $\gamma = \frac{1}{2}$, $\theta \in [0, 1]$:*
$$\bar{A} = 0, \tag{57}$$
- *(ii) $\theta = \frac{1}{2}$, $\gamma = \frac{1}{2}$ (midpoint method).*

Proof: The proof of (i) (resp. (ii)) follows from (iii) (resp. (iv)) in Proposition 4. Notice that since $\bar{A}$ is skew-symmetric then $\bar{A}^T\bar{A} = -\bar{A}^2$ and $\bar{A}^2 = 0$ if and only if $\bar{A} = 0$ since the matrices are real. Incidentally we note that $\bar{A} = 0 \Rightarrow A = 0$. The conditions (i) and (ii) of Proposition 4 imply that $\bar{C}\bar{C}^T = 0 \Rightarrow \bar{C} = 0$. ■

Therefore (i) means that when $D = 0$ and $A = 0$ the choice for θ is free, which is obvious from (11).

Example 4 *The scalar case $(A, B, C, D) \equiv (a, b, c, d)$: losslessness implies that $a = 0$, $p = \frac{c}{b} > 0$, $d = 0$. The lossless discrete-time LMI then gives*

$$\left(\begin{array}{cc} 0 & rhb - c \\ \\ rhb - c & -2\gamma chb + rh^2b^2 \end{array}\right) = 0,$$

so that $r = \frac{c}{hb} = \frac{p}{h} > 0$ and $chb = 2\gamma chb$ that is satisfied for $\gamma = \frac{1}{2}$. The $(\theta,\gamma)-$method with $\gamma = \frac{1}{2}$ preserves the energy and the dissipation, whatever θ. The storage function of the continuous-time system is $V(x) = \frac{1}{2}px^2$ and it is $V(x) = \frac{1}{2}\frac{p}{h}x^2$ for the discrete-time system.

Example 5 *Let us consider the triple (A, B, C) of Example 1 in the lossless case, with $\theta = \gamma = \frac{1}{2}$. Noting that A is nilpotent of order n, we can write $\tilde{A} = I + 2 \cdot \sum_{i=1}^{n-1} (\frac{h}{2}A)^i$. Let $R = \frac{1}{h}P$. We obtain $\tilde{A}^T P\tilde{A} - P = 0$ and $-\tilde{D} - \tilde{D}^T + \tilde{B}^T R\tilde{B} = 0$. Thus $\theta = \gamma = \frac{1}{2}$ allows to exactly satisfy (15) of the discrete-time LMI. However $\tilde{A}^T R\tilde{B} - \tilde{C}^T = (0, 0)^T$. The discrete-time reciprocal variables are $\lambda_{k+\frac{1}{2}}$ and $w_{k+\frac{1}{2}} = x_{n,k}$,* i.e. *$\tilde{C} = (0 \ldots 0\ 1)$.*

The conditions of the proposition guarantee the exact preservation of both the energy (equal to $\frac{1}{2}z^Tz$) and the dissipation functions. They yield quite stringent conditions when $\theta \neq \frac{1}{2}$. In practice one may be rather interested in approximate but less stringent conditions, or to preserve the exact energy while not preserving the dissipation, or *vice versa*.

3.2 Preservation of input losslessness ($L \neq 0$, $W = 0$)

Given a continuous passive input-lossless system,

$$\begin{pmatrix} A^TP + PA & PB - C^T \\ B^TP - C & -D - D^T \end{pmatrix} = -\begin{pmatrix} LL^T & 0 \\ 0 & 0 \end{pmatrix} \qquad P = P^T \geqslant 0, \tag{58}$$

with $L \neq 0$ and taking $P = hR$, the conditions needed to conserve the structure of (3) and (5) upon discretization are (see Lemma 1):

$$\begin{cases} \tilde{L}\tilde{L}^T = (I - h\theta A)^{-T}LL^T(I - h\theta A)^{-1} \text{ and either } A^TRA = 0 \text{ or } \theta = \frac{1}{2} \\ hB^TR\big(h(1-\theta-\gamma)A - h^2\theta(\gamma-\theta)A^2\big) - \theta B^T\tilde{L}\tilde{L}^T(I - h\theta A)(I + h(1-\theta)A) = 0 \\ h^2B^T(I - h\theta A)^{-T}((1-2\gamma)R - \gamma\theta LL^T)(I - h\theta A)^{-1}B = 0. \end{cases} \tag{59}$$

Similarly to the previous section (59) is obtained by using ansatz (31) and Lemma 1 with $W = \tilde{W} = 0$.

Proposition 6 *For the following choices of θ and γ the conditions listed preserve input lossless passivity upon discretization:*

- *(i) For $\theta, \gamma \in [0,1]$:*

$$\tilde{L}\tilde{L}^T = (I - h\theta A)^{-T}LL^T(I - h\theta A), \quad A^TRA = 0, \quad B^T\tilde{L} = 0, \quad B^TRB = 0, \quad B^TRA = 0 \text{ or } (1 - \theta - \gamma = 0 \text{ and } B^TRA^2 = 0). \tag{60}$$

- *(ii) For $\theta = \frac{1}{2}$, $\gamma \in [0,1]$:*

$$\tilde{L}\tilde{L}^T = (I - \tfrac{h}{2}A)^{-T}LL^T(I - \tfrac{h}{2}A), \quad B^T\tilde{L} = 0, \quad \tilde{B}^TR\tilde{B} = 0, \quad B^TRA = 0. \tag{61}$$

- *(iii) For $\gamma = \frac{1}{2}$, $\theta \in [0,1]$:*

$$\tilde{L}\tilde{L}^T = (I - h\theta A)^{-T}LL^T(I - h\theta A), \quad A^TRA = 0, \quad B^T\tilde{L} = 0, \quad B^TRA = 0. \tag{62}$$

- *(iv) For $\theta = \gamma = \frac{1}{2}$ (midpoint method):*

$$\tilde{L}\tilde{L}^T = (I - \tfrac{h}{2}A)^{-T}LL^T(I - \tfrac{h}{2}A), \quad B^T\tilde{L} = 0. \tag{63}$$

Proof: The proof follows from Lemma 1 and is similar to the proof of Proposition 4.

Example 6 *Consider the RLCZD circuit given by:*

$$\begin{cases} \dot{x}(t) = Ax(t) + B\lambda(t) \\ w = Cx(t) + D\lambda(t) + a \\ 0 \leqslant \lambda(t) \perp w(t) \geqslant 0, \end{cases} \tag{64}$$

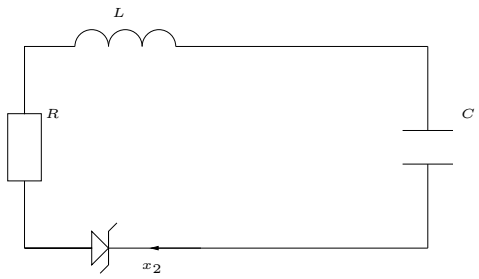


Figure 3: RLCZD circuit.

where

$$A=\begin{pmatrix} 0 & 1 \\ -\frac{1}{lc} & -\frac{r}{l} \end{pmatrix},\quad B=\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},\quad C=\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\quad D=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\quad a=\begin{pmatrix} 0 \\ V_z \end{pmatrix} \tag{65}$$

where x_1 is the charge of the capacitors, x_2 is the current through the inductors, λ is the voltage, $r=0.1$, $l=1$, $c=\frac{1}{4\pi^2}$ and $V_z=5$. The initial conditions of the system are taken to be $x_1(0)=x_2(0)=1$. The system satisfies the continuous-time LMI (2) with

$$P=hR=\begin{pmatrix} 4\pi^2 & 0 \\ 0 & 1 \end{pmatrix},\quad L=\begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix},\quad W=0. \tag{66}$$

Under the midpoint discretization $\theta=\gamma=\frac{1}{2}$, the system becomes

$$\begin{cases} x_{k+1}=\tilde{A}x_k+\tilde{B}\lambda_{k+\frac{1}{2}} \\ w_{k+\frac{1}{2}}=\tilde{C}x_k+\tilde{D}\lambda_{k+\frac{1}{2}}+a \\ 0\leqslant\lambda_{k+\frac{1}{2}}\perp w_{k+\frac{1}{2}}\geqslant 0 \end{cases} \tag{67}$$

where $(\tilde{A},\tilde{B},\tilde{C},\tilde{D})$ are defined as in (12), and $\tilde{L}=(I-\frac{h}{2}A)^{-T}L$. In Figure 5, the phase portrait of the exact solution as well as (θ,γ)-method approximations for $\theta=\frac{1}{2}$ and $\gamma=\frac{1}{2}$ or 1 are given. It is easy to verify that equality one of (59) is satisfied, and equalities two and three of (59) are not satisfied. One may note that under the conditions of $\theta=\gamma=\frac{1}{2}$, from Proposition 6 (iv), that this system fails the extra condition that $B\tilde{L}=0$. The discretized LMI conditions (16) and (17) yield:

$$\tilde{B}^TR\tilde{A}-\tilde{C}=\frac{40h}{(20+h+20h^2\pi^2)^2}\begin{pmatrix} 2h\pi^2 & -1 \\ 0 & 0 \end{pmatrix} \tag{68}$$

and

$$\tilde{B}^TR\tilde{B}-\tilde{D}-\tilde{D}^T=\frac{-20h^2}{(20+h+20h^2\pi^2)^2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{69}$$

The above discrete system (as a consequence of Theorem 1) is dissipative (with $\tilde{W}=\frac{\sqrt{20}h}{20+h+20h^2\pi^2}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $\tilde{L}=\frac{2\sqrt{20}}{20+h+20h^2\pi^2}\begin{pmatrix} 0 & -2h\pi^2 \\ 0 & 1 \end{pmatrix}$). Referring to Definition 1 we have to compare the two dissipation matrices $\mathcal{Q}$ and $\tilde{\mathcal{Q}}$. Obviously for $h=0$ one has $\mathcal{Q}=\tilde{\mathcal{Q}}$. For $h>0$ the plot of the non-zero eigenvalues of $Q-\tilde{Q}$ in Figure 4 show that we positive and negative eigenvalues, thus we can conclude that the midpoint discretization is neither NED, NOD or NUD (according to Definition 1) and is NID. The cumulative dissipation function appears to be decently approximated with the $(\frac{1}{2},\frac{1}{2})$-method (see Figure 6). The relative error of the storage function of the $(\frac{1}{2},\frac{1}{2})$-method,

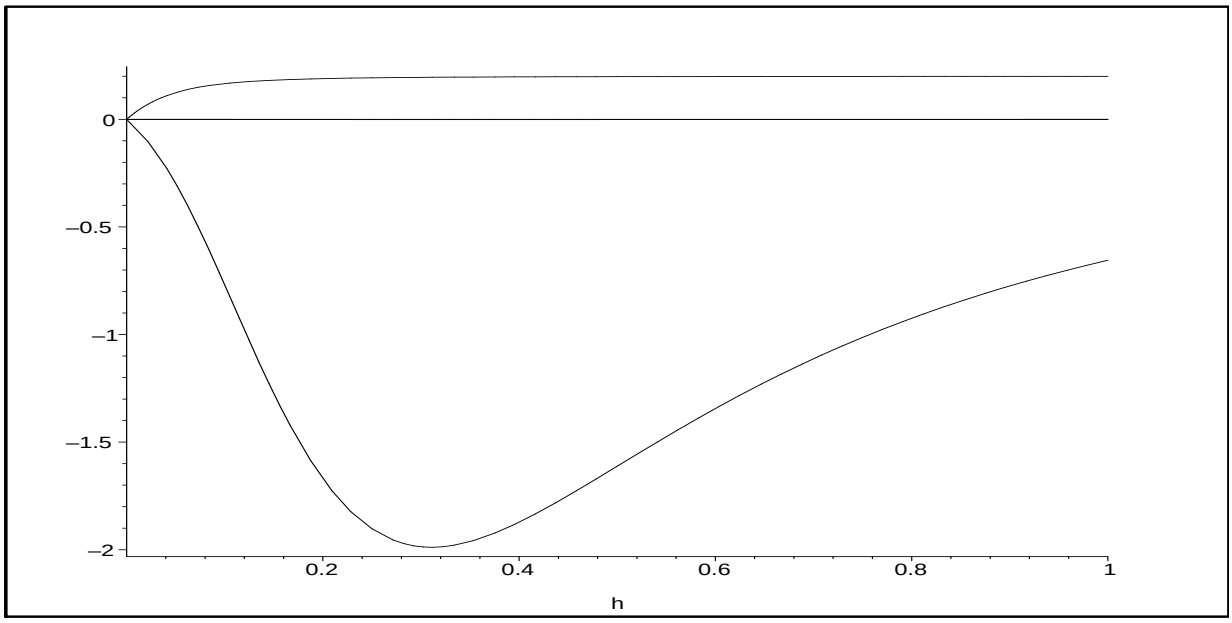


Figure 4: Example 6. RLCZD circuit: Eigenvalues of $Q - \tilde{Q}$ for step size $h \in [0, 1]$ and $\theta = \gamma = \frac{1}{2}$.

and the energy function of the $(\frac{1}{2}, 1)$*-method in comparison to the exact storage/energy function is shown in Figure 6; Clearly, the* $(\frac{1}{2}, \frac{1}{2})$*-method yields a better numerical approximation. For the midpoint method, since* $\tilde{L} \neq L$ *and* $\tilde{W} \neq 0 = W$*, we have both state and input dissipation. By Proposition 2 we calculate the (non-zero) eigenvalues of* $A^T LL^T + LL^T A$ *to be:*

$$-\frac{1 \pm \sqrt{1 + 1600\pi^4}}{50}. \tag{70}$$

Thus the system is NISD. Determining the eigenvalues of $W^T W - \tilde{W}^T \tilde{W}$ *yields that the system is NOID.*

Remark 5 *For the RLCZD circuit under the stated configuration,* $\gamma = \frac{10(h^2\pi^2+1)}{20+h+20h^2\pi^2} \in [0, 1]$ *and* $\theta = \frac{1}{2}$ *guarantee that conditions one and three of (59) are satisfied. Although such a* γ *prevents the discrete system from maintaining the discrete passivity property (since condition two of (59) is not satisfied), it does ensure that the system is state dissipative and NEID.*

Remark 6 *The computation of the exact solution from the various examples consisted of first solving the initial ODE system (with given initial condition)* via MAPLE*'s symbolic ODE solver, then determining the event time (the time that the dynamics switches) by finding the first time in which either the conditions* $Cx + D\lambda \geqslant 0$ *or* $\lambda \geqslant 0$ *are violated, and then solving the new re-initialized ODE system. We continue to employ this procedure until a desired final time is reached.*

Example 7 *For an illustration that* **does** *satisfy (59) (and the conditions from Proposition 6) consider the following system,*

$$\begin{cases} \dot{x}(t) = Ax(t) + B\lambda(t) \\ w(t) = Cx(t) + D\lambda(t) \\ 0 \leqslant \lambda(t) \perp w(t) \geqslant 0, \end{cases} \tag{71}$$

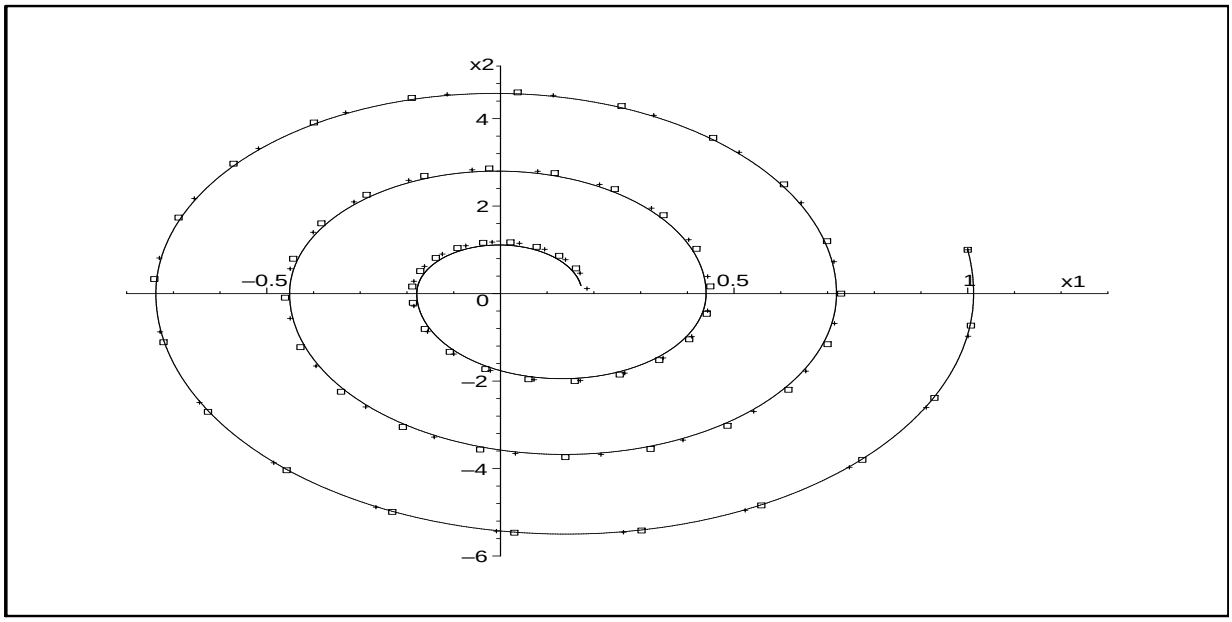


Figure 5: Example 6. RLCZD circuit phase portrait: Exact solution (solid line), $\theta = \gamma = \frac{1}{2}$ (cross) and $\theta = \frac{1}{2}, \gamma = 1$ (box). Time step $h = 0.05$.

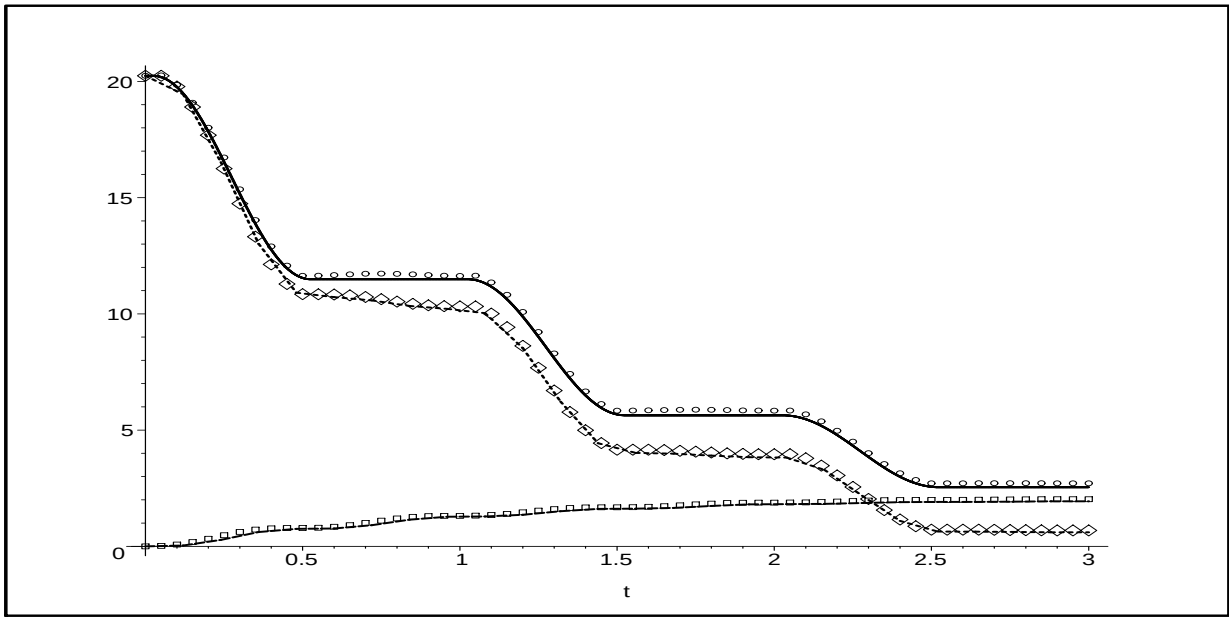


Figure 6: Example 6. RLCZD circuit: Exact storage function (dotted line), exact cumulative dissipation function (dashed line), exact storage function + cumulative dissipation function (solid line), $(\frac{1}{2}, \frac{1}{2})$-method approximation of storage function (diamond), $(\frac{1}{2}, \frac{1}{2})$-method approximation of cumulative dissipation function (box), $(\frac{1}{2}, \frac{1}{2})$-method approximation of storage function + cumulative dissipation function (circle). Time step $h = 0.05$.

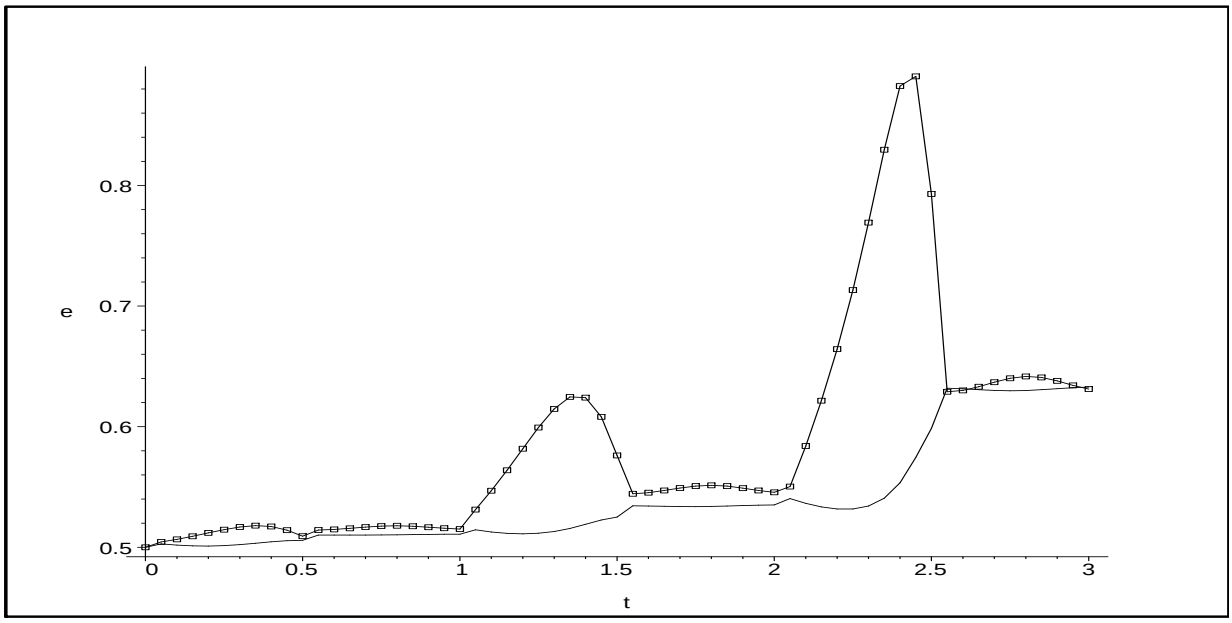


Figure 7: Example 6. RLCZD circuit: Relative error plot of the storage function approximations by the (θ,γ)-method with $\theta=\gamma=\frac{1}{2}$ (solid line) and $\theta=\frac{1}{2},\gamma=1$ (box-line) with respect to the exact storage function. Time step $h=0.05$.

with (A,B,C,D) *defined as:*

$$A=\begin{pmatrix} 0 & 1 \\ -\frac{1}{lc} & -\frac{r}{l} \end{pmatrix},\quad B=\begin{pmatrix} \frac{1}{\pi} \\ 2h\pi \end{pmatrix},\quad C=\begin{pmatrix} 4\pi & 2h\pi \end{pmatrix},\quad D=0 \tag{72}$$

with $r=0.1$, $l=1$, *and* $c=\frac{1}{4\pi^2}$. *The system satisfies the continuous LMI conditions (3-6) with*

$$P=hR=\begin{pmatrix} 4\pi^2 & 0 \\ 0 & 1 \end{pmatrix},\quad L=\left(0,\frac{1}{\sqrt{5}}\right)^T,\quad W=0. \tag{73}$$

Under the midpoint discretization, the system becomes

$$\begin{cases} x_{k+1}=\tilde{A}x_k+\tilde{B}\lambda_{k+\frac{1}{2}} \\ w_{k+\frac{1}{2}}=\tilde{C}x_k+\tilde{D}\lambda_{k+\frac{1}{2}}+a \\ 0\leqslant\lambda_{k+\frac{1}{2}}\perp w_{k+\frac{1}{2}}\geqslant 0, \end{cases} \tag{74}$$

where $(\tilde{A},\tilde{B},\tilde{C},\tilde{D})$ *are defined as in (12) and* $\tilde{L}=(I-\frac{h}{2}A)^{-T}L$ *and* $\tilde{W}=W=0$. *The choice of* $P=hR$ *and* $\theta=\gamma=\frac{1}{2}$ *ensures that the discrete LMI conditions (15-17) are satisfied (as well as the preservation conditions (59), and the extra condition from Proposition 6, that* $B^T\tilde{L}\tilde{L}^T=0$*). Referring to Definition 1 we have to compare the two dissipation matrices* $\mathcal{Q}$ *and* $\tilde{\mathcal{Q}}$. *Obviously for* $h=0$ *one has* $\mathcal{Q}=\tilde{\mathcal{Q}}$. *From figure 8 for* $h>0$ *we have one positive and one negative eigenvalue of* $Q-\tilde{Q}$, $\tilde{W}=W$, *and that* $-(A^TLL^T+LL^TA)$ *is the same as in Example 6, we conclude that the midpoint discretization is NID and NISD, but it is NEID (see Definition 1).*

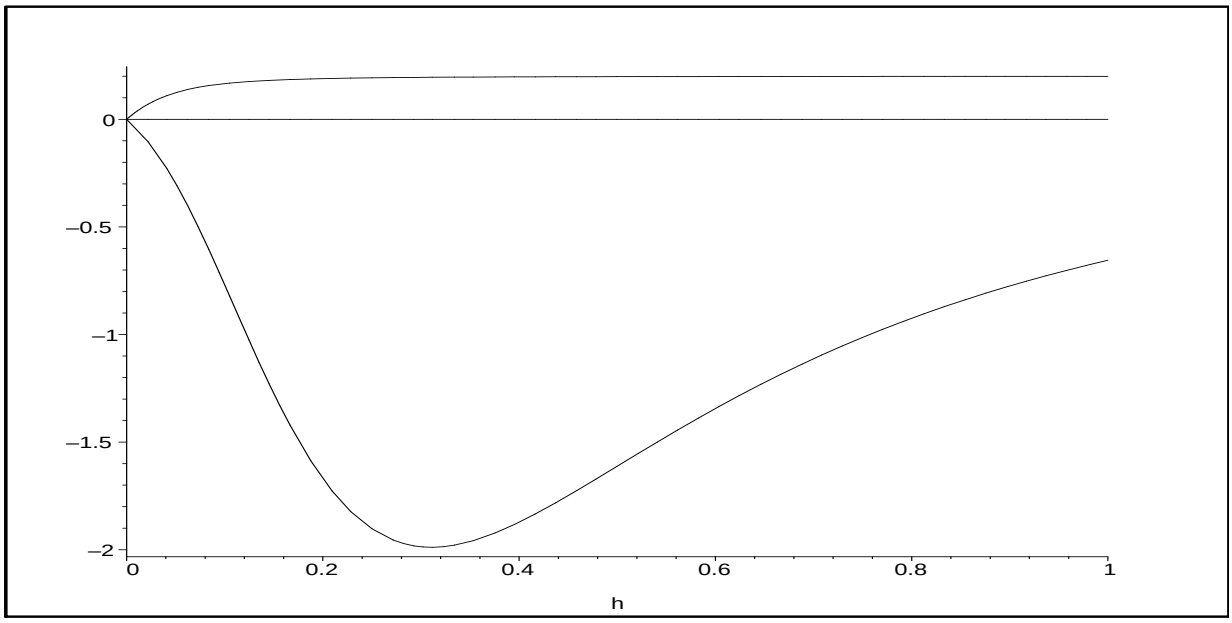


Figure 8: Example 7: Preservation of input lossless structure under (θ,γ)-discretization. Plot of eigenvalues of $Q-\tilde{Q}$ for step size $h\in[0,1]$ and $\theta=\gamma=\frac{1}{2}$.

3.3 Preservation of state losslessness ($L=0, W\neq 0$)

The form of this class of passive system is (3-5) with $L=0$ and $W\neq 0$. Taking $P=hR$, the conditions needed to conserve the passive 'structure' upon discretization are:

$$\begin{cases} A^TRA=0 \text{ or } \theta=\frac{1}{2} \\ h^2(1-\theta-\gamma)B^TRA=0 \\ h^2(1-2\gamma)B^T(I-h\theta A)^{-T}R(I-h\theta A)^{-1}B=W^TW-\tilde{W}^T\tilde{W}. \end{cases} \tag{75}$$

These conditions are once again obtained as the special case of Lemma 1 and ansatz (31) with $L=\tilde{L}=0$. In the special case of $\theta=\gamma=\frac{1}{2}$ the conditions in (75) reduce to,

$$\tilde{W}^T\tilde{W}=W^TW \;\Rightarrow\; \tilde{W}=W. \tag{76}$$

Thus the midpoint preserves both storage and dissipation functions.

Proposition 7 *For the following choices of θ and γ the conditions listed preserve passivity upon discretization:*

- *(i) For $\theta,\gamma\in[0,1]$:*

$$A^TRA=0,\quad (B^TRA=0 \text{ or } 1-\theta-\gamma=0),\quad B^TRB=0,\quad \tilde{W}=W. \tag{77}$$

- *(ii) For $\theta=\frac{1}{2}$, $\gamma\in[0,1]$:*

$$B^TRA=0,\quad B^TRB=0,\quad \tilde{W}=W. \tag{78}$$

- *(iii) For* $\gamma = \frac{1}{2}$, $\theta \in [0,1]$*:*

$$A^T R A = 0, \quad B^T R A = 0, \quad \tilde{W} = W. \tag{79}$$

- *(iv) For* $\theta = \gamma = \frac{1}{2}$ *(midpoint method): The equality (75) is satisfied for any* (A, B, C, D) *with* $\tilde{W} = W$*, so the algorithm is NED.*

Once again (iv) extends Theorem 1 since it characterizes the dissipation function.
Proof: The proof follows from Lemma 1 and is similar to the proof of Proposition 4.

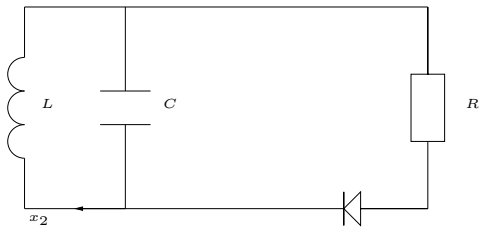


Figure 9: RLCD circuit.

Example 8 *Consider the RLCD circuit from Figure 9 whose dynamics is given by:*

$$\begin{cases} \dot{x}(t) = Ax(t) + B\lambda(t) \\ w(t) = Cx(t) + D\lambda(t) \\ 0 \leqslant \lambda(t) \perp w(t) \geqslant 0, \end{cases} \tag{80}$$

where

$$A = \begin{pmatrix} 0 & -1 \\ \frac{1}{lc} & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad C = \begin{pmatrix} \frac{1}{c} & 0 \end{pmatrix}, \quad D = r, \tag{81}$$

where x_1 *is the charge of the capacitors,* x_2 *is the current through the inductors and* λ *is —,* $r = 10$*,* $l = 1$*, and* $c = \frac{1}{4\pi^2}$*. The initial conditions of the system are* $x_1(0) = x_2(0) = 1$*. Here we have* $P = \begin{pmatrix} \frac{1}{c} & 0 \\ 0 & l \end{pmatrix}$*,* $W = \sqrt{2r}$*,* $L = 0$*. Thus* $\tilde{L} = 0$ *since* $L = 0 \Leftrightarrow \tilde{L} = 0$*. Furthermore, taking* $\tilde{W} = W$*,* $\theta = \gamma = \frac{1}{2}$ *yields that all conditions of (75) are satisfied. For this example, since* $\tilde{L} = L$ *and* $\tilde{W} = W$ *we have necessarily that* $\tilde{Q} = Q$*, and hence the RLCD circuit falls into the class of NED systems (see Definition 1). The plot of the cumulative dissipation (as seen in Figure 11) demonstrates that the* $(\frac{1}{2}, \frac{1}{2})$*-method yields decent numerical approximation of the exact dissipation function. The plot of the relative errors between the* $(\frac{1}{2}, \frac{1}{2})$*-method approximation of the storage function and the exact storage/energy function, as well as the energy function of the* $(\frac{1}{2}, 1)$*-method and the exact storage/energy function is depicted in Figure 12. Clearly the* $(\frac{1}{2}, \frac{1}{2})$*-method yields a better approximation.*

Example 9 *Let us consider the configuration of the 4-diode bridge illustrated in figure 13. The resistor inside the bridges is supplied by a LC oscillator. The dynamical equations are stated choosing:*

$$x = \begin{pmatrix} V_L \\ I_L \end{pmatrix}, w = \begin{pmatrix} I_{DR1} \\ I_{DF2} \\ V_2 - V_1 \\ V_1 - V_3 \end{pmatrix}, \quad \lambda = \begin{pmatrix} V_2 \\ -V_3 \\ I_{DF1} \\ I_{DR2} \end{pmatrix}, \tag{82}$$

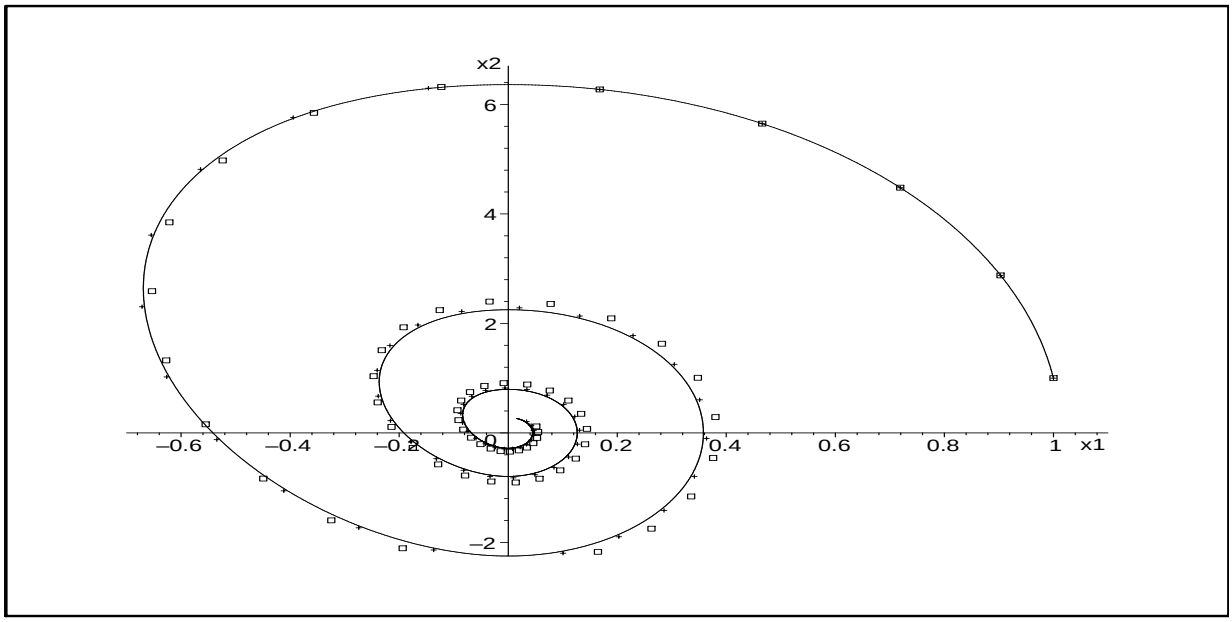


Figure 10: Example 8. RLCD circuit phase portrait: Exact solution (solid line), $\theta = \gamma = \frac{1}{2}$ (cross) and $\theta = \frac{1}{2}, \gamma = 1$ (box). Time step $h = 0.05$

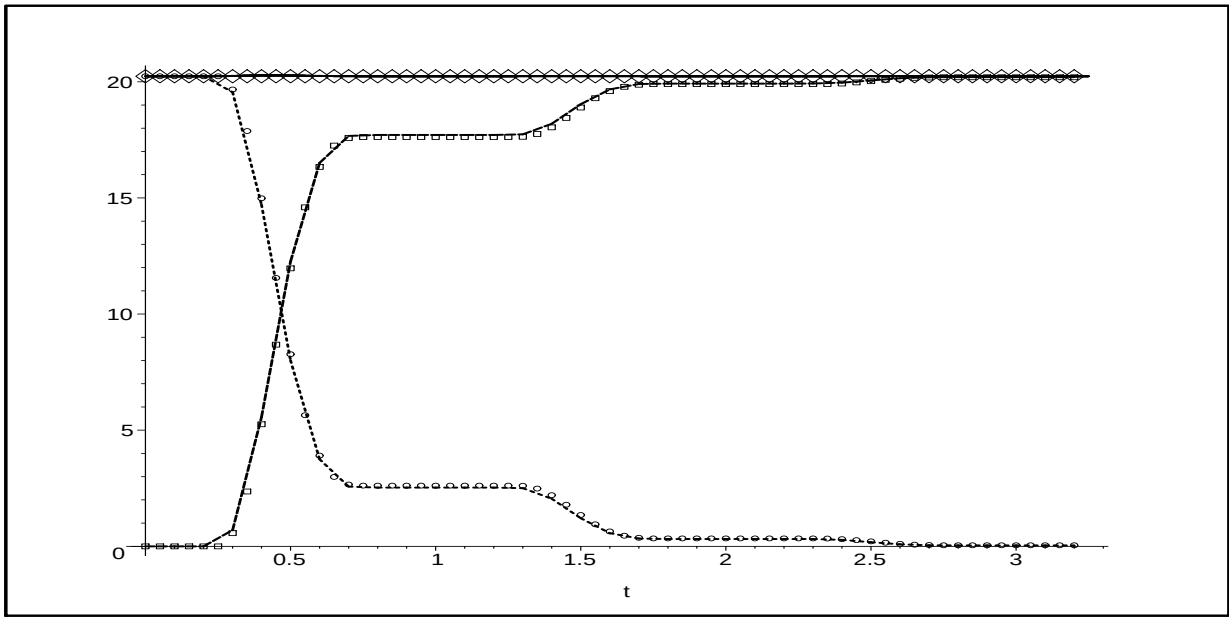


Figure 11: Example 8. RLCD circuit: Exact storage function (dotted line), exact cumulative dissipation function (dashed line), exact storage function + cumulative dissipation function (solid line), $(\frac{1}{2}, \frac{1}{2})$-method approximation of storage function (circle), $(\frac{1}{2}, \frac{1}{2})$-method approximation of cumulative dissipation function (box), $(\frac{1}{2}, \frac{1}{2})$-method approximation of storage function + cumulative dissipation function (diamond). Time step $h = 0.05$.

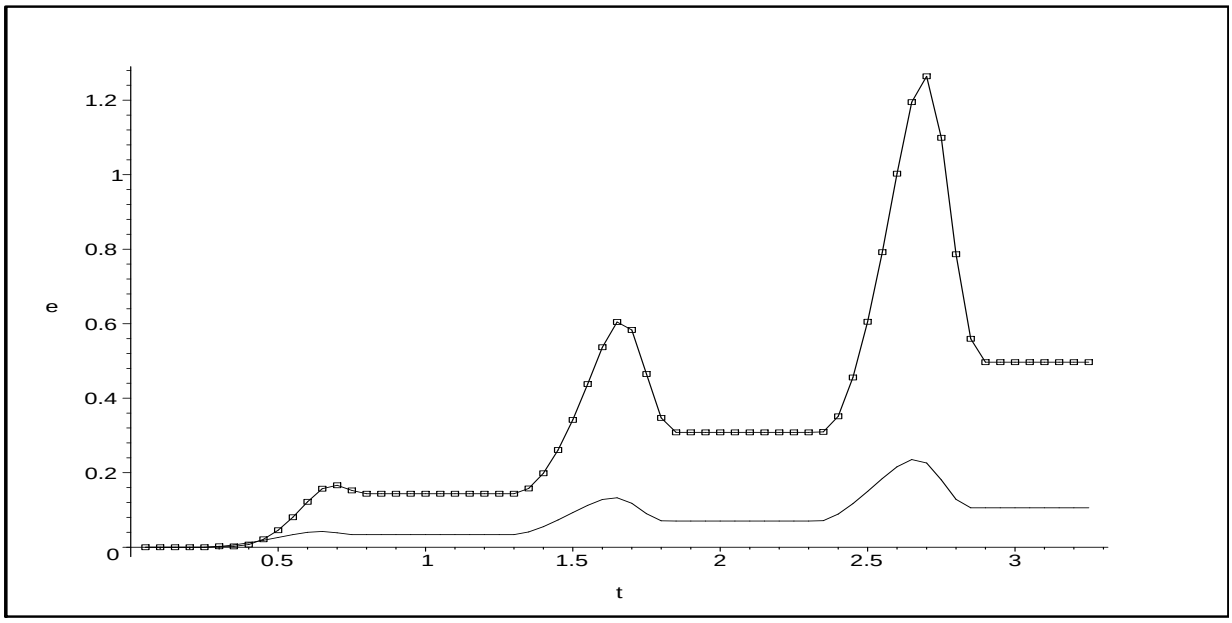


Figure 12: Example 8 RLCD circuit: Relative error plot of the storage function approximations by the (θ, γ)-method with $\theta = \gamma = \frac{1}{2}$ (solid line) and $\theta = \frac{1}{2}, \gamma = 1$ (box-line) with respect to the exact storage function. Time step $h = 0.05$.

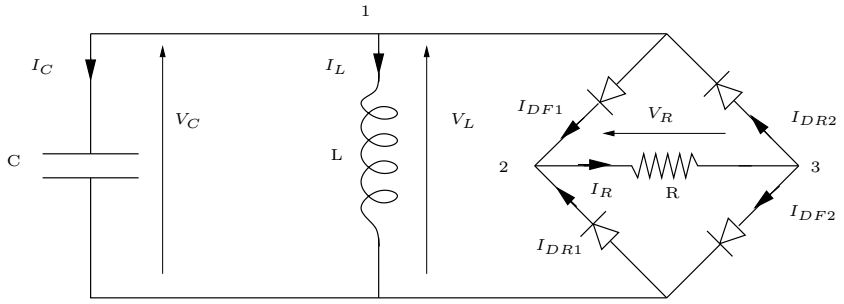


Figure 13: LC oscillator with a load resistor

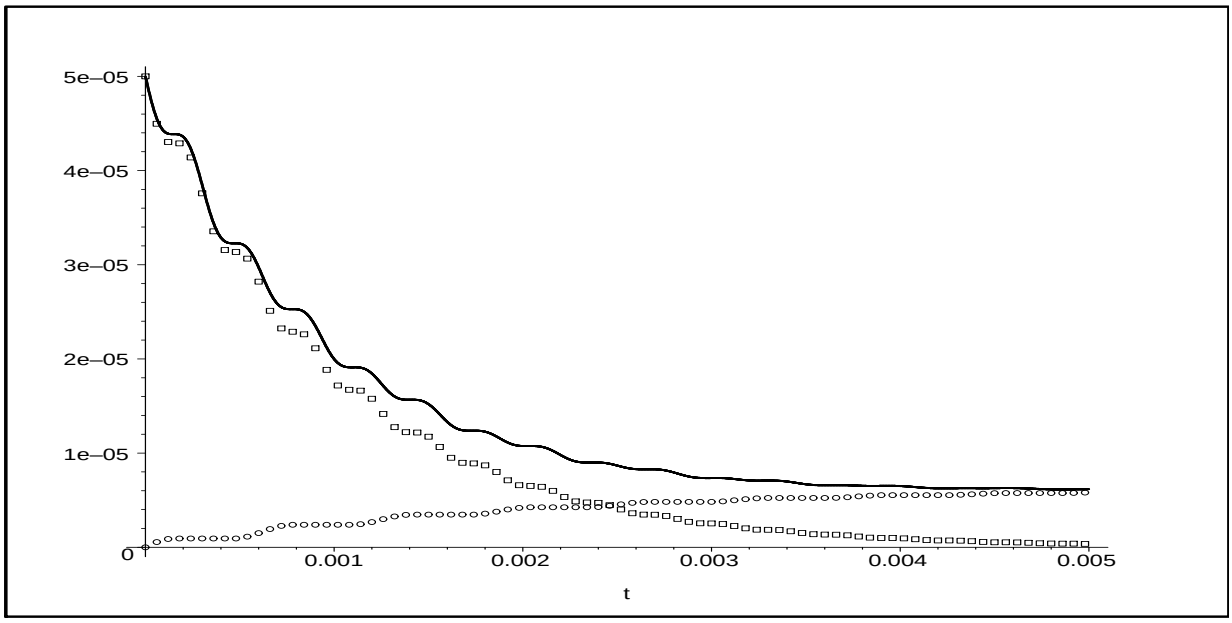


Figure 14: Example 9. LC oscillator: $(\frac{1}{2},\frac{1}{2})$-method approximation of storage function (box), $(\frac{1}{2},\frac{1}{2})$-method approximation of cummulative dissipation function (circle), $(\frac{1}{2},\frac{1}{2})$-method approximation of storage function + cummulative dissipation function (solid line). Time step $h = 1\times10^{-6}$.

and with

$$A = \begin{pmatrix} 0 & -1/c \\ 1/L & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & -1/c & 1/c \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$C = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ 1 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 1/R & 1/R & -1 & 0 \\ 1/R & 1/R & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \tag{83}$$

The storage function matrix is:

$$P = \begin{pmatrix} c & 0 \\ 0 & l \end{pmatrix}. \tag{84}$$

Also:

$$W = \begin{pmatrix} \frac{1}{\sqrt{r}} & \frac{1}{\sqrt{r}} & 0 & 0 \\ \frac{1}{\sqrt{r}} & \frac{1}{\sqrt{r}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{85}$$

and $L = 0$. The initial conditions and parameters of the system are taken as, $x_1(0) = 10$, $x_2(0) = 0$, $l = 1\times10^{-2}$, $c = 1\times10^{-6}$, and $r = 1\times10^{3}$. Using the $(\frac{1}{2},\frac{1}{2})$-method, one can verify that $\tilde{L} = 0$ and $\tilde{W} = W$. Hence we have that the system is NED for this particular discretization (since $\mathcal{Q} - \tilde{\mathcal{Q}} = 0$). For this configuration, the matrix D has full rank, so the solution $x(t)$ is a function of class C^1 [8, 15].

3.4 General preservation conditions ($L \neq 0, W \neq 0$)

Taking $P = hR$, we recall the general conditions stated in Lemma 1 for preservation of dissipativity under discretization under ansatz (31) are:

$$\begin{cases} \tilde{L}\tilde{L}^T = (I - h\theta A)^{-T} L L^T (I - h\theta A)^{-1} \text{ and either } A^T R A = 0 \text{ or } \theta = \frac{1}{2} \\ \\ hB^T R(h(1-\theta-\gamma)A - h^2\theta(\gamma-\theta)A^2) - \theta B^T \tilde{L}\tilde{L}^T (I - h\theta A)(I + h(1-\theta)A) \\ = W^T L^T (I + h(\gamma-\theta)A)(I + h\theta A) - \tilde{W}^T \tilde{L}^T (I - h\theta A)(I + h\theta A) \\ \\ h^2 B^T (I - h\theta A)^{-T} \big((1-2\gamma)R - \gamma\theta L L^T\big)(I - h\theta A)^{-1} B \\ = W^T W - \tilde{W}^T \tilde{W} + h\gamma(B^T (I - h\theta A)^{-T} L W + W^T L^T (I - h\theta)A)^{-1} B) \end{cases} \tag{86}$$

For the special case of $\theta = \gamma = \frac{1}{2}$ the general conditions reduce to:

$$\begin{cases} \tilde{L}\tilde{L}^T = (I - \frac{h}{2}A)^{-T} L L^T (I - \frac{h}{2}A)^{-1} \\ \\ -\frac{1}{2} B^T \tilde{L}\tilde{L}^T (I - \frac{h}{2}A) = W^T L^T - \tilde{W}^T \tilde{L}^T (I - \frac{h}{2}A) \\ \\ \frac{h^2}{4} B^T \tilde{L}\tilde{L}^T B = \tilde{W}^T \tilde{W} - W^T W - h\gamma(B^T (I - \frac{h}{2}A)^{-T} L W + W^T L^T (I - \frac{h}{2}A)^{-1} B) \end{cases} \tag{87}$$

Proposition 8 *Suppose $\tilde{W} = W$ and $\tilde{L} = (I - \frac{h}{2}A)^{-T} L$, then the midpoint method conditions in (87) reduce to satisfying the two equations:*

$$B^T \tilde{L}\tilde{L}^T = 0 \quad B^T \tilde{L} W + W^T \tilde{L}^T B = 0 \tag{88}$$

Proof: Direct substitution of $\tilde{W}$ and $\tilde{L}$ into (87) yield the desired result. Thus for $\theta = \gamma = \frac{1}{2}$, if $\tilde{L} = (I - \frac{h}{2}A)^{-T} L$, $B^T \tilde{L}\tilde{L}^T = 0$ and $B^T \tilde{L} W + W^T \tilde{L}^T B = 0$, then all three equations in (87) are satisfied with $\tilde{W} = W$ and $P = hR$. ■
Proposition 3 can be applied to guarantee the state dissipation preservation.

Example 10 *Consider the system given by:*

$$A = -\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} -\frac{3}{4} \\ 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & \frac{1}{2} \end{pmatrix}, \quad D = \frac{1}{4}. \tag{89}$$

The initial conditions of the system are taken to be $x_1(0) = x_2(0) = 1$. The system satisfies the continuous-time LMI (2) and the discrete-time LMI (14) with

$$P = hR = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}, \quad L = \sqrt{2}\,(2\ 1)^T, \quad W = \tfrac{1}{\sqrt{2}}, \quad \tilde{L} = \sqrt{2}(\tfrac{4}{2+h}\ \tfrac{2}{2+h})^T, \quad \tilde{W} = \tfrac{\sqrt{2}}{2+h}. \tag{90}$$

Determining the type of numerical dissipativity of the system, one finds that the (only) non-zero eigenvalue of $\mathcal{Q} - \tilde{\mathcal{Q}}$ is:

$$\frac{21h(4+h)}{2(2+h)^2}. \tag{91}$$

For $h > 0$ the non-zero eigenvalue is positive, thus the system is NUD. The numerical under dissipation can clearly be seen in Figure 15, as the exact dissipation function is greater than the $(\frac{1}{2}, \frac{1}{2})$-method approximation for all $t \in (0, 20]$.

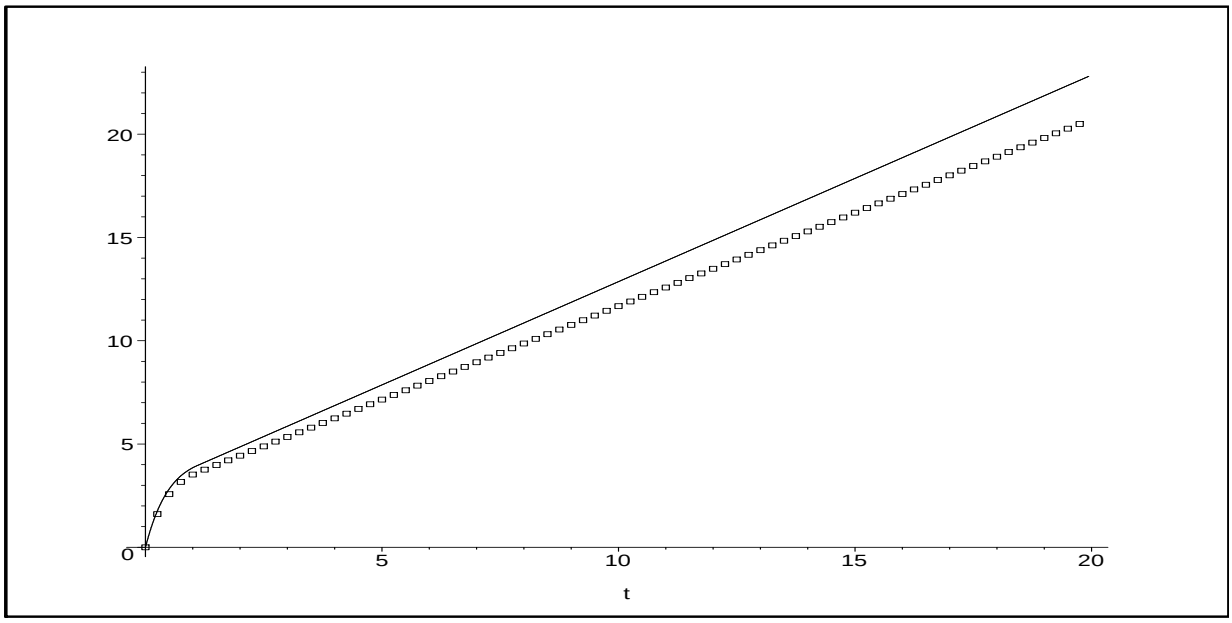


Figure 15: Example 10. Cumulative dissipation of a numerical under dissipative system: Exact Solution (solid line), and $\theta = \gamma = \frac{1}{2}$ (cross). Time step $h = 0.05$.

3.4.1 Approximations

In order to maintain the proper form of continuous dissipativity conditions when we isolate for powers of h, the substitution $hR = P$ is made in (15-16). We also assume that $\tilde{L} = (I - h\theta A)^{-T}L$ and $\tilde{W} = W$, and that $h\theta A$ has eigenvalues with modulus < 1 so that the expansions are meaningful [23, p.329]. Thus (15) reduces to:

$$(A^T P + PA) + h(1 - 2\theta)A^T PA = -(I - h\theta A)^T \tilde{L}\tilde{L}^T (I - h\theta A). \tag{92}$$

Taylor expanding about h to second order and isolating for the various powers of h yields:

Order of h	Condition
0	$A^T P + PA = -LL^T$
1	$(1 - 2\theta)A^T PA = 0$
2	$(1 - 2\theta)A^T(A^T P + PA)A - \theta(PA^3 + (A^3)^T P) = 0$

(93)

Following the same proceedure for (16):

Order of h	Condition
0	$B^T P - C = -W^T L^T$
1	$(1 - \theta)B^T PA + \theta B^T A^T P - (\gamma - \theta)CA = 0$
2	$\theta(1 - \theta)B^T A^T PA + \theta^2 B^T (A^2)^T P = 0$

(94)

and for (17):

Order of h	Condition
0	$D + D^T = W^T W^T$
1	$B^T PB - \gamma(CB + B^T C^T) = 0$
2	$\theta[B^T(A^T P + PA)B - \gamma(CAB + B^T A^T C^T)] = 0$

(95)

Some comments follow:

- The zero order conditions are equivalent to the continuous-time conditions, as expected.
- The conditions under which the first and second order conditions are satisfied implies that the matrices A, B, P (or R), C, D, L and W satisfy the same type of constraints as for the above exact preservation conditions.
- These results nevertheless indicate at which accuracy the dissipativity properties may be satisfied after discretization.

4 Systems with state jumps

Complementarity systems as in (1) may undergo state jumps (for instance initially if $D = 0$ and $Cx(0^-) < 0$). They may be seen as a switching system that switches between DAEs, where the number of constraints of the DAEs may vary: complementarity systems may live on lower-dimensional subspaces. The switches are ruled by complementarity conditions. The state jumps are necessary to re-initialize the system so that the right limit of the state is an admissible initial data for the new mode (the new DAE). The first point to fix here is a modelling issue. Depending on the application the state re-initialization may take different forms. In [39, section 1.1.5] it is shown on a circuit example that the $\theta-$method with $\theta = \gamma = 1$ is able to approximate state jumps for inconsistent states. There are mainly two sources of state discontinuities: the first one is associated with inconsistent initial values and the second one is due to the external excitation term $Fv(t)$ in (1) which may move the state outside the feasible region defined by the complementarity condition, see [9, 13].

If some state jumps are expected, the state $x(t)$ is usually assumed to be a right continuous function of local bounded variations (RCLBV)[34, 35, 36], or of special bounded variations (RCLSBV) [6]. The variable λ has to be replaced by a measure that contains Dirac distributions. In the same vein, the time–derivative of the state $x(t)$ cannot be considered in the usual sense but as a differential measure dx associated with a RCLBV function $x(t)$ [34]. In the following we shall assume that the solution of (1) is RCLSBV. Consequently the dynamics in (1) is written in terms of a measure differential equation as:

$$dx = Ax(t)dt + Eu(t)dt + Bd\Lambda, \tag{96}$$

where $d\Lambda$ is a measure associated in the following way with $\lambda(t)$. The absolutely continuous function $\lambda(t)$ is the density of $d\Lambda$ with respect to the Lebesgue measure, *i.e.*:

$$\frac{d\Lambda}{dt}(t) = \lambda(t). \tag{97}$$

Since we assume that solutions are RCLSBV, a decomposition of the measure can be written as [36, 40]:

$$d\Lambda = \lambda(t)dt + \sum_i \sigma_i \delta_{t_i}, \tag{98}$$

where δ_{t_i} is the Dirac measure at time of discontinuities t_i and σ_i the amplitude. Using this decomposition, the differential measure equation (96) can be written as a smooth dynamics:

$$\dot{x}(t) = Ax(t) + Eu(t) + B\lambda(t), \quad dt - \text{almost everywhere}, \tag{99}$$

and a jump dynamics at t_i:

$$x(t_i^+) - x(t_i^-) = B\sigma_i. \tag{100}$$

The jump dynamics (100) is not sufficient to determine uniquely the state $x(t_i^+)$ after a discontinuity. A jump rule needs to be stated which has to be consistent with the complementarity conditions. In the sequel, the following energy-based jump rule in Definition 2 will be used. This jump rule follows from [9, 11, 18], and is inspired by Moreau's generalized impact law for lagrangian systems (see [37] and [40, §2.7]).

Definition 2 (State Jump Law) *Let us consider the dynamics in (1), and suppose that (A, B, C, D) is passive with storage function $V(x) = \frac{1}{2}x^T Px$, $P = P^T > 0$. Let us define the set $K(t) = \{z \in \mathbb{R}^n \mid Cz + Fv(t^+) \in Q_D^*\}$, with $Q_D = \{z \in \mathbb{R}^m \mid z \geqslant 0, Dz \geqslant 0, z^T Dz = 0\}$. For any $x(t^-)$ the state after the state discontinuities $x(t^+)$ is given by:*

$$P(x(t^+) - x(t^-)) \in -N_K(x(t^+)). \tag{101}$$

To simplify the notation in the following we denote $K(t)$ as K. Notice that Q_D is a closed convex cone. Equivalent expressions for (101) are given in appendix C. It can be shown that the state jump law uniquely defines the state after the discontinuities provided that the qualification constraint $Fv(t) \in Q_D^* + \mathrm{Im}(C)$, where Q_D^* is the dual cone of Q_D, is satisfied (see also [9, Proposition 3.2] for similar conditions that hold when $D = 0$). Notice that this qualification constraint is equivalent to $K(t) \neq \emptyset$ which by (221) assures indeed that the projection is unique. Furthermore, the post-jump state $x(t^+)$ is consistent with the complementarity system's dynamics on the right of t.

Lemma 2 *The state jump law in (101) guarantees that $V(x(t^+)) - V(x(t^-)) \leqslant 0$ provided that $0 \in K$.*

Proof: Direct from (221) because $0 \in K$ assures that the projection makes the norm of $x(t^+)$ in the metric defined by P, smaller than that of $x(t^-)$. ■

If $v(t) = 0$ (autonomous system) then the jump occurs initially and dissipates energy since the condition is always satisfied [13]. The condition $0 \in K$ may also be satisfied for $v(t) \neq 0$, see example 11. The time-discretization of (97) has to take into account the nature of the solution to avoid point-wise evaluations of measures at atoms. A direct application of the scheme (11) is not consistent with possible jumps in the state. Let us consider that the scheme (11) is used with $x_k = x(t^-)$ and we expect to have a jump at time t such that $x(t^+) - x(t^-) = \sigma \neq 0$. If the scheme is consistent, we expect to have $\lim_{h\to 0} x_{k+1} = x(t^+)$. For $B \neq 0$, the scheme implies that $\lim_{h\to 0} \lambda_{k+\gamma} = \infty$. This reveals a point–wise evaluation of a measure. Only the measures of the time–intervals $(t_k, t_{k+1}]$ are considered such that:

$$dx((t_k, t_{k+1}]) = \int_{t_k}^{t_{k+1}} Ax(t) + Eu(t)\, dt + Bd\Lambda((t_k, t_{k+1}]). \tag{102}$$

By definition of a differential measure, we have:

$$dx((t_k, t_{k+1}]) = x(t_{k+1}^+) - x(t_k^+). \tag{103}$$

The measure of the time–interval by $d\Lambda$ is kept as an unknown variable denoted by:

$$\sigma_{k+1} \approx d\Lambda((t_k, t_{k+1}]). \tag{104}$$

Finally, the remaining Lebesgue integral in (102) is approximated by the θ-method:

$$\int_{t_k}^{t_{k+1}} Ax(t) + Eu(t)\,dt \approx h(Ax_{k+\theta} + Eu_{k+\theta}), \tag{105}$$

yielding the following integration formula for (96):

$$x_{k+1} - x_k = h(Ax_{k+\theta} + Eu_{k+\theta}) + B\sigma_{k+1}. \tag{106}$$

In the following sections, we try to answer the following questions: Is the scheme based on the integration rule (106) able to consistently approximate the jump rule of Definition 2? We will also consider the direct application of the scheme (11).

In this section, we will also assume that the RCLSBV solution $x(t)$ exists and that the following schemes based on (106) generate a bounded sequences $\{x_k\}$ and $\{\sigma_k\}$ for a sufficiently small h. Especially, for one time–step, given the values of x_k and σ_k, we assume that

$$\lim_{h\to 0} x_{k+1} < +\infty \text{ and } \lim_{h\to 0} \sigma_{k+1} < +\infty. \tag{107}$$

Three cases are analyzed in the following: $D = 0$, $D \geqslant 0$ with a special structure, and the general case $D \geqslant 0$. Indeed $D > 0$ implies that the solutions are contiunous of class C^1, and is of no interest in this section.

4.1 The case $D = 0$

We consider in this section the following time–stepping scheme whenever $D = 0$:

$$\begin{cases} x_{k+1} - x_k = h\left(Ax_{k+\theta} + Eu_{k+\theta}\right) + B\sigma_{k+1} \\ w_{k+1} = Cx_{k+1} + Fv_{k+1} \\ 0 \leqslant w_{k+1} \perp \sigma_{k+1} \geqslant 0. \end{cases} \tag{108}$$

Proposition 9 *Let us assume that $D = 0$ and that (A, B, C) is passive with storage function $V(x) = \frac{1}{2}x^T Px$, $P = P^T > 0$. The scheme (108) consistently approximates the jump rule of Definition 2 in the sense that for $h = 0$, x_{k+1} solves (101) for $x_k = x(t^-)$ and $v(t^+) = v_{k+1}$. Moreover, if $K \neq \emptyset$, we have*

$$\lim_{h\to 0} \|x_{k+1} - x(t^+)\| = 0 \tag{109}$$

for $x(t^+)$ given by the state jump rule (101) and any $x_k = x(t^-)$.

Proof: If $D = 0$ then $Q_D = I\!R^m_+ = Q^*_D$. Therefore K reduces to:

$$K = \{x \in I\!R^n \mid Cx + Fv(t^+) \geqslant 0\}. \tag{110}$$

For the jump law (101), we get:

$$\begin{cases} P(x(t^+) - x(t^-)) = C^T\sigma \\ w = Cx(t^+) + Fv(t^+) \\ 0 \leqslant w \perp \sigma \geqslant 0. \end{cases} \tag{111}$$

For $h = 0$ in (108), x_{k+1} solves:

$$\begin{cases} x_{k+1} - x_k = B\sigma_{k+1} \\ w_{k+1} = Cx_{k+1} + Fv_{k+1} \\ 0 \leqslant w_{k+1} \perp \sigma_{k+1} \geqslant 0. \end{cases} \tag{112}$$

When $D=0$, the passivity assumption implies that $PB=C^T$. Since $P>0$, we get:

$$\left\{\begin{array}{l} P(x_{k+1}-x_k)=C^T\sigma_{k+1} \\ \\ w_{k+1}=Cx_{k+1}+Fv_{k+1} \\ \\ 0\leqslant w_{k+1}\perp\sigma_{k+1}\geqslant 0, \end{array}\right. \tag{113}$$

that is

$$-P(x_{k+1}-x_k)\in N_K(x_{k+1}). \tag{114}$$

Let us now consider the case $h\neq 0$. The jump law (101) is an Affine Variational Inequality (AVI) written as an inclusion into a normal cone, that is:

$$-(Mz+q)\in N_K(z), \tag{115}$$

with $M=P$ and $q=-Px(t^-)$. Let us denote by $\mathrm{AVI}(K,q,M)$ the problem (115) and by $\mathrm{SOL}(K,q,M)$ the solution set of $\mathrm{AVI}(K,q,M)$. Since $M=P>0$, the AVI is strongly monotone. In our case, the set of solution is reduced to a singleton for any value of $x(t^-)$ if $K\neq\emptyset$, that is:

$$\mathrm{SOL}(K,q,M)=\{x(t^+)\}. \tag{116}$$

For $h\neq 0$ and multiplying the first line of (108) by P, the one–step problem can be rewritten as:

$$\left\{\begin{array}{l} P(I-h\theta A)x_{k+1}=P(I+h(1-\theta)A)x_k+hPEu_{k+\theta}+C^T\sigma_{k+1} \\ \\ w_{k+1}=Cx_{k+1}+Fv_{k+1} \\ \\ 0\leqslant w_{k+1}\perp\sigma_{k+1}\geqslant 0. \end{array}\right. \tag{117}$$

The system (117) is the solution of the $\mathrm{AVI}(K,q_{k+1},M_{k+1})$ with

$$\begin{array}{l} M_{k+1}=P(I-h\theta A) \\ q_{k+1}=-P(I+h(1-\theta)A)x_k+hPEu_{k+\theta}. \end{array} \tag{118}$$

Since M_{k+1} is positive definite for sufficiently small h, and since $K\neq\emptyset$, the set of solution of $\mathrm{AVI}(K,q_{k+1},M_{k+1})$ reduces to a singleton:

$$\mathrm{SOL}(K,q_{k+1},M_{k+1})=\{x_{k+1}\}. \tag{119}$$

Furthermore, since

$$\lim_{h\to 0}\|M-M_{k+1}\|=0, \qquad \lim_{h\to 0}\|q-q_{k+1}\|=0, \tag{120}$$

and we assume that $\lim_{h\to 0}x_{k+1}<+\infty$ and $\lim_{h\to 0}\sigma_{k+1}=\sigma_\infty<+\infty$ we have that $\lim_{h\to 0}\|x_{k+1}-x(t^+)\|=0$ from (117).■

Let us now discuss the direct application of the scheme (11). The scheme in this form cannot be consistent if a jump is expected. Indeed, the variable $\lambda_{k+\gamma}$ diverges to $+\infty$ as h vanishes. Mimicking scheme (108), we can define the following:

$$\left\{\begin{array}{l} x_{k+1}-x_k=h\left(Ax_{k+\theta}+Eu_{k+\theta}\right)+B\sigma_{k+\gamma} \\ \\ w_{k+\gamma}=Cx_{k+\gamma}+Fv_{k+\gamma} \\ \\ 0\leqslant w_{k+\gamma}\perp\sigma_{k+\gamma}\geqslant 0. \end{array}\right. \tag{121}$$

For $h = 0$, x_{k+1} solves:

$$\left\{ \begin{array}{l} P(x_{k+1} - x_k) = C^T \sigma_{k+\gamma} \\ \\ w_{k+\gamma} = \gamma C x_{k+1} + (1-\gamma) C x_k + F v_{k+\gamma} \\ \\ 0 \leqslant w_{k+\gamma} \perp \sigma_{k+\gamma} \geqslant 0. \end{array} \right. \tag{122}$$

which is equivalent to the inclusion:

$$-P(x_{k+1} - x_k) = N_{K_\gamma}(x_{k+1}). \tag{123}$$

where:

$$K_\gamma = \{x \in I\!R^n \mid \gamma C x + (1-\gamma) x_k + F v_{k+\gamma}\}. \tag{124}$$

This prevents a consistent scheme for $\gamma \neq 1$ because the cone K_γ depends not only on γ but also on x_k. For $\gamma = 1$, we exactly recover the scheme (108) which is consistent with the jump law.

Remark 7 *The midpoint discretization with $\theta = \gamma = \frac{1}{2}$ therefore does not consistently approximate the jumps according to Proposition 9. One gets $x_{k+1} = proj_P[K_{\frac{1}{2}}(x_k, v_{k+\frac{1}{2}}); x_k]$.*

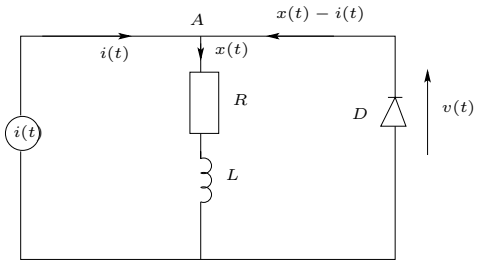


Figure 16: A circuit with an ideal diode, a resistor, an inductor and a current source.

Example 11 *Let us consider the circuit in Figure 16, that is composed of an ideal diode mounted in parallel with an inductor/resistor (L/R) and a current source $i(t)$. The current through the inductor/resistor is denoted as $x(t)$ and the following dynamics is obtained:*

$$\left\{ \begin{array}{l} \dot{x}(t) = -\dfrac{R}{L} x(t) + \lambda(t) \\ \\ 0 \leqslant w(t) = x(t) - i(t) \perp \lambda(t) \geqslant 0. \end{array} \right. \tag{125}$$

The variable $\dfrac{\lambda(t)}{L}$ is the voltage across the diode, and $(A, B, C, D, E, F) = (-\frac{R}{L}, 1, 1, 0, 0, -1)$. Notice now from the complementarity condition of (125) that the state $x(t)$ is unilaterally constrained as $x(t) \geqslant i(t)$ for all $t \geqslant 0$. Suppose that at some time $t \geqslant 0$ this constraint is violated, due to a jump in $i(t)$. The jump law in Definition 2 has to be applied and amounts to solving in this special case:

$$x(t^+) = x(t^-) + \max[0, i(t^+) - x(t^-)] \;\Leftrightarrow\; x(t^+) = proj[K; x(t^-)], \tag{126}$$

with $K = \{z \in I\!R \mid z - i(t^+) \geqslant 0\}$. Let us now apply the scheme (108), i.e.*:*

$$\left\{ \begin{array}{l} x_{k+1} - x_k = -h\left(\frac{R}{L} x_{k+\theta}\right) + \sigma_{k+1} \\ \\ y_{k+1} = x_{k+1} - i_{k+1} \\ \\ 0 \leqslant y_{k+1} \perp \sigma_{k+1} \geqslant 0. \end{array} \right. \tag{127}$$

The assumption of Proposition 9 holds because the transfer function of the system $(A,B,C) = (-\frac{R}{L},1,1)$ *is equal to* $\frac{1}{s+\frac{R}{L}}$, $s\in\mathbb{C}$, *that is positive real hence the system is passive [27]. One can verify that for* $h=0$ *we solve:*

$$\left\{\begin{array}{l} x_{k+1}-x_k=\sigma_{k+1}\\ \\ w_{k+1}=x_{k+1}-i_{k+1}\\ \\ 0\leqslant w_{k+1}\perp\sigma_{k+1}\geqslant 0. \end{array}\right. \tag{128}$$

that is:

$$\left\{\begin{array}{l} \sigma_{k+1}=\max[0,i_{k+1}-x_k]\\ \\ x_{k+1}=x_k+\sigma_{k+1}=x_k+\max[0,i_{k+1}-x_k]. \end{array}\right. \tag{129}$$

which is exactly the discrete counterpart of (126). For $h\neq 0$, *we have:*

$$\left\{\begin{array}{l} \sigma_{k+1}=\max[0,\left(1+h\theta\dfrac{R}{L}\right)i_{k+1}-\left(1-h(1-\theta)\dfrac{R}{L}\right)x_k]\\ \\ x_{k+1}=\left(1+h\theta\dfrac{R}{L}\right)^{-1}\left[\left(1-h(1-\theta)\dfrac{R}{L}\right)x_k+\sigma_{k+1}\right]. \end{array}\right. \tag{130}$$

and the limit (109) holds. For the scheme (121) and $h=0$, *we get:*

$$\left\{\begin{array}{l} \sigma_{k+\gamma}=\max[0,\dfrac{1}{\gamma}(i_{k+1}-x_k)]\\ \\ x_{k+1}=x_k+\sigma_{k+\gamma}=x_k+\max[0,\dfrac{1}{\gamma}(i_{k+1}-x_k)]. \end{array}\right. \tag{131}$$

If there is a jump, the magnitude of σ_{k+1} *is proportional to* $\dfrac{1}{\gamma}$. *This scheme cannot yield a consistent state jump with* $\gamma\neq 1$.

4.2 The case when D has a special structure

In this section, a special structure of the matrix $D\in\mathbb{R}^{m\times m}$ is considered as:

$$D=\left[\begin{array}{cc}\tilde{D}&0\\0&0\end{array}\right],\tilde{D}\in\mathbb{R}^{d\times d},d<m,\quad \tilde{D}>0 \tag{132}$$

(see [8] for the analysis of such cases). The following scheme is considered, that is a variation of (11):

$$\left\{\begin{array}{l} x_{k+1}-x_k=h\left(Ax_{k+\theta}+u_{k+\theta}\right)+h\tilde{B}\lambda_{k+\gamma}+\hat{B}\sigma_{k+1}\\ \tilde{w}_{k+\gamma}=\tilde{C}x_{k+\gamma}+\tilde{F}v_{k+\gamma}+\tilde{D}\lambda_{k+\gamma}\\ \hat{w}_{k+1}=\hat{C}x_{k+1}+\hat{F}v_{k+1}\\ 0\leqslant\tilde{w}_{k+\gamma}\perp\lambda_{k+\gamma}\geqslant 0\\ 0\leqslant\hat{w}_{k+1}\perp\sigma_{k+1}\geqslant 0. \end{array}\right. \tag{133}$$

Mimicking [8] the following decompositions are performed:

$$C=\left[\begin{array}{c}\tilde{C}\\ \hat{C}\end{array}\right],F=\left[\begin{array}{c}\tilde{F}\\ \hat{F}\end{array}\right]\text{ with }\tilde{C},\tilde{F}\in\mathbb{R}^{d\times n},\quad B=\left[\begin{array}{cc}\tilde{B}&\hat{B}\end{array}\right]\text{ with }\tilde{B}\in\mathbb{R}^{n\times d}. \tag{134}$$

Due to the structure of D, the cone Q_D is given by:

$$Q_D=\{0\}^d\times\mathbb{R}_+^{m-d}, \tag{135}$$

and then

$$K=\mathbb{R}^d\times\hat{K}\text{ with }\hat{K}=\{x\in\mathbb{R}^n|\hat{C}x+\hat{F}v(t^+)\geqslant 0\}. \tag{136}$$

Proposition 10 *Let us assume that the matrix D has the special structure (132) and that (A, B, C, D) is passive with storage function $V(x) = \frac{1}{2}x^T Px$, $P = P^T > 0$. The scheme (133) consistently approximates the jump rule of Definition 2 in the sense that for $h = 0$, x_{k+1} solves the jump rule (101) for $x_k = x(t^-)$ and $v_{k+1} = v(t^+)$. Moreover, if $\hat{K} \neq \emptyset$, we have*

$$\lim_{h\to 0} \|x_{k+1} - x(t^+)\| = 0, \tag{137}$$

for $x(t^+)$ given by the state jump rule (101) and any $x_k = x(t^-)$.

Proof: The jump law of Definition 2 reduces to:

$$\begin{cases} P(x(t^+) - x(t^-)) = \hat{C}^T\sigma \\ \\ 0 \leqslant \hat{C}x(t^+) + \hat{F}v(t^+) \perp \sigma \geqslant 0. \end{cases} \tag{138}$$

If the system is passive and D has the structure (132), from [27, Lemma A.64] we have:

$$P\hat{B} = \hat{C}^T. \tag{139}$$

Hence since $P > 0$ (138) is equivalent to:

$$\begin{cases} x(t^+) - x(t^-) = \hat{B}\sigma, \\ \\ 0 \leqslant \hat{C}x(t^+) + \hat{F}v(t^+) \perp \sigma \geqslant 0. \end{cases} \tag{140}$$

For $h = 0$, the scheme (133) reads as:

$$\begin{cases} x_{k+1} - x_k = \hat{B}\sigma_{k+1} \\ \tilde{y}_{k+\gamma} = \tilde{C}x_{k+\gamma} + \tilde{F}v_{k+\gamma} + \tilde{D}\lambda_{k+\gamma} \\ \hat{w}_{k+1} = \hat{C}x_{k+1} + \hat{F}v_{k+1} \\ 0 \leqslant \tilde{w}_{k+\gamma} \perp \lambda_{k+\gamma} \geqslant 0 \\ 0 \leqslant \hat{w}_{k+1} \perp \sigma_{k+1} \geqslant 0. \end{cases} \tag{141}$$

Since $\lambda_{k+\gamma}$ does not play any role in the first line of (141), we can simplify to:

$$\begin{cases} x_{k+1} - x_k = \hat{B}\sigma_{k+1} \\ \\ 0 \leqslant \hat{C}x_{k+1} + \hat{F}v_{k+1} \perp \sigma_{k+1} \geqslant 0. \end{cases} \tag{142}$$

which amounts to solving the jump rule (140) for x_{k+1}. Since $\tilde{D} > 0$, $\mathrm{SOL}(\tilde{D}, q)$ is a singleton for any q, we can denote the solution of the LCP:

$$0 \leqslant \tilde{C}x_{k+\gamma} + \tilde{F}v_{k+\gamma} + \tilde{D}\lambda_{k+\gamma} \perp \lambda_{k+\gamma} \geqslant 0 \tag{143}$$

as follows:

$$\lambda_{k+\gamma} = W(x_{k+\gamma}, v_{k+\gamma}) \tag{144}$$

where W is a Lipschitz continuous function of its argument[17]. The scheme (133) reads as:

$$\begin{cases} x_{k+1} - x_k = h\left(Ax_{k+\theta} + u_{k+\theta}\right) + h\tilde{B}W(x_{k+\gamma}, v_{k+\gamma}) + \hat{B}\sigma_{k+1} \\ \hat{w}_{k+1} = \hat{C}x_{k+1} + \hat{F}v_{k+1} \\ 0 \leqslant \hat{w}_{k+1} \perp \sigma_{k+1} \geqslant 0. \end{cases} \tag{145}$$

Equivalently:

$$-\mathcal{F}(x_{k+1}) \in N_{\hat{K}}(x_{k+1}), \tag{146}$$

with

$$\mathcal{F}(x) = P(x - x_k - h\left(A(\theta x + (1-\theta)x_k) + u_{k+\theta}\right) - h\tilde{B}W(\gamma x + (1-\gamma)x_k, v_{k+\gamma}). \tag{147}$$

Since $P > 0$ one has:

$$\begin{array}{ll}(x-y)^T(\mathcal{F}(x)-\mathcal{F}(y)) & = (x-y)^TP(x-y) - h\theta(x-y)^TPA(x-y) \\ & -h(x-y)^TP\tilde{B} \quad [W(\gamma x + (1-\gamma)x_k, v_{k+\gamma}) - W(\gamma y + (1-\gamma)x_k, v_{k+\gamma})] \\ & \geqslant \alpha\|x-y\|^2 \end{array} \tag{148}$$

for some $\alpha > 0$ and for sufficiently small h. Consequently the variational inequality (146) is strongly monotone. A solution exists and is unique if $\hat{K} \neq \emptyset$. Since the following trivially holds

$$\lim_{h\to 0} \sup_{x\in V(x(t^+))} (\mathcal{F}(x) - P(x - x(t^-))) = 0, \tag{149}$$

where $V(x)$ is a neighborhood of x, we conclude that (109) is satisfied. ■

Remark 8 *In all section 4 the dissipativity of (A,B,C,D) is used but it is not necessary. In fact only the properties that $PB = C^T$ (or its variants) and D positive semi definite are used in the developments, similarly to [8, 9]. But the system needs not be stable.*

Example 12 *Let us consider the electrical system of Figure 17 that is composed of two resistors R with voltage/current law $u(t) = Ri(t)$, four capacitors C with voltage/current law $C\dot{u}(t) = i(t)$, and two ideal diodes with characteristics $0 \leqslant v_1(t) \perp i_1(t) \geqslant 0$ and $0 \leqslant v_2(t) \perp i_3(t) \geqslant 0$ respectively. The state variables are $x_1(t) = \int_0^t i_1(t)dt$, $x_2(t) = \int_0^t i_2(t)dt$, $x_3(t) = v_2(t)$, and $\lambda_2(t) = i_3(t)$, $\lambda_1(t) = v_1(t)$. The dynamics of this circuit is given by:*

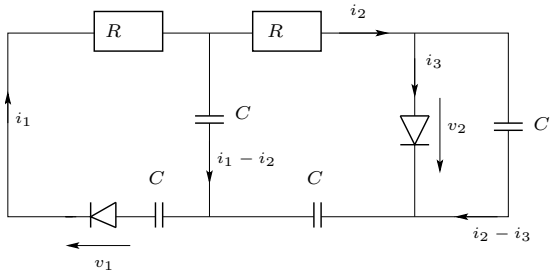


Figure 17: Electrical circuit with capacitors, resistors and ideal diodes.

$$\begin{array}{l}\begin{pmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\\ \dot{x}_3(t)\end{pmatrix} = \begin{pmatrix}\frac{-2}{RC} & \frac{1}{RC} & 0\\ \frac{1}{RC} & \frac{-2}{RC} & \frac{1}{R}\\ -\frac{1}{RC^2} & \frac{2}{RC^2} & -\frac{1}{RC}\end{pmatrix}\begin{pmatrix}x_1(t)\\ x_2(t)\\ x_3(t)\end{pmatrix} + \begin{pmatrix}\frac{1}{R} & 0\\ 0 & 0\\ 0 & \frac{1}{C}\end{pmatrix}\lambda(t) \\ 0 \leqslant \lambda(t) \perp w(t) = \begin{pmatrix}\frac{-2}{RC} & \frac{1}{RC} & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}x_1(t)\\ x_2(t)\\ x_3(t)\end{pmatrix} + \begin{pmatrix}\frac{1}{R} & 0\\ 0 & 0\end{pmatrix}\lambda(t) \geqslant 0\end{array} \tag{150}$$

The matrices P, L and W that solve the LMI (3)–(5) are given by:

$$P = \begin{pmatrix}\frac{2}{c} & \frac{-1}{c} & 0\\ \frac{-1}{c} & \frac{2}{c} & 0\\ 0 & 0 & c\end{pmatrix} \quad L = \sqrt{\tfrac{2}{r}}\begin{pmatrix}\frac{1}{c} & \frac{-2}{c}\\ \frac{-2}{c} & \frac{1}{c}\\ 1 & 0\end{pmatrix} \quad W = \begin{pmatrix}\sqrt{\frac{2}{r}} & 0\\ 0 & 0\end{pmatrix} \tag{151}$$

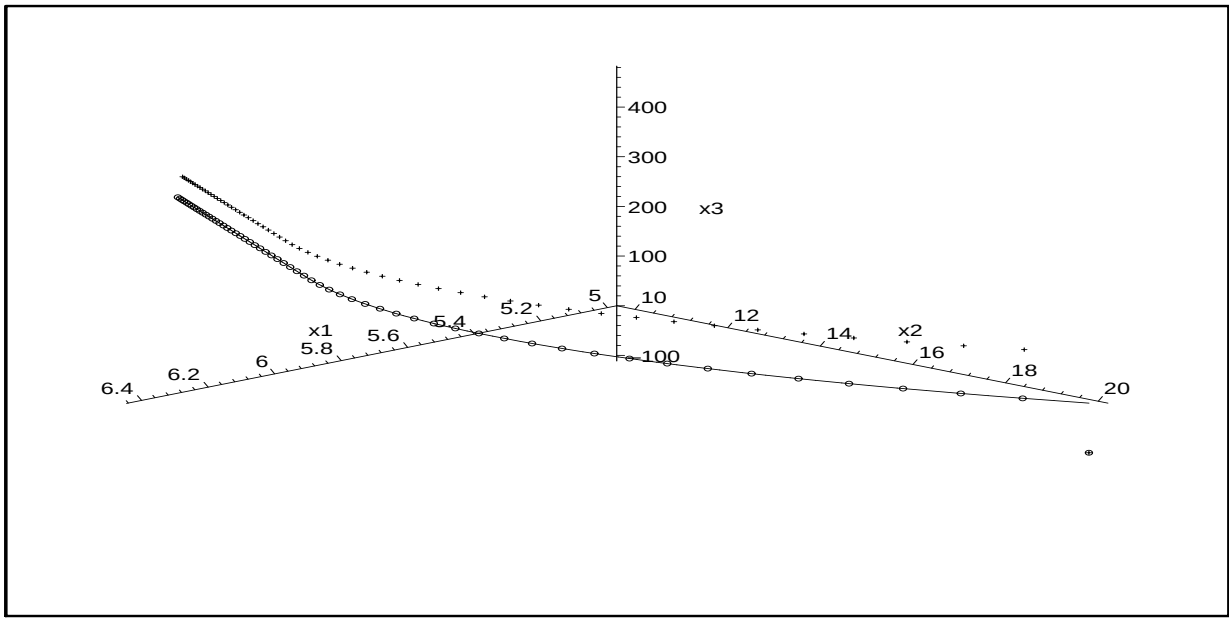


Figure 18: Example 12. RCD circuit phase portrait: Exact solution (solid line), $\theta = \gamma = \frac{1}{2}$ (cross), and $\theta = \frac{1}{2}, \gamma = 1$ (box).

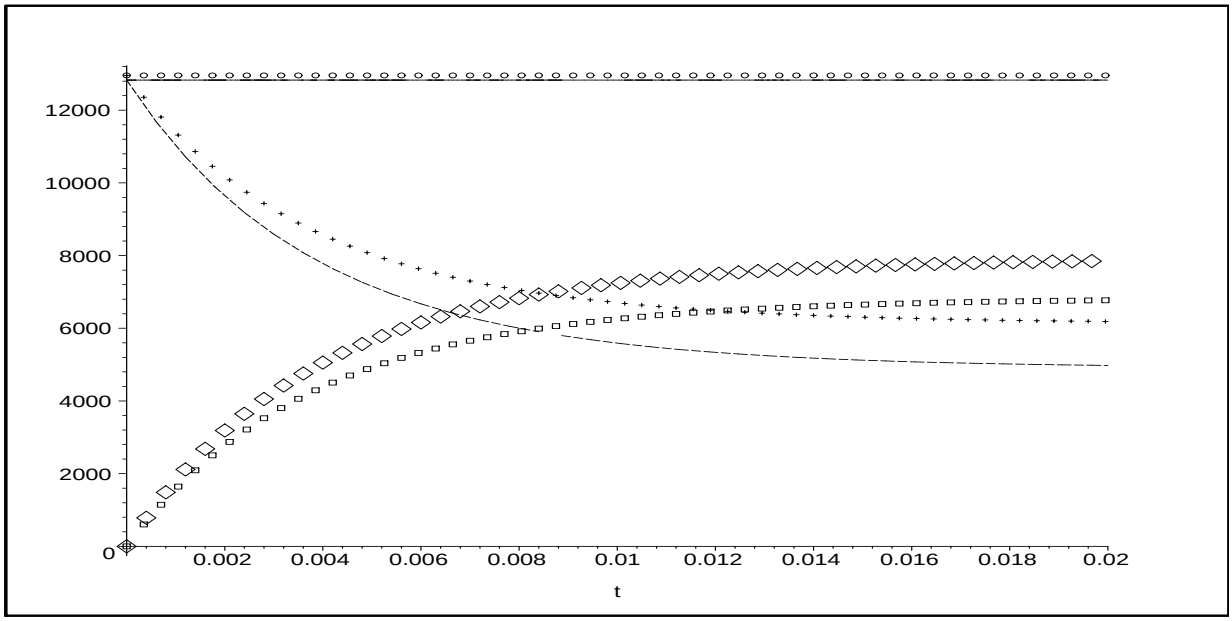


Figure 19: Example 12. RCD circuit: Exact storage function (dashed line), exact cumulative dissipation function (diamond), exact storage function + cumulative dissipation function (solid line), $(\frac{1}{2}, \frac{1}{2})$-method approximation of storage function (cross), $(\frac{1}{2}, \frac{1}{2})$-method approximation of cumulative dissipation function (box), $(\frac{1}{2}, \frac{1}{2})$-method approximation of storage function + cumulative dissipation function (circle). Time step $h = 1 \times 10^{-6}$.

This system has a unilateral constraint $x_3(t) \geqslant 0$ and $\tilde{D} = \frac{1}{R}$. Figure 18 depicts a trajectory starting at $x(0^-) = (5, 20, -100)^T$ with $h = 1 \times 10^{-6}$, and for $\theta = \gamma = \frac{1}{2}$ and $\theta = \frac{1}{2}$, $\gamma = 1$. The solution has an initial jump to since $x_3(0^-) < 0$, and the jump varies with γ.

Remark 9 *When D has the special structure of section 4.2, then the algorithm (133) is more general than (152), because it leaves some freedom for the choice of γ in one part of the variables. This is why it is worth studying the special structure despite the results of section 4.3 deal with the more general case $D \geqslant 0$.*

4.3 The case when $D \geqslant 0$

Considering the most general case $D \geqslant 0$, we propose to use the following $(\theta, 1)$-scheme:

$$\begin{cases} x_{k+1} - x_k = h\left(Ax_{k+\theta} + Eu_{k+\theta}\right) + B\sigma_{k+1} \\ \\ w_{k+1} = Cx_{k+1} + Fv_{k+1} + \dfrac{D}{h}\sigma_{k+1} \\ \\ 0 \leqslant w_{k+1} \perp \sigma_{k+1} \geqslant 0. \end{cases} \tag{152}$$

Let us denote by $\mathbf{K}(D)$ is the set of vectors q such that LCP(q, D) has a solution, that is:

$$\mathbf{K}(D) = \{q \in I\!R^m \mid \mathrm{SOL}(q, D) \neq \emptyset\}. \tag{153}$$

From Lemma 3 in the Appendix one has:

$$Q_D = \{z \in I\!R^m \mid z \geqslant 0, D^T z \leqslant 0\} = [\mathbf{K}(D)]^*, \tag{154}$$

and

$$\mathbf{K}(D) = Q_D^* = I\!R_+^m - DI\!R_+^m. \tag{155}$$

Proposition 11 *Let us assume that $\lim_{h\to 0} x_{k+1} < +\infty$ and $\lim_{h\to 0} \sigma_{k+1} = \sigma_\infty \neq 0$ with $\sigma_\infty < +\infty$. Then σ_∞ solves the following LCCP:*

$$Q_D \ni \sigma_\infty \perp Fv_{k+1} + Cx_k + CB\sigma_\infty \in Q_D^*, \tag{156}$$

for $x_k = x(t^-)$ and $v_{k+1} = v(t^+)$, which is equivalent to the jump law in Definition 2. Furthermore, we obtain that if $K \neq \emptyset$:

$$\lim_{h\to 0} \|x_{k+1} - x(t^+)\| = 0, \tag{157}$$

where $x(t^+)$ is given by the state jump rule (101) and any $x_k = x(t^-)$.

Proof: From (152) let us consider the associated LCP one–step problem:

$$\begin{cases} hw_{k+1} = h(Cx_{k+1} + Fv_{k+1}) + D\sigma_{k+1} \\ \\ 0 \leqslant hw_{k+1} \perp \sigma_{k+1} \geqslant 0. \end{cases} \tag{158}$$

If we assume that $\lim_{h\to 0} \sigma_{k+1} = \sigma_\infty < +\infty$ and $\lim_{h\to 0} x_{k+1} < +\infty$, we have that:

$$\lim_{h\to 0} hw_{k+1} = D\sigma_\infty, \tag{159}$$

and σ_∞ satisfies:

$$0 \leqslant D\sigma_\infty \perp \sigma_\infty \geqslant 0. \tag{160}$$

This implies that $\sigma_\infty \in Q_D$. From (155), we note that

$$w_{k+1} - \frac{D}{h}\sigma_{k+1} \in \mathbf{K}(D) = Q_D^*. \tag{161}$$

From (152) one has:

$$w_{k+1} - \frac{D}{h}\sigma_{k+1} = Fv_{k+1} + C(I - h\theta A)^{-1}\left[(I + h(1-\theta)A)x_k + hEu_{k+\theta}\right] + C(I - h\theta A)^{-1}B\sigma_{k+1}, \tag{162}$$

Let us denote the following limit by

$$w_\infty \triangleq \lim_{h\to 0} w_{k+1} - \frac{D}{h}\sigma_{k+1} \tag{163}$$

So it follows that:

$$w_\infty = Fv_{k+1} + Cx_k + CB\sigma_\infty \in Q_D^*, \tag{164}$$

since Q_D^* is a closed set and we assume $\lim_{h\to 0}\sigma_{k+1} = \sigma_\infty$. It remains to prove that $w_\infty \perp \sigma_\infty$. Since $\lim_{h\to 0}\sigma_{k+1} = \sigma_\infty < +\infty$, hence $w_\infty < +\infty$, we can write:

$$\begin{array}{rcll} \sigma_\infty^T w_\infty & = & \lim_{h\to 0}\sigma_{k+1}^T\left(w_{k+1} - \dfrac{D}{h}\sigma_{k+1}\right) & \\ & = & \lim_{h\to 0} -\dfrac{1}{h}\sigma_{k+1}^T D\sigma_{k+1} & \text{due to (152)} \\ & \leqslant & 0, & \end{array} \tag{165}$$

due to the positive semi–definiteness of D. Since $\sigma_\infty^T \in Q_D$ and $w_\infty \in Q_D^*$, we also have $\sigma_\infty^T w_\infty \geqslant 0$ and therefore we conclude that $\sigma_\infty^T w_\infty = 0$. To summarize, w_∞ and σ_∞ solve the following LCCP:

$$\begin{cases} w_\infty = Fv_{k+1} + Cx_k + CB\sigma_\infty \\ Q_D^* \ni w_\infty \perp \sigma_\infty \in Q_D, \end{cases} \tag{166}$$

or equivalenlty the jump law (223) hence (222). From (223) and (152), we get:

$$\|x_{k+1} - x(t^+)\| = \|h\left(Ax_{k+\theta} + Eu_{k+\theta}\right) + B(\sigma_{k+1} - \sigma)\| \tag{167}$$

since $x_k = x(t^-)$. Hence:

$$\lim_{h\to 0}\|x_{k+1} - x(t^+)\| = \lim_{h\to 0}\|B(\sigma_{k+1} - \sigma)\| \tag{168}$$

for a bounded sequence of $x_{k+\theta}$ and $u_{k+\theta}$. If $K \neq \emptyset$, $x(t^+)$ is uniquely defined by the jump law (222) and therefore $B\sigma$ is also uniquely defined. We can therefore conclude that $\lim_{h\to 0}\|B(\sigma_{k+1} - \sigma)\| = 0$ which ends the proof. ■

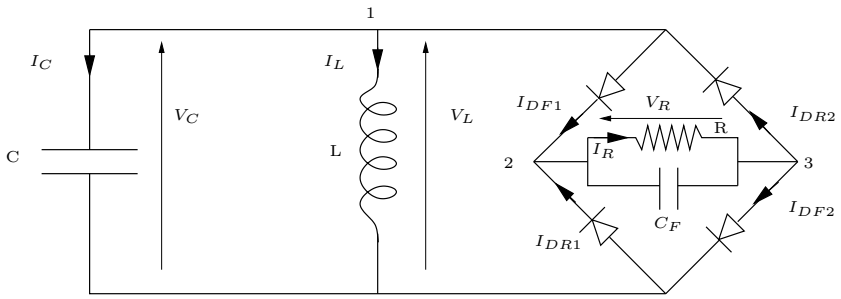


Figure 20: LC oscillator with a load resistor filtered by a capacitor

Example 13 *[Diode bridge cap filter] Let us consider the circuit in figure 20. Its dynamics is given by the following data:*

$$A = \begin{pmatrix} 0 & -\frac{1}{c} & 0 \\ \frac{1}{l} & 0 & 0 \\ 0 & 0 & -\frac{1}{rc_f} \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & -\frac{1}{c} & \frac{1}{c} \\ 0 & 0 & 0 & 0 \\ \frac{1}{c_f} & 0 & \frac{1}{c_f} & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \quad D = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & -1 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \tag{169}$$

with

$$L=\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&\sqrt{\frac{2}{r}}\end{pmatrix}\quad \textit{and}\quad W=0 \tag{170}$$

The parameter values and initial conditions of the system are taken as, $x_1(0)=10.0$, $x_2(0)=0$, $r=10^3$, $c=10^{-6}$, $c_f=300.10^{-9}$. *The discrete system (for* $\theta=\gamma=\frac{1}{2}$*) has:*

$$\tilde{L}=\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&\frac{2rc_f}{2rc_f+h}\sqrt{\frac{2}{r}}\end{pmatrix}\quad \textit{and}\quad \tilde{W}=\frac{\sqrt{2r}h}{(2rc_f+h)}\begin{pmatrix}1&0&1&0\\0&0&0&0\\1&0&1&0\\0&0&0&0\end{pmatrix}. \tag{171}$$

The non-zero eigenvalues of $Q-\tilde{Q}$ *are:*

$$\frac{-2hr^2+4rc_f+h\pm\sqrt{(1+2r^2)(2h^2r^2+16r^2c_f^2+8rc_fh+h^2)}h}{r(4r^2c_f^2+4rc_fh+h^2)}. \tag{172}$$

Let us start with some computation for this specific example:

$$Q_D=\{z\in I\!R^m \mid z\geqslant 0, Dz\geqslant 0, z^TDz=0\}. \tag{173}$$

Since $D+D^T=0$ *and* $z^TDz=\frac{1}{2}z^T(D+D^T)z$, *the condition* $z^TDz=0$ *holds for any* $z\in I\!R^m$. *The computation of* Q_D *yields*

$$Q_D=\{z\in I\!R^4 \mid z\geqslant 0, z_2=0, z_1+z_3-z_4\geqslant 0\}, \tag{174}$$

and the cone K *given by*

$$K=\{x\in I\!R^3 \mid Cx\in Q_D\}, \tag{175}$$

is given in our example by

$$K=\{x\in I\!R^3 \mid x_1\geqslant 0, x_3\geqslant 0, 2(x_3-x_1)\geqslant 0\}, \tag{176}$$

that is

$$K=\{x\in I\!R^3 \mid Cx\geqslant 0\}. \tag{177}$$

We can check that

$$P=\begin{pmatrix}c&0&0\\0&L&0\\0&0&c_F\end{pmatrix} \tag{178}$$

solves the LMI in (2). The jump law is given by:

$$-P(x(t^+)-x(t^-))\in N_K(x(t^+)). \tag{179}$$

Since $K=\{x\in I\!R^3 \mid Cx\geqslant 0\}$, *we get:*

$$\begin{cases}P(x(t^+)-x(t^-))=C^T\sigma,\\ w=Cx(t^+),\\ 0\leqslant w\perp\sigma\geqslant 0.\end{cases} \tag{180}$$

Since $PB=C^T$ *and* $P>0$, *we get:*

$$\begin{cases}(x(t^+)-x(t^-))=B\sigma\\ w=Cx(t^+)\\ 0\leqslant w\perp\sigma\geqslant 0,\end{cases} \tag{181}$$

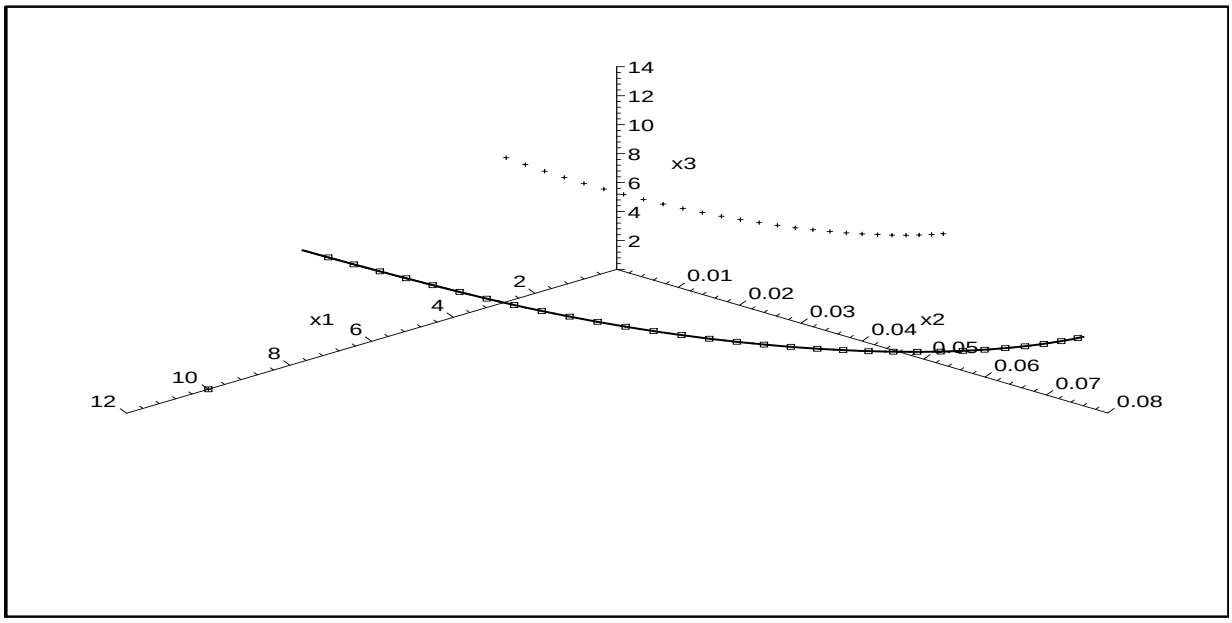


Figure 21: Example 13: Capfilter phase portrait: Exact Solution (solid line), $\theta = \gamma = \frac{1}{2}$ (cross) and $\theta = \frac{1}{2}, \gamma = 1$ (box). Initial state (10,0,0), time step $h = 1.0 \times 10^{-6}$.

and conclude that the jump law amounts to solving

$$\begin{cases} w = Cx(t^-) + CB\sigma \\ 0 \leqslant w \perp \sigma \geqslant 0, \end{cases} \tag{182}$$

which is exactly what is solved by the time-stepping scheme at the first step for $h \to 0$. *If the* $(\theta, \gamma)-$*scheme is used, we solve:*

$$\begin{cases} w_{k+\gamma} = M\lambda_{k+\gamma} + q \\ \\ K^* \in w_{k+\gamma} \perp \lambda_{k+\gamma} \in K, \end{cases} \tag{183}$$

with

$$M = D + h\gamma C(I_n - h\theta A)^{-1}B, q = a_{k+\gamma} + \gamma C(I_n - h\theta A)^{-1}\left[(I_n + h(1-\theta)A)x_k + hu_{k+\theta}\right] + C(1-\gamma)x_k. \tag{184}$$

that is for $h \to 0$ *and* $\sigma_{k+\gamma} = h\lambda_{k+\gamma}$

$$\begin{cases} w_{k+\gamma} = \gamma CB\sigma_{k+\gamma} + a_{k+\gamma} + Cx_k \\ \\ K^* \in w_{k+\gamma} \perp \sigma_{k+\gamma} \in K, \end{cases} \tag{185}$$

We can see that the matrix of the LCP is multiplied by γ. *Since* D *is not full rank and* $Cx(0^-) \leqslant 0$ *the system initially undergoes a state jump. One can see in Figure 21 that the* $(\frac{1}{2}, \frac{1}{2})$*-method fails to estimate the jump properly, whereas the* $(\frac{1}{2}, 1)$*-method jumps to the correct state. Unsurpisingly, due to the incorrect jump approximation, the storage and dissipation functions (Figures 21 and 22) fail to be approximated by the* $(\frac{1}{2}, \frac{1}{2})$*-method.*

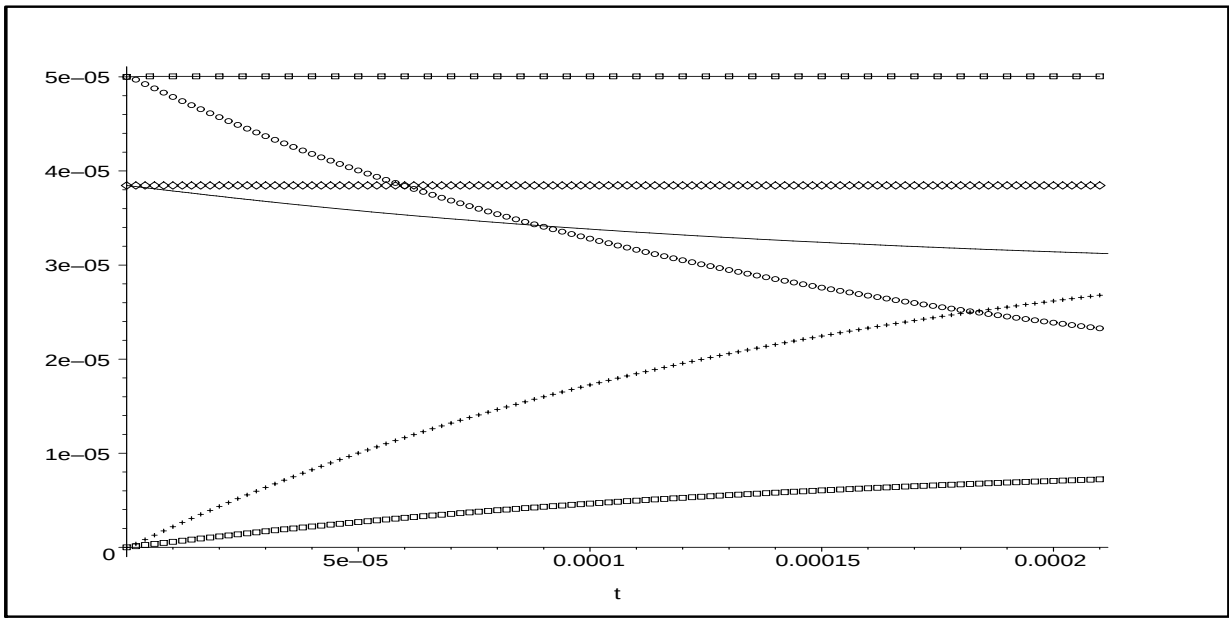


Figure 22: Example 13. Capfilter : Exact storage function (solid line), exact cumulative dissipation function (box), exact storage function + cumulative dissipation function (diamond), $(\frac{1}{2}, \frac{1}{2})$-method approximation of storage function (circle), $(\frac{1}{2}, \frac{1}{2})$-method approximation of cumulative dissipation function (cross), $(\frac{1}{2}, \frac{1}{2})$-method approximation of storage function + cumulative dissipation function (box-line). Time step $h = 1.0 \times 10^{-6}$.

4.4 Conclusions

It follows from the previous sections that the $(\theta, \gamma)-$method is consistently approximating the state jumps only if $\gamma = 1$, *i.e.* for schemes that are fully implicit in the multiplier.

5 Conclusions and perspectives

The main results of this paper and some perspectives may are summarized now:

- It offers a systematic framework to study the dissipativity properties preservation after the discretization with the $(\theta, \gamma)-$method;
- It provides a rigorous definition of the numerical dissipation;
- It explains why state lossless continuous-time systems are more easily transformed into state lossless discrete-time systems, than state dissipative systems;
- It examines the consistency of state jumps approximations, and shows that only fully implicit (in the multiplier) methods yield consistency;
- It presents several examples of circuits containing multivalued nonsmooth components (ideal diodes, Zener diodes) to illustrate the developments;
- The framework may be extended to other numerical schemes like the zero order hold method that is used for feedback control purposes; it may also be used to study if other methods like multistep methods (applied on the state only) may improve the dissipativity preservation.

- From a practical point of view it may be recommended to choose $\theta = \gamma = \frac{1}{2}$ for systems with no state jumps, and $\theta = \frac{1}{2}$, $\gamma = 1$ for systems with state jumps.

A Some results on LCPs

Lemma 3 *Let us assume that $D \in I\!R^{m\times m}$ is a semi-definite positive matrix. Let us define $Q_D = \mathrm{SOL}(D,0) = \{z, z \geqslant 0, Dz \geqslant 0, z^T Dz = 0\}$ and $\mathbf{K}(D) = \{q \mid \mathrm{SOL}(D,q) \neq \emptyset\}$. Then, we have*

a) $z^T Dz = 0 \Longleftrightarrow (D + D^T)z = 0$

b) $Q_D = \{z \in I\!R^m \mid D^T z \leqslant 0, z \geqslant 0\}$

c) $\mathbf{K}(D) = Q_D^* = I\!R_+^m - DI\!R_+^m$

d) $[\mathbf{K}(D)]^* = Q_D$

Proof:

a) Let us consider the following convex quadratic programing problem

$$\min \frac{1}{2} z^T Dz \tag{186}$$

which is equivalent to:

$$\min z^T (D + D^T)z. \tag{187}$$

Since $z^T(D+D^T)z \geqslant 0$ and the bound is reached for $z = 0$, the solution of (187) is then equivalent to $z^T Dz = 0$. Since the problem is convex, the KKT conditions are:

$$(D + D^T)z = 0, \tag{188}$$

and they are equivalent to (187). Finally, we conclude that:

$$z^T Dz = 0 \Longleftrightarrow \min \frac{1}{2} z^T Dz \Longleftrightarrow (D + D^T)z = 0. \tag{189}$$

b) If $v \in \{z \mid D^T z \leqslant 0, z \geqslant 0\}$, we have:

$$z^T D^T z \leqslant 0, \tag{190}$$

which implies:

$$z^T D^T z = 0 \tag{191}$$

since $D^T \geqslant 0$. Hence, $Dz = -D^T z \geqslant 0$ and $v \in Q_D$. Conversely, if $v \in Q_D$, we have $Dz = -D^T z \geqslant 0$.

c) If $q \in Q_D^*$, then $\mathrm{LCP}(D,q)$ is solvable [17, Theorem 3.8.6]. Hence, $Q_D^* \subset \mathbf{K}(D)$. If $q \in \mathbf{K}(D)$, $\exists x, w \in I\!R^m$ such that

$$\begin{cases} w = Dx + q \\ 0 \leqslant x \perp w \geqslant 0. \end{cases} \tag{192}$$

Since $Q_D = \{z \in I\!R^m \mid D^T z \leqslant 0, z \geqslant 0\}$, the dual cone Q_D^* can be expressed as [43, p 122]

$$Q_D^* = \{v \in I\!R^m \mid v = [I - D]\alpha, \alpha \geqslant 0\}. \tag{193}$$

that is

$$Q_D^* = I\!R_+^m - DI\!R_+^m. \tag{194}$$

From (192), we get

$$q = w - Dx, w \geqslant 0, x \geqslant 0. \tag{195}$$

Hence, $q \in Q_D^*$ if we choose $\alpha^T = [w^T x^T]$.

d) Since $D \geqslant 0$, the set of solution of $\mathrm{LCP}(D,0)$ is a closed convex cone; therefore $[Q_D^*]^* = Q_D = [\mathbf{K}(D)]^*$ [43, Theorem 14.1] . ■

B Proof of Lemma 1

(i) In the derivation of equation (27) we make use of the fact that $(I-\mu A)(I+\eta A) = (I+\eta A)(I-\mu A)$ for any reals μ and η. Recalling equation (15) we have,

$$\tilde{A}^T R\tilde{A} - R = -\tilde{L}\tilde{L}^T. \tag{196}$$

Using the definition that $\tilde{A} = (I-h\theta A)^{-1}(I+h(1-\theta)A)$ yields:

$$(I+h(1-\theta)A)^T(I-h\theta A)^{-T}R(I-h\theta A)^{-1}(I+h(1-\theta)A) - R = -\tilde{L}\tilde{L}^T. \tag{197}$$

Multiplying (on the left) by $(I-h\theta A)^T$ and (on the right) by $(I-h\theta A)$ yields:

$$\begin{array}{l}(I-h\theta A)^T\big((I+h(1-\theta)A)^T(I-h\theta A)^{-T}R(I-h\theta A)^{-1}(I+h(1-\theta)A) - R\big)(I-h\theta A)\\ = -(I-h\theta A)^T\tilde{L}\tilde{L}^T(I-h\theta A)\end{array} \tag{198}$$

and thus:

$$\begin{array}{l}(I-h\theta A)^T(I+h(1-\theta)A)^T(I-h\theta A)^{-T}R(I-h\theta A)^{-1}(I+h(1-\theta)A)(I-h\theta A)\\ -(I-h\theta A)^TR(I-h\theta A) = -(I-h\theta A)^T\tilde{L}\tilde{L}^T(I-h\theta A)\end{array} \tag{199}$$

Using the commutativity of $(I+h(1-\theta)A)(I-h\theta A)$ (and likewise the commutativity of its transpose), we obtain:

$$\begin{array}{l}(I+h(1-\theta)A)^T(I-h\theta A)^T(I-h\theta A)^{-T}R(I-h\theta A)^{-1}(I-h\theta A)(I+h(1-\theta)A) - (I-h\theta A)^TR(I-h\theta A)\\ = -(I-h\theta A)^T\tilde{L}\tilde{L}^T(I-h\theta A).\end{array} \tag{200}$$

Simplifying:

$$(I+h(1-\theta)A)^TR(I+h(1-\theta)A) - (I-h\theta A)^TR(I-h\theta A) = -(I-h\theta A)^T\tilde{L}\tilde{L}^T(I-h\theta A) \tag{201}$$

Expanding yields:

$$\begin{array}{l}\big(R+h(1-\theta)(A^TR+RA)+h^2(1-\theta)^2A^TRA\big) - \big(R-h\theta(A^TR+RA)+h^2\theta^2A^TRA\big)\\ = -(I-h\theta A)^T\tilde{L}\tilde{L}^T(I-h\theta A).\end{array} \tag{202}$$

Collecting terms by powers of h yields (27).

(ii) Recalling equation (16) we have,

$$\tilde{B}^TR\tilde{A} - \tilde{C} = -\tilde{W}^T\tilde{L}^T \tag{203}$$

Using the definitions of $(\tilde{A},\tilde{B},\tilde{C},\tilde{D})$ from (12) yields:

$$\begin{array}{l}hB^T(I-h\theta A)^{-T}R(I-h\theta A)^{-1}(I+h(1-\theta)A) - \gamma C(I-h\theta A)^{-1}(I+h(1-\theta)A) - (1-\gamma)C\\ = -\tilde{W}^T\tilde{L}^T\end{array} \tag{204}$$

Using $h(A^TR+RA) = -LL^T$, we have that $(I-h\theta A)^TR = R(I+h\theta A)+\theta LL^T$ and thus $(I-h\theta A)^{-T}R = R(I+h\theta A)^{-1} - \theta(I-h\theta A)^{-T}LL^T(I+h\theta A)^{-1}$, so that we obtain:

$$\begin{array}{l}B^T[hR(I+h\theta A)^{-1} - \theta(I-h\theta A)^{-T}LL^T(I+h\theta A)^{-1}](I-h\theta A)^{-1}(I+h(1-\theta)A)\\ -\gamma C(I-h\theta A)^{-1}(I+h(1-\theta)A) - (1-\gamma)C = -\tilde{W}^T\tilde{L}^T.\end{array} \tag{205}$$

For the next step we multiply (on the right) by $(I-h\theta A)$ and $(I+h\theta A)$ and use the commutativity feature of matrices of the form $(I+\mu A)(I-\eta A)$ in order to cancel out matrices $(I+h\theta A)^{-1}$ and $(I-h\theta A)^{-1}$, and thus obtain:

$$\begin{array}{l}B^T[hR-\theta(I-h\theta A)^{-T}LL^T](I+h(1-\theta)A) - [\gamma C(I+h(1-\theta)A)+(1-\gamma)C(I-h\theta A)](I+h\theta A)\\ = -\tilde{W}^T\tilde{L}^T(I-h\theta A)(I+h\theta A).\end{array} \tag{206}$$

Simplifying the part of the equality involving C yields:

$$\begin{aligned} &B^T\big(hR - \theta(I-h\theta A)^{-T}LL^T\big)(I+h(1-\theta)A) - C(I+h(\gamma-\theta)A)(I+h\theta A) \\ &= -\tilde{W}^T\tilde{L}^T(I-h\theta A)(I+h\theta A). \end{aligned} \tag{207}$$

Imposing the condition (4), that is $C = hB^TR + W^TL^T$, we obtain:

$$\begin{aligned} &B^T\big(hR - \theta(I-h\theta A)^{-T}LL^T\big)(I+h(1-\theta)A) - (hB^TR + W^TL^T)(I+h(\gamma-\theta)A)(I+h\theta A) \\ &= -\tilde{W}^T\tilde{L}^T(I-h\theta A)(I+h\theta A). \end{aligned} \tag{208}$$

Moving the part of the equality involving W^TL^T on the left-hand side to the right-hand side yields:

$$\begin{aligned} &B^T\big(hR - \theta(I-h\theta A)^{-T}LL^T\big)(I+h(1-\theta)A) - hB^TR(I+h(\gamma-\theta)A)(I+h\theta A) \\ &= W^TL^T(I+h(\gamma-\theta)A)(I+h\theta A) - \tilde{W}^T\tilde{L}^T(I-h\theta A)(I+h\theta A). \end{aligned} \tag{209}$$

Collecting terms involving hB^TR yields:

$$\begin{aligned} &hB^TR\big((I+h(1-\theta)A) - (I+h(\gamma-\theta)A)(I+h\theta A)\big) - \theta B^T(I-h\theta A)^{-T}LL^T(I+h(1-\theta)A) \\ &= W^TL^T(I+h(\gamma-\theta)A)(I+h\theta A) - \tilde{W}^T\tilde{L}^T(I-h\theta A)(I+h\theta A). \end{aligned} \tag{210}$$

Simplifying yields:

$$\begin{aligned} &hB^TR\big(h(1-\theta-\gamma)A - h^2\theta(\gamma-\theta)A^2\big) - \theta B^T(I-h\theta A)^{-T}LL^T(I+h(1-\theta)A) \\ &= W^TL^T(I+h(\gamma-\theta)A)(I+h\theta A) - \tilde{W}^T\tilde{L}^T(I-h\theta A)(I+h\theta A). \end{aligned} \tag{211}$$

Noting that $(I-h\theta A)^{-T}LL^T(I+h(1-\theta)A) = (I-h\theta A)^{-T}LL^T(I-h\theta A)^{-1}(I-h\theta A)(I+h(1-\theta)A) = \tilde{L}\tilde{L}^T(I-h\theta A)(I+h(1-\theta)A)$ (by ansatz (31)) we finally obtain (28).
(iii) For (17) we once again use $(\tilde{A},\tilde{B},\tilde{C},\tilde{D})$ defined by (12) to get:

$$h^2B^T(I-h\theta A)^{-T}R(I-h\theta A)^{-1}B - h\gamma C(I-h\theta A)^{-1}B - h\gamma B^T(I-h\theta A)^{-T}C^T - D^T - D = -\tilde{W}^T\tilde{W}. \tag{212}$$

Using the continuous conditions $D^T + D = W^TW$ and $C = hB^TR + W^TL^T$ we obtain:

$$\begin{aligned} &h^2B^T(I-h\theta A)^{-T}R(I-h\theta A)^{-1}B - h\gamma(hB^TR + W^TL^T)(I-h\theta A)^{-1}B \\ &-h\gamma B^T(I-h\theta A)^{-T}(hB^TR + W^TL^T)^T - W^TW = -\tilde{W}^T\tilde{W}. \end{aligned} \tag{213}$$

Rearranging so that terms involving W and L are on the right-hand side yields:

$$\begin{aligned} &h^2B^T(I-h\theta A)^{-T}R(I-h\theta A)^{-1}B - h^2\gamma B^TR(I-h\theta A)^{-1}B - h^2\gamma B^T(I-h\theta A)^{-T}RB \\ &= W^TW - \tilde{W}^T\tilde{W} + h\gamma W^TL^T(I-h\theta A)^{-1}B + h\gamma B^T(I-h\theta A)^{-T}LW. \end{aligned} \tag{214}$$

Factoring the left-hand side by $B^T(I-h\theta A)^{-T}$ (from the left) and $(I-h\theta A)^{-1}B$ (from the right) yields:

$$\begin{aligned} &h^2B^T(I-h\theta A)^{-T}(R - \gamma R(I-h\theta A) - \gamma(I-h\theta A)^TR)(I-h\theta A)^{-1}B \\ &= W^TW - \tilde{W}^T\tilde{W} + h\gamma W^TL^T(I-h\theta A)^{-1}B + h\gamma B^T(I-h\theta A)^{-T}LW. \end{aligned} \tag{215}$$

Collecting the left-hand side by powers of h:

$$\begin{aligned} &h^2B^T(I-h\theta A)^{-T}((1-2\gamma)R + h\theta\gamma(A^TR + RA))(I-h\theta A)^{-1}B \\ &= W^TW - \tilde{W}^T\tilde{W} + h\gamma W^TL^T(I-h\theta A)^{-1}B + h\gamma B^T(I-h\theta A)^{-T}LW. \end{aligned} \tag{216}$$

Finally using $h(A^TR + RA) = -LL^T$ yields (29).

C Equivalent formulations of the state jump law

Proposition 12 *Under the conditions of Definition 2, the following holds:*

$$P(x(t^+) - x(t^-)) \in -N_K(x(t^+)). \tag{217}$$

$$\Updownarrow$$

$$P(x(t^+) - x(t^-))(x(t^+) - y) \geqslant 0, \quad \textit{for all } y \in K. \tag{218}$$

$$\Updownarrow$$

$$x(t^+) = \text{argmin}_{x\in K}\, \frac{1}{2}(x - x(t^-))^T P(x - x(t^-)), \tag{219}$$

$$\Updownarrow$$

$$K \ni x(t^+) \perp P(x(t^+) - x(t^-)) \in K^* \tag{220}$$

$$\Updownarrow$$

$$x(t^+) = proj_P[K; x(t^-)] \tag{221}$$

$$\Updownarrow$$

$$\begin{cases} P(x(t^+) - x(t^-)) = C^T\sigma \\ w = Cx(t^+) + Fv(t^+) \\ Q_D^* \ni w \perp \sigma \in Q_D. \end{cases} \tag{222}$$

$$\Updownarrow$$

$$\begin{cases} (x(t^+) - x(t^-)) = B\sigma \\ w = Cx(t^+) + Fv(t^+) \\ Q_D^* \ni w \perp \sigma \in Q_D. \end{cases} \tag{223}$$

$$\Updownarrow$$

$$\begin{cases} w = Cx(t^-) + Fv(t^+) + CB\sigma \\ Q_D^* \ni w \perp \sigma \in Q_D. \end{cases} \tag{224}$$

Proof: The equivalence between (217) and (218) follows from the definition of a normal cone to a convex set [44, Definition 5.2.3]. The equivalences between (218), (219) and (220) can be shown using the material in [50, Chapter 1]. The equivalence between (220) and (217) is direct from convex analysis: for any convex non empty closed cone $K \subset I\!R^n$ and any two vectors x and y in $I\!R^n$, $K \ni x \perp y \in K^* \;\Leftrightarrow\; y \in -N_K(x)$. Notice that (221) is just a rewriting of (219). The equivalence between (222) and (217) can be shown as follows: the complementarity conditions in (222) are equivalent to $\sigma \in -N_{Q_D^*}(w)$ that is equivalent (since $P > 0$) $P(x(t^+) - x(t^-)) \in -C^T N_{Q_D^*}(Cx(t^+) + Fv(t^+)) = -N_K(x(t^+))$, where the last equality follows from the chain rule of convex analysis [44, Theorem 4.2.1] and the definitions of K and Q_D. The equivalence between (222) and (223) is true since $\sigma \in Q_D$ implies that $\sigma^T(D + D^T)\sigma = 0$ and then $C^T\sigma = PB\sigma$ (see *e.g.* [28, Lemma 2.b]). Finally (224) is just a rewriting of (223). ■

References

[1] M. Gonzalez, B. Schmidt, M. Ortiz, "Energy-stepping integrators in Lagrangian mechanics", Int. Journal for Numerical Methods in Engineering, vol.82, pp.205-241, 2010.

[2] T.A. Laursen, V. Chawla, "Design of energy conserving algorithms for frictionless dynamic contact problems", Int. Journal for Numerical Methods in Engineering, vol.40, pp.863-886, 1997.

[3] T. Eirola, J.M. Sanz-Serna, "Conservation of integrals and symplectic structure in the integration of differential equations by multistep methods", Numer. Math., vol.61, pp.281-290, 1992.

[4] E. Hairer, C. Lubich, "Symmetric multistep methods over long times", Numer. Math., vol.97, pp.699-723, 2004.

[5] E. Celledoni, R.I. McLachlan, D.I. McLaren, B. Owren, G.R.W. Quispel, W.M. Wright, "Energy-preserving Runge-Kutta methods", ESAIM: Mathematical Modelling and Numerical Analysis, vol.43, pp.645-649, 2009.

[6] V. Acary, B. Brogliato, D. Goeleven, "Higher order Moreau's sweeping process: Mathematical formulation and numerical simulation", Mathematical Programming A, vol.113, pp.133-217, 2008.

[7] A. Bemporad, G. Bianchini, F. Brogi, "Passivity analysis and passification of discrete-time hybrid systems", IEEE Transactions on Automatic Control, vol.53, no 4, pp.1004-1009, May 2008.

[8] B. Brogliato, D. Goeleven, "Well-posedness, stability and invariance results for a class of multivalued Lur'e dynamical systems", Nonlinear Analysis: Theory, Methods and Applications, vol.74, pp.195-212, 2011.

[9] B. Brogliato, L. Thibault, "Well-posedness results for non-autonomous complementarity systems", Journal of Convex Analysis, vol.17, no 3-4, pp.961-990, 2010.

[10] B. Brogliato, "Absolute stability and the Lagrange-Dirichlet theorem with monotone multivalued mappings", Systems and Control Letters, vol.51, no 5, pp.343-353, April.

[11] M.K. Camlibel, L. Iannelli, F. Vasca, "Passivity and complementarity", GRACE Internal Report available at www.grace.ing.unisannio.it, no 352, 2006.

[12] M.K. Camlibel, W.P.M.H. Heemels, J.M. Schumacher, "Consistency of a time-stepping method for a class of piecewise-linear networks", IEEE Transactions on Circuits and Systems–I: Fundamental Theory and Applications, vol.49, no 3, pp.349-357, March 2002.

[13] M.K. Camlibel, W.P.M.H. Heemels, J.M. Schumacher, "On linear passive complementarity systems", European Journal of Control, vol.8, no 3, pp.220-237, 2002.

[14] M.K. Camlibel, W.P.M.H. Heemels, van der Schaft, A.J, J.M. Schumacher, "Switched networks and complementarity",IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications , vol.50, no 8, pp.1036–1046, 2003.

[15] M.K. Camlibel, J.S. Pang, J. Shen, "Lyapunov stability of complementarity and extended systems", SIAM J. Optimization, vol.17, no 4, pp.1056-1101, 2006.

[16] R. Costa Castello, E. Fossas, "On preserving passivity in sampled-data linear systems", European Journal of Control, vol.13, no 6, pp.583-590, 2007.

[17] R. W. Cottle, J. Pang, and R. E. Stone. *The Linear Complementarity Problem.* Academic Press, Inc., Boston, MA, 1992.

[18] R. Frasca, M.K. Camlibel, I.C. Goknar, L. Iannelli, and F. Vasca. Linear passive networks with ideal switches: Consistent initial conditions and state discontinuities. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 57(12):3138 –3151, 2010.

[19] R. Frasca, M.K. Camlibel, I.C. Goknar, and F. Vasca. State jump rules in linear passive networks with ideal switches. In *IEEE Int. Symp. on Circuits and Systems (ISCAS)*, Seattle, Washington, May 2008.

[20] J. Jiang, "Preservations of positive realness under discretizations", Journal of the Franklin Institute, vol.330, no 4, pp.721-734, 1993.

[21] L. Iannelli, F. Vasca, and K. Camlibel. Complementarity and passivity for piecewise linear feedback systems. In *Decision and Control, 2006 45th IEEE Conference on*, pages 4212 –4217, 2006.

[22] L. Iannelli, F. Vasca, G. Angelone, "Computation of steady-state oscillations in power converters through complementarity", IEEE Transactions on Circuits and Systems–I Regular Papers, in press, 2011.

[23] P. Lancaster, M. Tismenetsky, *The Theory of Matrices*, 2nd Ed., 1985, Academic Press.

[24] P. Faurre, M. Clerget, F. Germain, *Opérateurs Rationnels Positifs*, Dunod, 1979.

[25] P. Faurre, *Réalisations Markoviennes de Processus Stationnaires*, PhD Thesis, university Paris VI, 1972.

[26] M. Bruschetta, G. Picci, A. Saccon , "How to Sample Linear Mechanical Systems", in Persp. in Math. Sys. Theory, Ctrl. and Sign. Pro., J.C. Willems et al (Eds.), LNCIS 398, pp.343-353, Springer-Verlag Berlin 2010.

[27] B. Brogliato, R. Lozano, B. Maschke, O. Egeland, *Dissipative Systems Analysis and Control. Theory and Applications*, 2^{nd} Edition, Springer-Verlag Berlin 2007.

[28] L. Han, A. Tiwari, M.K. Camlibel, J.S. Pang, "Convergence of time-stepping schemes for passive and extended linear complementarity systems", SIAM J. Numer. Anal., vol.47, no 5, pp.3768-3796, 2009.

[29] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems.* Springer, 1996.

[30] J.B. Hoagg, S.L. Lacy, R.S. Erwin, D.S. Bernstein, "First-order-hold sampling of positive real systems and subspace identification of positive real models", Proc. of the American Control Conference, Boston, Massachussetts, June 30-July 2, pp.861-866, 2004.

[31] Y. Oishi, "Passivity Degradation under the Discretization with the Zero-Order Hold and the Ideal Sampler", Proc. of the 49^{th} IEEE Conference on Decision and Control, Atlanta, pp.7613-7617, December 15-17, 2010.

[32] D. Nesic, D. S. Laila, A. R. Teel, "On preservation of dissipation inequalities under sampling", Proc. of the 39^{th} IEEE Conference on Decision and Control, Sydney, Australla, pp.2472-2477, December 2000.

[33] D. S. Laila, D. Nesic, A. R. Teel, "Open and closed loop dissipation inequalities under sampling and controller emulation", European Journal of Control, vol.8, no 2, pp.109-125, 2002.

[34] M.D.P. Monteiro Marques. *Differential Inclusions in Nonsmooth Mechanical Problems. Shocks and Dry Friction.* Progress in Nonlinear Differential Equations and their Applications, vol.9. Birkhauser, Basel, 1993.

[35] J.J. Moreau. Evolution problem associated with a moving convex set in a Hilbert space. *Journal of Differential Equations*, 26:347–374, 1977.

[36] J.J. Moreau. Bounded variation in time. In J.J Moreau, P.D. Panagiotopoulos, and G. Strang, editors, *Topics in Nonsmooth Mechanics*, pages 1–74, Basel, 1988. Birkaüser.

[37] J.J. Moreau, "Liaisons unilatérales sans frottement et chocs inélastiques", C. R. Acad. Sci. Paris, Serie II, vol.296, pp.1473–1476, 1983.

[38] R. Dzonou, M.D.P. Monteiro Marques, "A sweeping process approach to inelastic contact problems with general inertia operators", European Journal of Mechanics A/Solids, vol.26, no 3, pp.474-490, May-June 2007.

[39] V. Acary, O. Bonnefon, B. Brogliato, *Nonsmooth Modeling and Simulation for Switched Circuits*, LNEE vol.69, Springer-Verlag Berlin, 2011.

[40] V. Acary, B. Brogliato, *Numerical Methods for Nonsmooth Dynamical Systems*, LNACM vol.35, Springer Verlag Heidelberg, 2008.

[41] http://siconos.gforge.inria.fr/

[42] Robinson, S.M. "Generalized equations and their solutions. I. Basic theory", Mathematical Programming Study, vol.10, pp.128–141, 2010.

[43] R.T. Rockafellar. *Convex Analysis*, Princeton University Press, 1970.

[44] J.B. Hiriart-Urruty, C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer, Berlin Heidelberg 2001.

[45] J.S. Pang, "Three modeling paradigms in mathematical programming", Mathematical Programming B, vol.125, no 2, pp.297-323, 2010.

[46] J. Shen, J.S. Pang, "Semicopositive linear complementarity systems", Int. J. Robust Nonlinear Control,vol.17, no 15, pp.1367-1386, 2007.

[47] M. de la Sen, "Preserving positive realness through discretization", Positivity, vol.6, pp.31-45, 2002.

[48] J.C. Simo, N. Tarnow, "The discrete energy-momentum method. Conserving algorithms for nonlinear elastodynamics", ZAMP, vol.43, pp.757-793, 1992.

[49] C. Studer. *Numerics of Unilateral Contacts and Friction. – Modeling and Numerical Time Integration in Non-Smooth Dynamics*, LNACM vol.47, Springer Verlag Heidelberg, 2009.

[50] F. Facchinei, J.S. Pang *Finite-Dimensional Variational Inequalities and Complementarity Problems; Volume 1*, Springer Series in Operations Research, 2003.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399