



**HAL**  
open science

# Ensemble Approximations for Constrained Dynamical Systems using Liouville Equation

Marianne Souaiby, Aneel Tanwani, Didier Henrion

► **To cite this version:**

Marianne Souaiby, Aneel Tanwani, Didier Henrion. Ensemble Approximations for Constrained Dynamical Systems using Liouville Equation. 2021. hal-03167458v1

**HAL Id: hal-03167458**

**<https://laas.hal.science/hal-03167458v1>**

Preprint submitted on 12 Mar 2021 (v1), last revised 24 Nov 2022 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ENSEMBLE APPROXIMATIONS FOR CONSTRAINED DYNAMICAL SYSTEMS USING LIOUVILLE EQUATION

MARIANNE SOUAIBY, ANEEL TANWANI, AND DIDIER HENRION

ABSTRACT. For a class of state-constrained dynamical systems described by evolution variational inequalities, we study the time evolution of a probability measure which describes the distribution of the state over a set. In contrast to smooth ordinary differential equations, where the evolution of this probability measure is described by the Liouville equations, the flow map associated with the nonsmooth differential inclusion is not necessarily invertible and one cannot directly derive a continuity equation to describe the evolution of the distribution of states. Instead, we consider Lipschitz approximation of our original nonsmooth system and construct a sequence of measures obtained from Liouville equations corresponding to these approximations. This sequence of measures converges in weak-star topology to the measure describing the evolution of the distribution of states for the original nonsmooth system. This allows us to approximate numerically the evolution of moments (up to some finite order) for our original nonsmooth system, using a solver that uses finite order moment approximations of the Liouville equation. Our approach is illustrated with the help of an academic example.

## 1. INTRODUCTION

In the theory of dynamical systems, studying the evolution of state trajectories, both qualitatively and quantitatively, is a common occurrence. For ordinary differential equations, with a *fixed* initial condition described by a point in the finite-dimensional vector space, the tools for analyzing the behavior of trajectories are widely available. However, for many applications, it is of interest to consider the evolution of dynamical systems when the initial condition is described by a distribution over some set in the state space. This article explores this latter direction for a particular class of nonsmooth dynamical systems.

If we consider a probability measure to describe the distribution of the initial conditions of a dynamical system, then the time evolution of this initial probability measure with respect to underlying dynamics is the object of our interest. For an autonomous dynamical system described by an ordinary differential equation (ODE) with Lipschitz continuous vector field, the time evolution of this measure is described by a linear partial differential equation (PDE) called the *Liouville equation* or the continuity equation, see e.g. [33, Section 5.4]. The solution to the Liouville equation, that is the probability measure describing the distribution at time  $t$ , is the pushforward or image measure of the initial probability measure through the flow map at time  $t$ . Lipschitz continuity of the vector field ensures that the flow map of the ODE is invertible, which in turn ensures that the pushforward measure is the unique solution to the Liouville equation. This approach of associating the continuity equation with finite dimensional ODEs has found relevance in numerical optimal control [24, 19] as well as in several control-theoretic problems [4, 8, 9].

---

This work was supported by the ANR project CONVAN with grant number ANR-17-CE40-0019-01.

When the vector field is not Lipschitz continuous, then the study of the evolution of the initial distribution is more involved. The first occurrence of continuity equations corresponding to nonsmooth ODEs occurs in [15]. Continuity equations corresponding to one-sided Lipschitz vector fields have been studied in [6, 7]. In [2], the authors consider less regular ODEs and study uniqueness of solutions for (Lebesgue) almost-all initial conditions by using the Liouville equation.

The dynamical systems for which we want to study the evolution of probability measures (describing the distribution of states) are the so-called *constrained systems* described by differential inclusions. In particular, given a closed convex set  $S \subset \mathbb{R}^n$ , and a continuously differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , we describe the evolution of constrained systems via the differential inclusion

$$(1) \quad \dot{x} \in f(x) - \mathcal{N}_S(x)$$

where  $\mathcal{N}_S(x) \in \mathbb{R}^n$  denotes the outward normal cone to the set  $S$  at the point  $x \in \mathbb{R}^n$ . Since the normal cone takes a zero value in the interior of  $S$ , it is clear that the right-hand side of (1) is potentially discontinuous at the boundary of the set  $S$ . One can also think of (1) as an evolution variational inequality, described as

$$\langle \dot{x}(t) - f(x(t)), y - x(t) \rangle \geq 0, \quad \text{for all } y \in S, x(t) \in S, t \geq 0.$$

Such dynamical systems have been a matter of extensive study in past decades due to their relevance in engineering and physical systems. A recent survey article [10], and a research monograph [1], provide an overview of different research oriented directions in the literature pertaining to system (1) and its connections to different classes of nonsmooth mathematical models. Analysis of such systems requires tools from variational analysis, nonsmooth analysis, set-valued analysis [3, 27, 30]. Results based on stability analysis with computational aspects have been addressed recently by the authors in [31].

For a fixed initial condition,  $x(0) \in S$ , the question of existence and uniqueness of solution to system (1) has already been well-established in the literature, and the origins of such works can be found in [28], see [16] for a recent exposition. However, if we consider the initial conditions described by a probability measure, then the evolution of this measure under the dynamics of (1) has not received any attention in the literature. Since the right-hand side of (1) is a discontinuous mapping with respect to the state, one cannot write down the corresponding Liouville equation directly. In fact, it can be shown that the solutions of system (1) are forward unique, but not backward unique, and hence the flow map of (1) is not invertible in general. One possible approach to describe the evolution of probability measure could be to formulate a set-valued version of the continuity equation and study the solutions of such equations under appropriate hypothesis. This approach has been adopted in [5] but it is not clear to us how to derive the corresponding set-valued partial differential equation for system (1) and whether the resulting inclusion would satisfy the sort of hypothesis required in [5]. We therefore study different techniques, partly with the motivation of numerical simulation, which allow us to approximate the distribution of states as a function of time.

Inspired by the concepts presented in [2], our basic idea is to consider Lipschitz approximations of system (1). The particular approximations that we work with are the ones obtained by *Yosida-Moreau* regularization and are parameterized by a positive scalar converging to zero, see [10]. We can then associate a Liouville equation to each of these approximants, and establish convergence of the resulting sequence of measures. Furthermore, we also show that the support of the sequence of measures converges (with respect to the Hausdorff distance) to the support of the pushforward measure for the nonsmooth system. These analytical results allow us to get an approximation of the actual solution.

Since the pushforward measure, at each time instant, is an infinite-dimensional object, it can be challenging to approximate it numerically. A possibility – that we do not explore here – could be to use Monte-Carlo probabilistic algorithms. Instead, we would like to investigate a purely deterministic approach: in order to get a quantitative measure of the distribution of state at any time instant, we apply the moment-SOS hierarchy [20] which allows us to approximate numerically the moments (up to some finite order) associated with the pushforward measure. Also, using the recent developments on approximating the support of a measure with the Christoffel-Darboux kernel [23], we can approximate the support of the pushforward measure, and hence the trajectories corresponding to a certain initial distribution.

The remainder of the article is organized as follows: In Section 2, we formalize the problem and introduce the basic mathematical elements necessary for doing so. In Section 3, we construct Lipschitz approximations of our initial dynamical system. In Sections 4 and 5, we study certain properties of the sequence of measures associated with approximations constructed in Section 3. Numerical aspects for approximating the moments, and support, of the probability measure describing the evolution of system dynamics are also discussed in Sections 4 and 5. We illustrate our results with the help of an academic example in Section 6. Some concluding remarks with possible future directions appear in Section 7, followed by an Appendix which collects some additional tools used in the development of our results.

## 2. PRELIMINARIES AND PROBLEM FORMULATION

**2.1. Evolution of ensembles.** Let us consider the time-varying ODE

$$(2) \quad \dot{z}(t) = g(t, z(t)), \quad z(0) = z_0,$$

over a given time interval  $[0, T]$ , where  $g : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a given vector field and  $z(t) \in \mathbb{R}^n$  is the state. For each  $t \in [0, T]$ , let us consider the flow map  $G_t : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , so that the mapping  $z_0 \mapsto G_t(z_0)$  provides the value of state trajectory of (2) at time  $t$ , and moreover it satisfies

$$(3) \quad \partial_t G_t(z_0) = g(t, G_t(z_0)), \quad G_0(z_0) = z_0, \quad (t, z_0) \in [0, T] \times \mathbb{R}^n.$$

In this article, we consider the evolution of dynamical systems when the initial condition is defined probabilistically. In particular, we use the notation  $z(0) \sim \xi_0$  to mean that  $z(0)$  is a random variable whose law is a given probability measure, or density function  $\xi_0 \in \mathcal{P}(\mathbb{R}^n)$ , where  $\mathcal{P}(S)$  denotes the set of probability measures supported on  $S$ .

This model allows to capture an initial spatial distribution of particles. To define the corresponding density function at time  $t \geq 0$ , denoted by  $\xi_t \in \mathcal{P}(\mathbb{R}^n)$ , we consider the pushforward or image measure of  $\xi_0$  through the flow map  $G_t(\cdot)$ . That is, let

$$(4) \quad \xi_t := G_t\#\xi_0,$$

so that, for every Borel subset  $B \subset \mathbb{R}^n$ , it holds that

$$\xi_t(B) = \xi_0(G_t^{-1}(B)) = \xi_0(\{z \in \mathbb{R}^n : G_t(z) \in B\}).$$

The evolution of  $\xi_t$  is described by the following PDE, called the continuity or *Liouville* equation:

$$(5) \quad \partial_t \xi_t + \operatorname{div}(\xi_t g) = 0,$$

with the initial condition:

$$(6) \quad \xi|_{t=0} = \xi_0.$$

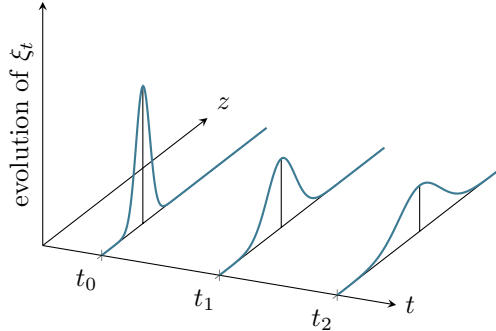


FIGURE 1. Evolution of probability measure  $\xi_t$  w.r.t. time and space.

The Liouville equation (5) should be understood in the sense of distributions, i.e.

$$\int_{\mathbb{R}^n} (\partial_t v(t, z) + \partial_z v(t, z) \cdot g(t, z)) d\xi_t(z) = 0$$

for all continuously differentiable functions  $v$  from  $\mathbb{R}_+ \times \mathbb{R}^n$  to  $\mathbb{R}$ . The equivalence between the solutions of ODE (2) and PDE (5), is established in the following result, see e.g. [33, Theorem 5.34]:

**Theorem 1.** *For each  $t \in [0, T]$ , let  $G_t : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a diffeomorphism so that (3) holds. Given  $\xi_0 \in \mathcal{P}(\mathbb{R}^n)$ , let  $\xi_t$  be defined as in (4). Then,  $\xi_t$  is the unique solution of the Liouville equation (5)-(6) over the time interval  $[0, T]$ .*

The importance of the *Liouville* PDE relies on its linearity in the probability measure  $\xi_t$ , whereas the *Cauchy* ODE is nonlinear in the state trajectory  $z(t)$ . This PDE governs the time evolution of a measure transported by the flow of a nonlinear dynamical system. The nonlinear dynamics is then replaced by a linear equation on measures. It is important to note that, in Theorem 1, the equivalence is established under the assumption that  $G_t$  is a diffeomorphism for each  $t \in [0, T]$ , which in particular requires that the flow map  $G_t$  is invertible. ODEs with Lipschitz vector fields have this property, but when the vector field is not Lipschitz continuous in state variable, the backward invertibility assumption may not hold, or the flow map  $G_t$  may itself not be uniquely defined.

**2.2. Ensembles of constrained system.** In this paper, we are interested in studying a class of dynamical systems described by the variational inequalities

$$(7) \quad \dot{z}(t) \in f(t, z(t)) - \mathcal{N}_{S(t)}(z(t)), \quad z(0) \sim \xi_0,$$

over an interval  $[0, T]$  for some given  $T > 0$ , where  $f : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a given vector field,  $S : [0, T] \rightrightarrows \mathbb{R}^n$  a compact and convex-valued mapping, and we recall that the normal cone to  $S$  at  $z$  is defined by

$$(8) \quad \mathcal{N}_S(z) := \{\lambda \in \mathbb{R}^n \mid \langle \lambda, z' - z \rangle \leq 0, \forall z' \in S\}.$$

If  $z \in \text{int}(S)$ , the interior of  $S$ , then  $\mathcal{N}_S(z) = 0$  and by convention, we let  $\mathcal{N}_S(z) := \emptyset$  for all  $z \notin S$ . The formalism of system (7) with inclusion naturally allows us to describe dynamics constrained to evolve in set  $S$ . Using the depiction in Figure 2, it is seen that, during the evolution of a trajectory, if  $z(t)$  is in interior of  $S$ , then  $\mathcal{N}_S(z(t)) = 0$  and the motion of the trajectory continues according to the differential equation  $\dot{z}(t) = f(t, z(t))$ . Whenever  $z(t)$  is on the boundary, we add a vector from the set  $-\mathcal{N}_S(z(t))$ , which restricts the motion of the state trajectory in tangential direction on the boundary of the constraint set  $S$ .

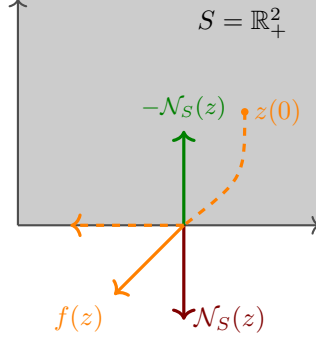


FIGURE 2. State trajectories in constrained system with  $S = \mathbb{R}_+^n$ .

We emphasize that, in (7),  $\xi_0 \in \mathcal{P}(S(0))$  is a probability measure that specifies the distribution of the initial state. For each  $t \in [0, T]$ , let us denote the flow map by  $F_t : S(t) \rightarrow S(t)$ , so that  $z_0 \mapsto F_t(z_0)$  is the value at time  $t$  of the state trajectory of (7) with  $z(0) = z_0$ . Given this random initial condition, the state at each time  $t$  can also be interpreted as a random variable in  $S(t)$ , i.e.  $z(t) \sim \xi_t \in \mathcal{P}(S(t))$  defined by  $\xi_t := F_t \# \xi_0$ . However, unlike Lipschitz continuous ODEs, the mapping  $F_t$  is not invertible in general. An example illustrating this fact is given next.

**Example 1** (Flow map not invertible). Let  $f(z) = Az$  with  $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$  and  $S = \mathbb{R}_+^2$  and let  $z_0$  be a given initial condition, with angle  $\theta_0$ . For  $t \leq \theta_0$ , we have  $z(t) = F_t(z_0) = e^{At}z_0 = \begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix} z_0$ . And for  $t \geq \theta_0$ , we have  $z(t) = [|z_0| \ 0]^\top$ . For example if  $z_0 = [1 \ 1]^\top$ , it holds  $\theta_0 = \frac{\pi}{4}$  and then for  $t \geq \theta_0$ , we have  $z(t) = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = [\sqrt{2} \ 0]^\top$ . The flow map reads

$$z(t) = F_t(z_0) = \begin{cases} e^{At}z_0 & \text{if } t \leq \theta_0 \\ [|z_0| \ 0]^\top & \text{if } t \geq \theta_0. \end{cases}$$

Indeed, as we can observe, the flow map is not invertible since given a state  $z(t)$  for a given time  $t \geq \theta_0$ , it is not possible to retrieve the initial condition  $z_0$ .

As a consequence of Example 1, it is seen that the flow map associated with dynamical system (7) is not necessarily invertible, and hence the conditions of Theorem 1 are not satisfied in general for such systems. Therefore, to study the propagation of the probability measure  $\xi_0$ , it is not possible to write down the Liouville equation directly to obtain  $\xi_t := F_t \# \xi_0$ . In this article, our goal is to find alternate methods to approximate the measure  $\xi_t$  and propose computational algorithms to calculate such approximations numerically.

**2.3. Problem Formulation.** We consider the dynamical system (7) with flow map  $F_t : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . For a given  $\xi_0 \in \mathcal{P}(S(0))$ , since there is no direct derivation of the PDE for characterizing the evolution of  $\xi_t := F_t \# \xi_0$ , we compute an approximation of  $\xi_t$  as follows:

- Construct a sequence of ODEs with Lipschitz continuous right-hand sides which approximate the solution of ODE (7) for a fixed initial condition. This construction is based on a regularization of (7), and results in a sequence parameterized by a scalar  $\lambda > 0$ .
- Exploit the regularity of the approximating ODE to construct a sequence of measures  $\xi_t^\lambda := F_t^\lambda \# \xi_0$ .

- When  $\lambda$  tends to 0, prove that  $\xi_t^\lambda$  converges to  $\xi_t := F_t \# \xi_0$  in the weak-star topology. In particular, all finite order moments of  $\xi_t^\lambda$  converge to the moments of  $\xi_t$ .
- When  $\lambda$  tends to 0, prove the convergence of the support of  $\xi_t^\lambda$  to the support of  $\xi_t$  in the Hausdorff metric.

From a computational viewpoint, the by-product of the above results is that, for a fixed  $\lambda > 0$ , one can invoke efficient numerical methods for computing moments associated with the probability measure  $\xi_t^\lambda$  and the support of  $\xi_t^\lambda$ . This allows us to compute an approximation of the moments and support of  $\xi_t$  associated with nonsmooth system (7).

### 3. LIPSCHITZ APPROXIMATION

The first step in our analysis is to compute an approximation of the solutions of (7) by using Moreau-Yosida regularization. The development carried out here is inspired by [10]. We introduce a sequence of approximate solutions, the so-called Moreau-Yosida approximants  $\{z_\lambda\}_{\lambda>0}$ , which are obtained by solving the following ODE parameterized by  $\lambda > 0$ :

$$(9) \quad \dot{z}_\lambda(t) = f(t, z_\lambda(t)) - \frac{1}{\lambda}(z_\lambda(t) - \text{proj}(z_\lambda(t), S(t))), \quad z_\lambda(0) = z_0 \in S(0)$$

over the interval  $[0, T]$ , where  $\text{proj}(z, S)$  is the (unique) Euclidean projection of vector  $z$  onto convex set  $S$ . It is observed that, for each  $\lambda > 0$ , the right-hand side of (9) is (globally) Lipschitz continuous, and therefore, there exists a continuously differentiable trajectory  $z_\lambda : [0, T] \rightarrow \mathbb{R}^n$  such that (9) holds for every  $t \in [0, T]$ . The relation between the solution of the inclusion (7) and the approximants  $\{z_\lambda\}_{\lambda>0}$  holds under the following assumptions:

**Assumption 1.** There exists a constant  $L_f > 0$  such that, for each  $t \in [0, T]$ ,

$$\begin{aligned} |f(t, z)| &\leq L_f(1 + |z|), \quad \forall z \in \mathbb{R}^n \\ |f(t, z_1) - f(t, z_2)| &\leq L_f|z_1 - z_2|, \quad \forall z_1, z_2 \in \mathbb{R}^n. \end{aligned}$$

**Assumption 2.** The mapping  $S : [0, T] \rightrightarrows \mathbb{R}^n$  is closed and convex-valued for each  $t \in [0, T]$ , and  $S(\cdot)$  varies in a Lipschitz continuous manner with time, that is, there exists a constant  $L_S \geq 0$ , such that

$$d_H(S(t_1), S(t_2)) \leq L_S|t_1 - t_2|, \quad \forall t_1, t_2 \in [0, T].$$

The notation  $d_H(A, B)$  means the Hausdorff distance between sets  $A$  and  $B$ , that is,

$$(10) \quad d_H(A, B) := \max \left\{ \sup_{y \in A} d(y, B), \sup_{x \in B} d(x, A) \right\}$$

where  $d(x, A)$  denotes the Euclidean distance between vector  $x$  and set  $A$ .

**Theorem 2.** *Under Assumptions 1–2, consider the sequence of solutions  $\{z_\lambda\}_{\lambda>0}$  to parameterized ODE (9) on an interval  $[0, T]$ . Then, as  $\lambda \rightarrow 0$ , the sequence converges uniformly to a Lipschitz continuous function  $z : [0, T] \rightarrow \mathbb{R}^n$ , the unique solution to the differential inclusion (7).*

The proof of this theorem is described in the remainder of this section. Certain calculations, leading to the intermediate lemmas used in the proof, have been included in the appendix.

*Proof.* The basic idea of the proof is to show that the sequence  $\{z_\lambda\}_{\lambda>0}$  satisfies bounds ensuring uniform convergence to a function  $z(\cdot)$  solving (7). This development is carried out in four steps.

*Step 1: Estimates on the sequence  $\{z_\lambda\}_{\lambda>0}$ .* As a first step, to obtain bounds on the norm of  $z_\lambda(\cdot)$ , let us begin by computing bounds on the norm of  $\dot{z}_\lambda(\cdot)$  as stated in the following lemma, whose proof is relegated to Appendix A.

**Lemma 3.** *For each  $\lambda > 0$ , it holds*

$$(11) \quad |\dot{z}_\lambda(t)| \leq 2L_f + L_f|z_\lambda(t)| + L_f \max_{0 \leq s \leq t} |z_\lambda(s)| + L_S,$$

where  $L_f, L_S$  were introduced in Assumptions 1 and 2 respectively.

Based on Lemma 3, let us now calculate  $\frac{d}{dt}|z_\lambda(t)|^2$  for getting an estimate on  $|z_\lambda(\cdot)|$ . First, we observe that

$$(12) \quad \frac{d}{dt}|z_\lambda(t)|^2 = 2\langle z_\lambda(t), \dot{z}_\lambda(t) \rangle \leq 2|z_\lambda(t)||\dot{z}_\lambda(t)|.$$

Substituting (11) in (12) yields

$$\frac{d}{dt}|z_\lambda(t)|^2 \leq 2L_f|z_\lambda(t)|^2 + 2L_f|z_\lambda(t)| \cdot \max_{0 \leq s \leq t} |z_\lambda(s)| + (4L_f + 2L_S)|z_\lambda(t)|.$$

Let  $y_\lambda(t) = |z_\lambda(t)|^2$ , so

$$\frac{d}{dt}y_\lambda(t) \leq 2L_f y_\lambda(t) + 2L_f \sqrt{y_\lambda(t)} \cdot \max_{0 \leq s \leq t} \sqrt{y_\lambda(s)} + (4L_f + 2L_S)\sqrt{y_\lambda(t)}.$$

Since the right-hand side of this differential inequality results in a nonnegative and nondecreasing function, it follows that  $y_\lambda(t) \leq \widehat{y}_\lambda(t)$ , for all  $t \in [0, T]$ , where  $\widehat{y}_\lambda$  satisfies

$$(13) \quad \begin{aligned} \frac{d}{dt}\widehat{y}_\lambda(t) &= 2L_f\widehat{y}_\lambda(t) + 2L_f\sqrt{\widehat{y}_\lambda(t)} \cdot \sqrt{\widehat{y}_\lambda(t)} + (4L_f + 2L_S)\sqrt{\widehat{y}_\lambda(t)} \\ &= 4L_f\widehat{y}_\lambda(t) + (4L_f + 2L_S)\sqrt{\widehat{y}_\lambda(t)}. \end{aligned}$$

By using the substitution  $v(t) = (\widehat{y}_\lambda(t))^{\frac{1}{2}}$  in (13), it yields

$$\dot{v}(t) = 2L_f v(t) + 2L_f + L_S.$$

The solution of this differential equation is  $v(t) = e^{2L_f t}v(0) + (e^{2L_f t} - 1)\frac{(2L_f + L_S)}{2L_f}$ . Consequently,  $|z_\lambda(t)|^2 = y_\lambda(t) \leq \widehat{y}_\lambda(t) = v(t)^2$ , and we obtain

$$(14) \quad |z_\lambda(t)| \leq e^{2L_f T}|z_\lambda(0)| + (e^{2L_f T} - 1)\frac{(2L_f + L_S)}{2L_f},$$

so that  $|z_\lambda(t)|$  is bounded on the interval  $[0, T]$ , independently of  $\lambda$ .

*Step 2: Extracting a converging subsequence.* Based on the estimates in Step 1, there exists a subsequence of  $z_\lambda(\cdot)$  which converges to  $z(\cdot)$ . More formally, the following statement is obtained.

**Lemma 4.** *There exists a subsequence  $\{z_{\lambda_i}\}_{i \in \mathbb{N}}$  which converges uniformly to a Lipschitz continuous function  $z(\cdot)$  on  $[0, T]$ .*

The proof of Lemma 4 is a consequence of the Arzelà-Ascoli theorem since the sequence  $\{z_{\lambda_i}\}_{i \in \mathbb{N}}$  is continuously differentiable and  $\{\dot{z}_{\lambda_i}\}_{i \in \mathbb{N}}$  is uniformly bounded. The limit function  $z(\cdot)$  is also Lipschitz continuous in this case.

*Step 3: Limit is a solution.* To finish the proof of Theorem 2, we just need to show that the limit  $z(\cdot)$  satisfies the differential inclusion (7). This particular step requires a variational inequality, which is stated in the following lemma, and its proof is given in Appendix A.



**Lemma 5.** *If  $\varphi : [0, T] \rightarrow \mathbb{R}^n$  is a continuous function that satisfies  $\varphi(s) + \int_{t_1}^s f(r, z_\lambda(r)) dr \in S(s)$  for each  $s \in [t_1, t_2]$ , for  $t_1, t_2 \in [0, T]$ , then*

$$(15) \quad \int_{t_1}^{t_2} \langle \varphi(s), \dot{z}(s) - f(s, z(s)) \rangle ds \geq \frac{1}{2} \left( \left\| z(t_2) - \int_{t_1}^{t_2} f(r, z(r)) dr \right\|^2 - \|z(t_1)\|^2 \right).$$

We now complete the proof of Theorem 2 by showing that the limit of the converging subsequence  $z(\cdot)$  satisfies  $\dot{z}(t) \in f(t, z(t)) - \mathcal{N}_{S(t)}(z(t))$  that is,  $\langle \xi - z(t), \dot{z}(t) - f(t, z(t)) \rangle \geq 0$ , for any  $\xi \in S(t)$  and for almost every  $t \geq 0$ . This is indeed the case, since for every  $\xi \in S(t)$ , we can take a Lipschitz continuous function  $\varphi : [t, T] \rightarrow \mathbb{R}^n$  such that, due to Lemma 5, we get

$$\int_{[t, t+\epsilon[} \langle \varphi(s), \dot{z}(s) - f(s, z(s)) \rangle ds \geq \frac{1}{2} \left( \|z(t+\epsilon) - \int_t^{t+\epsilon} f(r, z(r)) dr\|^2 - \|z(t)\|^2 \right),$$

and by letting  $\varphi(s) = \xi - (\xi - \varphi(s))$ , we obtain

$$\begin{aligned} & \int_{[t, t+\epsilon[} \langle \xi, \dot{z}(s) - f(s, z(s)) \rangle ds - \int_{[t, t+\epsilon[} \langle \xi - \varphi(s), \dot{z}(s) - f(s, z(s)) \rangle ds \\ & \geq \frac{1}{2} \langle z(t+\epsilon) - \int_t^{t+\epsilon} f(r, z(r)) dr + z(t), z(t+\epsilon) - \int_t^{t+\epsilon} f(r, z(r)) dr - z(t) \rangle, \end{aligned}$$

which implies

$$\begin{aligned} & \langle \xi, z(t+\epsilon) - z(t) - \int_t^{t+\epsilon} f(s, z(s)) ds \rangle - \int_t^{t+\epsilon} \langle \xi - \varphi(s), \dot{z}(s) - f(s, z(s)) \rangle ds \\ & \geq \frac{1}{2} \langle z(t+\epsilon) - \int_t^{t+\epsilon} f(r, z(r)) dr + z(t), z(t+\epsilon) - \int_t^{t+\epsilon} f(r, z(r)) dr - z(t) \rangle. \end{aligned}$$

From this, we get

$$\begin{aligned} & \left\langle \xi - \frac{1}{2} \left( z(t+\epsilon) - \int_t^{t+\epsilon} f(r, z(r)) dr + z(t) \right), z(t+\epsilon) - z(t) - \int_t^{t+\epsilon} f(s, z(s)) ds \right\rangle \\ & \geq \int_t^{t+\epsilon} \langle \xi - \varphi(s), \dot{z}(s) - f(s, z(s)) \rangle ds \\ & \geq -\epsilon \max_{s \in [t, t+\epsilon[} |\xi - \varphi(s)| |\dot{z}(s) - f(s, z(s))| \\ & \geq -\epsilon \max_{s \in [t, t+\epsilon[} |\xi - \varphi(s)| |\dot{z}(s)| - \epsilon L_f \max_{s \in [t, t+\epsilon[} |\xi - \varphi(s)| (1 + |z(s)|). \end{aligned}$$

Since  $z(\cdot)$  is Lipschitz continuous,  $z(\cdot)$  is bounded on  $[0, T]$  and differentiable almost everywhere on  $[0, T]$ . Hence, for almost every  $t \in [0, T]$ , where  $z(\cdot)$  is differentiable, dividing the last inequality by  $\epsilon$ , we get

$$\begin{aligned} & \left\langle \xi - \frac{1}{2} \left( z(t+\epsilon) - \int_t^{t+\epsilon} f(r, z(r)) dr + z(t) \right), \frac{z(t+\epsilon) - z(t)}{\epsilon} - \frac{\int_t^{t+\epsilon} f(s, z(s)) ds}{\epsilon} \right\rangle \\ & \geq -M \max_{s \in [t, t+\epsilon[} |\xi - \varphi(s)| - ML_f \max_{s \in [t, t+\epsilon[} |\xi - \varphi(s)|, \end{aligned}$$

for some constant  $M > 0$ . Letting  $\epsilon$  tend to zero, we get

$$\langle \xi - z(t), \dot{z}(t) - f(t, z(t)) \rangle \geq 0, \text{ for each } \xi \in S(t),$$

and hence,  $z(\cdot)$  satisfies the differential inclusion (7).  $\square$

*Remark 6.* In the literature, we can find several proofs of convergence of solutions obtained from Moreau-Yosida regularization to the solution of systems closely related to (7), see for example [10, 22, 29]. The proof technique adopted here closely

follows the outline given in [10], but the difference here is that we add the Lipschitz perturbation  $f(t, z)$  on the right-hand side of (7), which modifies certain calculations.

#### 4. CONVERGENCE OF MEASURES

Using the results from the previous section on the convergence of solutions for fixed initial condition, we now study the evolution of probability measures for system (7). As before, let us assume that  $z(0)$  is a random variable whose law is a given probability measure  $\xi_0 \in \mathcal{P}(S(0))$ . We recall that the flow map for system (7) is denoted by  $F_t$ , so that  $t \mapsto z(t) := F_t(z_0)$  is the unique solution to (7).

For the Lipschitz approximation given in (9), consider the map  $F_t^\lambda : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , so that  $t \mapsto z_\lambda(t) := F_t^\lambda(z_0)$  defines the unique solution to (9). Since the right-hand side of (9) is Lipschitz continuous for each  $\lambda > 0$ , we can consider a sequence of probability measures  $\xi_t^\lambda \in \mathcal{P}(S(t))$  defined as

$$\xi_t^\lambda := F_t^\lambda \# \xi_0$$

for each  $t \in [0, T]$  and  $\lambda > 0$ . From Theorem 1, it follows that  $\xi_t^\lambda$  satisfies the partial differential equation:

$$(16) \quad \partial_t \xi_t^\lambda + \operatorname{div}(\xi_t^\lambda f_t^\lambda) = 0$$

in the sense of distributions, with the initial condition  $\xi|_{t=0} = \xi_0$ , and

$$(17) \quad f_t^\lambda(z) := f(t, z) - \frac{1}{\lambda} \left( z - \operatorname{proj}(z, S(t)) \right).$$

On the other hand, we do not know how to derive a meaningful PDE for  $\xi_t$ . However, in the sequel, we show that the probability measure  $\xi_t$  can be approximated by  $\xi_t^\lambda$  as  $\lambda \rightarrow 0$ . This way, a good numerical approximation of  $\xi_t^\lambda$  would also provide an approximation of  $\xi_t$ .

**4.1. Weak-star convergence.** We first show convergence in the weak-star topology. This allows us to approximate the evolution of the moments of the measure  $\xi_t$  using the moments of  $\xi_t^\lambda$ . Given a measure  $\xi$ , we denote its support by  $\operatorname{supp}(\xi)$ , defined as the smallest closed set whose complement has zero measure with respect to  $\xi$ . Equivalently, it is the smallest closed set for which every point has a neighborhood of positive measure with respect to  $\xi$ .

**Proposition 7.** *Let  $v : \mathbb{R}^n \rightarrow \mathbb{R}$  be a continuous function, and assume that  $\xi_0$  has bounded support. Then,*

$$(18) \quad \lim_{\lambda \rightarrow 0} \int_{\mathbb{R}^n} v(z) d\xi_t^\lambda(z) = \int_{\mathbb{R}^n} v(z) d\xi_t(z).$$

*Proof.* By definition of the pushforward measure  $\xi_t^\lambda$ , it holds

$$(19) \quad \int_{\mathbb{R}^n} v(z) d\xi_t^\lambda(z) = \int_{\mathbb{R}^n} v(F_t^\lambda(y)) d\xi_0(y)$$

for all continuous functions  $v$ . From Theorem 2, for each  $t \in [0, T]$ , we have  $\lim_{\lambda \rightarrow 0} z_\lambda(t) = z(t)$ , which is equivalent to

$$\lim_{\lambda \rightarrow 0} F_t^\lambda(y) = F_t(y), \quad \forall y \in S(0).$$

Since  $v$  is any continuous function, this implies

$$\lim_{\lambda \rightarrow 0} v(F_t^\lambda(y)) = v(F_t(y)).$$

By assumption,  $v \circ F_t^\lambda$  is bounded on the bounded set  $\text{supp}(\xi_0)$ . This allows us to invoke Lebesgue's dominated convergence theorem to get

$$(20) \quad \lim_{\lambda \rightarrow 0} \int_{\mathbb{R}^n} v(F_t^\lambda(y)) d\xi_0(y) = \int_{\mathbb{R}^n} v(F_t(y)) d\xi_0(y).$$

Hence, (19) and (20) yield

$$\lim_{\lambda \rightarrow 0} \int_{\mathbb{R}^n} v(z) d\xi_t^\lambda(z) = \int_{\mathbb{R}^n} v(F_t(y)) d\xi_0(y).$$

Using again the change of variables formula, we obtain

$$\lim_{\lambda \rightarrow 0} \int_{\mathbb{R}^n} v(z) d\xi_t^\lambda(z) = \int_{\mathbb{R}^n} v(z) d\xi_t(z)$$

for all continuous functions  $v$  on  $\mathbb{R}^n$ . Therefore, the equality in (18) is proved.  $\square$

*Remark 8.* In the proof of Proposition 7, the boundedness of  $\text{supp}(\xi_0)$  was used to invoke dominated convergence theorem. The result of Proposition 7 extends in some cases where  $\text{supp}(\xi_0)$  is unbounded. In particular, if it can be shown that there exists a function  $g : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$  such that, for each  $\lambda > 0$ ,

$$|F_t^\lambda(y)| \leq g(t, y), \quad t \in [0, T]$$

then the convergence in (18) holds for all continuous functions  $v$  which satisfy

$$\int_{\mathbb{R}^n} v(g(t, y)) d\xi_0(y) < \infty, \quad t \in [0, T].$$

**4.2. Relations Describing Moments.** An immediate consequence of Proposition 7 is that we can get a desired approximation of the moments of  $\xi_t$  by choosing appropriate test functions  $v$ . This amounts to computing the moments of  $\xi_t^\lambda$ . We will now explore numerical techniques which allow us to compute the solution of (16) by computing the desired moments.

Toward this end, we first observe that the Liouville equation (16) can be equivalently written as a linear PDE satisfied by the occupation measures  $d\mu^\lambda := dt d\xi_t^\lambda$ ,  $\mu_0^\lambda := \delta_0 \xi_0$ ,  $\mu_T^\lambda := \delta_T \xi_T$ , which is

$$(21) \quad \partial_t \mu^\lambda + \text{div}(\mu^\lambda f_\lambda) + \mu_T^\lambda = \mu_0^\lambda$$

which again should be understood in the sense of distributions, i.e.

$$\begin{aligned} & \int_{\mathbb{R}_+} \int_{\mathbb{R}^n} (\partial_t v(t, z) + \partial_z v(t, z) \cdot f_\lambda(t, z)) d\mu^\lambda(t, z) \\ &= \int_{\mathbb{R}_+} \int_{\mathbb{R}^n} v(t, z) (d\mu_T^\lambda(t, z) - d\mu_0^\lambda(t, z)) \end{aligned}$$

for all continuously differentiable functions  $v$ .

We compute approximate moments of  $\mu^\lambda$  by applying the moment-SOS hierarchy [20]. This method consists of minimizing a functional subject to the following constraints:

- (1) The *Liouville* equation (21) expressed in the sense of distributions, as a linear constraints on the moments of  $\mu^\lambda$  and  $\mu_T^\lambda$ .
- (2) Necessary linear matrix inequality (LMI) constraints based on the dual of Putinar's Positivstellensatz.

We will see in the following how to formulate the *Liouville* equation (21) as a linear moment constraint.

Let  $g$  be a polynomial vector field defined as

$$g : \underbrace{(z_1, z_2, \dots, z_n)}_z \in \mathbb{R}^n \mapsto (g_1, g_2, \dots, g_n) \in \mathbb{R}^n,$$

and  $v$  be a monomial test function, with a maximum degree  $d \in \mathbb{N}$ , defined as

$$v : (t, z) \mapsto t^a z^b := t^a z_1^{b_1} z_2^{b_2} \dots z_n^{b_n},$$

for all  $(a, b) \in \mathbb{N}^{n+1}$ , with  $a + b_1 + b_2 + \dots + b_n \leq d$ . The maximal degree  $d$  is called the relaxation degree.

Besides, let us denote

$$(22) \quad y_{a-1,b} := \int_0^T \int_{\mathbb{R}^n} t^{a-1} z^b d\mu^\lambda(t, z)$$

and

$$(23) \quad y_{a,b}^T := \int_0^T \int_{\mathbb{R}^n} t^a z^b d\mu_T^\lambda(t, z),$$

$$(24) \quad y_{a,b}^0 := \int_0^T \int_{\mathbb{R}^n} t^a z^b d\mu_0^\lambda(t, z).$$

Let  $e_i$  denote the vector whose only non-zero entry is equal to one at position  $i$ .

**Proposition 9.** *The Liouville equation (21) is equivalently expressed as:*

$$(25) \quad y_{a,b}^T - y_{a,b}^0 = ay_{a-1,b} + \sum_{i=1}^n \int_0^T \int_{\mathbb{R}^n} b_i t^a z^{b-e_i} g_i(z) d\mu^\lambda(t, z)$$

which are linear constraints that link the moments of the initial measure, terminal measure and occupation measure.

*Proof.* Choosing  $v(t, z) = t^a z^b$  as a monomial test function, the Liouville equation (21) is then written as

$$\langle \partial_t \mu^\lambda, v \rangle + \langle \operatorname{div}(\mu^\lambda g), v \rangle + \langle \mu_T^\lambda, v \rangle = \langle \mu_0^\lambda, v \rangle,$$

which implies

$$(26) \quad \int_0^T \int_{\mathbb{R}^n} (\partial_t v(t, z) + \partial_z v(t, z) \cdot g(z)) d\mu^\lambda(t, z) = \int_0^T \int_{\mathbb{R}^n} v(t, z) (d\mu_T^\lambda(t, z) - d\mu_0^\lambda(t, z)).$$

We have

$$\partial_t v(t, z) = at^{a-1} z^b,$$

and

$$\partial_z v(t, z) = (b_1 t^a z_1^{b_1-1} z_2^{b_2} \dots z_n^{b_n}, b_2 t^a z_1^{b_1} z_2^{b_2-1} \dots z_n^{b_n}, \dots, b_n t^a z_1^{b_1} z_2^{b_2} \dots z_n^{b_n-1}).$$

Replacing  $\partial_t v(t, z)$  and  $\partial_z v(t, z)$  by their expressions in (26) yields

$$\begin{aligned} & \int_0^T \int_{\mathbb{R}^n} (at^{a-1} z^b + \sum_{i=1}^n b_i t^a z^{b-e_i} g_i(z)) d\mu^\lambda(t, z) \\ &= \int_0^T \int_{\mathbb{R}^n} t^a z^b d\mu_T^\lambda(t, z) - \int_0^T \int_{\mathbb{R}^n} t^a z^b d\mu_0^\lambda(t, z) \end{aligned}$$

which is the expected statement by using the notations (22), (23) and (24).  $\square$

**4.3. Numerical computation.** Let  $\mathbb{N}_d^p := \{a \in \mathbb{N}^p; a_1 + \dots + a_p \leq d\}$  and let  $y = (y_a)_{a \in \mathbb{N}^p}$  be a given sequence. The moment matrix of degree  $d$ , denoted by  $M_d(y)$  is a matrix whose rows and columns are indexed by monomials of degree at most  $d$ . For  $a, b \in \mathbb{N}_d^p$ , the corresponding entry in  $M_d(y)$  is defined by  $(M_d(y))_{a,b} := y_{a+b}$ .

Let  $b_d(z) := (z^a)_{a \in \mathbb{N}_d^p} \in \mathbb{R}[z]_d^{s(d)}$  denote the vector of monomials of degree less or equal to  $d$ , where  $s(d) := \binom{p+d}{d}$  is the number of monomials of degree at most  $d$ . If the sequence  $y$  has a representing measure  $\xi$ , i.e.  $y_a = \int_{\mathbb{R}^p} z^a d\xi(z)$  for all  $a \in \mathbb{N}^p$ , we can use the equivalent definition  $M_d(y) := \int_{\mathbb{R}^p} b_d(z) b_d(z)^\top d\xi(z)$ , where the integral is understood entrywise. And we can also define the localizing matrix of degree  $d$  with respect to a given  $q(z) \in \mathbb{R}[z]$  by  $M_{d-\lceil \deg(q)/2 \rceil}(qy) := \int_{\mathbb{R}^p} q(z) b_d(z) b_d(z)^\top d\xi(z)$  where  $\lceil x \rceil$  denotes the smallest integer greater than  $x$ .

Assume that  $X \subset \mathbb{R}^n$  is a compact basic semialgebraic set i.e.

$$X := \{z \in \mathbb{R}^n : p_k(z) \geq 0, \quad k = 0, \dots, n_X\}$$

for given  $p_k \in \mathbb{R}[z]$ ,  $k = 0, \dots, n_X$ . Let  $p_0(z) = 1$  and let one of the inequalities  $p_k(z) \geq 0$  be of the form  $R - \sum_{i=1}^n z_i^2 \geq 0$  where  $R$  is a sufficiently large positive constant.

**Proposition 10.** (*Putinar's Theorem*) *The sequence of moments  $y$  has a representing measure supported on  $X$  if and only if  $M_{d-\lceil \deg p_k / 2 \rceil}(p_k y)$ ,  $k = 0, \dots, n_X$  are positive semidefinite for all  $d \in \mathbb{N}$ .*

The following corollary shows that for fixed  $d \in \mathbb{N}$ , we have an LMI necessary condition for a truncated sequence  $y$  to have a representing measure supported on  $X$ .

**Corollary 11.** *Given  $d \in \mathbb{N}$ , if the sequence of moments  $y$  has a representing measure supported on  $X$ , then  $M_{d-\lceil \deg(p_k) / 2 \rceil}(p_k y)$ ,  $k = 0, \dots, n_X$  are positive semidefinite.*

Numerically, for applying the moment-SOS hierarchy, we need first to define the initial, terminal and occupation measures associated to each time and state. Since the initial measure is given, the moments of the initial measure are given. The moments of the occupation measure as well as those of the terminal measure are the unknowns, which appear linearly in the relaxation (25).

In addition to those linear constraints that link the moments, there are also necessary LMI constraints based on Corollary 11. Solving the LMI problem gives us an approximation of the moments of the occupation measure and the terminal measure. The LMI constraints are automatically constructed by the `msdp` command in Gloptipoly for Matlab [21]. For more details about the LMI constraints, refer to [18, Section 3.3] or [20].

## 5. CONVERGENCE OF SUPPORT OF MEASURES

For several applications, it is important to approximate the support of the measure  $\xi_t$ , since it provides a probabilistic estimate of the state trajectories at time  $t \in [0, T]$ . Once again, our goal is to approximate the support of  $\xi_t$  by the support of  $\xi_t^\lambda$  where  $\xi_t^\lambda$  satisfies (16).

**5.1. Hausdorff convergence of support.** We first show that  $\text{supp}(\xi_t^\lambda)$  converges in the Hausdorff distance to  $\text{supp}(\xi_t)$ .

**Proposition 12.** *For each  $t \in [0, T]$ , it holds*

$$(27) \quad \lim_{\lambda \rightarrow 0} d_H(\text{supp}(\xi_t^\lambda), \text{supp}(\xi_t)) = 0.$$

*Proof.* First, let  $A_t^\lambda := \text{supp}(\xi_t^\lambda)$  and  $A_t := \text{supp}(\xi_t)$ . For proving that  $\lim_{\lambda \rightarrow 0} d_H(A_t^\lambda, A_t) = 0$ , we need to prove the following two limits:

$$(28) \quad \lim_{\lambda \rightarrow 0} \sup_{y_\lambda \in A_t^\lambda} d(y_\lambda, A_t) = 0,$$

and

$$(29) \quad \lim_{\lambda \rightarrow 0} \sup_{x \in A_t} d(x, A_t^\lambda) = 0.$$

For proving (28), we first observe that

$$\sup_{y_\lambda \in A_t^\lambda} d(y_\lambda, A_t) = \sup_{y_\lambda \in A_t^\lambda} \inf_{x \in A_t} |y_\lambda - x|,$$

and hence it needs to be shown that for every  $y_\lambda \in A_t^\lambda$ , there exists  $x \in A_t$  such that  $|x - y_\lambda|$  converges to zero as  $\lambda$  converges to zero. Since  $y_\lambda \in A_t^\lambda$ , there exists  $z_0 \in \text{supp}(\xi_0)$  such that  $y_\lambda = F_t^\lambda(z_0)$ . By choosing  $x = F_t(z_0) \in A_t$ , it follows from Theorem 2 that  $\lim_{\lambda \rightarrow 0} F_t^\lambda(z_0) = F_t(z_0)$ , or equivalently,  $|x - y_\lambda|$  converges to 0 as  $\lambda \rightarrow 0$ .

For proving (29), we similarly observe that

$$\sup_{x \in A_t} d(x, A_t^\lambda) = \sup_{x \in A_t} \inf_{y_\lambda \in A_t^\lambda} |x - y_\lambda|.$$

Following the same idea as before, let us take  $x \in A_t$ , then there exists  $z_0 \in \text{supp}(\xi_0)$  such that  $x = F_t(z_0)$ . By choosing  $y_\lambda = F_t^\lambda(z_0) \in A_t^\lambda$ , it again follows from Theorem 2 that  $|x - y_\lambda|$  converges to 0 as  $\lambda \rightarrow 0$ , and (29) is obtained.  $\square$

**5.2. Approximation of support.** Just like the approximation of moments, we can provide some numerical methods to approximate the support of the sequence of measures  $\xi_t^\lambda$ . By Proposition 12, by computing such an approximation for  $\lambda > 0$  sufficiently small, we get an approximation of the support of the probability measure  $\xi_t$  for the original system.

Let  $M_d^\lambda(y)$  denote the moment matrix of degree  $d$  of the sequence of moments  $y$  of the measure  $\xi_t^\lambda$ . Let  $\xi_t^\lambda$  be absolutely continuous with respect to the Lebesgue measure and let us suppose that  $\text{supp}(\xi_t^\lambda)$  is compact. The following statement shows the existence of a sublevel set that approximates the support of the sequence of measures  $\xi_t^\lambda$  when  $\lambda$  is fixed. This result and its proof are established in [23, Theorem 3.9].

**Proposition 13.** *For every  $\epsilon > 0$  (small enough), there exists  $d \in \mathbb{N}$  (large enough) and  $\gamma_d > 0$ , such that for  $\lambda > 0$  fixed, the sublevel set*

$$(30) \quad S_d^\lambda := \{z \in \mathbb{R}^p \mid b_d(z)^\top M_d^\lambda(y)^{-1} b_d(z) < \gamma_d\}$$

satisfies

$$(31) \quad d_H(S_d^\lambda, \text{supp}(\xi_t^\lambda)) \leq \epsilon,$$

as  $d \rightarrow +\infty$ .

## 6. ILLUSTRATIVE EXAMPLE

In this section, we give an example that illustrates the computation of the moments associated with  $\xi_t^\lambda$  of the regularized system (9) in the case where  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , by applying the moment-SOS hierarchy [20].

Consider the constrained system (7) of Example 1 where  $f(z) = Az$  with  $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$  and  $S = \mathbb{R}_+^2$ . Let us write the regularized system (9) in polar coordinates  $(r, \theta)$  as follows:

$$\begin{cases} \dot{r}(t) = 0, \\ \dot{\theta}_\lambda(t) = -1 - \frac{1}{\lambda}(\theta_\lambda(t) - \text{proj}(\theta_\lambda(t), S(t))). \end{cases}$$

or equivalently:

$$(32) \quad \begin{cases} \dot{r}(t) = 0, \\ \dot{\theta}_\lambda(t) = -1 - \frac{1}{\lambda}(\theta_\lambda(t) - \max(\theta_\lambda(t), 0)). \end{cases}$$

Let  $d = 4$  be the degree of relaxation, and let us choose different values of the regularization parameter  $\lambda \in \{0.05, 0.1, 0.5\}$ . We introduce the initial measure as a Dirac measure with respect to time product a uniform measure in  $[0, 1] \times [0, \frac{1}{2}]$  with respect to the state.

We calculate the moment of the initial measure to replace it directly in Liouville constraint (25), where the variables  $z_1$  and  $z_2$  in (25) are respectively  $r$  and  $\theta$ . For

all  $(a, b_1, b_2) \in \mathbb{N}^3$ , with  $a + b_1 + b_2 \leq d$ , the moment of the initial measure is then given as

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^n} t^a z_1^{b_1} z_2^{b_2} d\mu_0(t, z) &= \int_0^T \int_{\mathbb{R}^n} t^a z_1^{b_1} z_2^{b_2} \delta_0(dt) \lambda_{[0,1]}(dz_1) \lambda_{[0, \frac{1}{2}]}(dz_2) \\ &= 0^a \int_0^1 z_1^{b_1} dz_1 \int_0^{\frac{1}{2}} 2z_2^{b_2} dz_2 \\ &= 0^a \frac{1}{b_1 + 1} (1^{b_1+1} - 0^{b_1+1}) \frac{2}{b_2 + 1} \left( \left( \frac{1}{2} \right)^{b_2+1} - 0^{b_2+1} \right). \end{aligned}$$

Then we apply the moment-SOS hierarchy [20] which allows us to approximate numerically the moments of the unknown occupation measure and terminal measure. For different values of the terminal time  $T \in \{0, 0.1, 0.2, \dots, 1\}$ , this gives us:

- The evolution of the moment  $\int r(t)^2 d\mu_T^\lambda$  as a function of time, which we observe numerically is a constant for different values of the regularization parameter  $\lambda$ .
- The evolution of the moment  $\int \theta(t)^2 d\mu_T^\lambda$  as a function of time for different values of the regularization parameter  $\lambda$ , which is illustrated on Figure 3.

## 7. CONCLUSIONS

In this article, we studied the time evolution of nonsmooth constrained dynamical systems when the initial condition is described by a probability measure. Unlike conventional ODEs we do not know how to describe the time evolution of the image measure by the flow as a Liouville PDE. To circumvent this issue, we propose an approximation technique based on constructing Lipschitz approximations for the original nonsmooth system, and then using the Liouville equation for the approximate Lipschitz dynamics. Numerical methods for computing the approximation of solutions of Liouville equation then allow us to compute the moments and support of the probability measures associated to the original system.

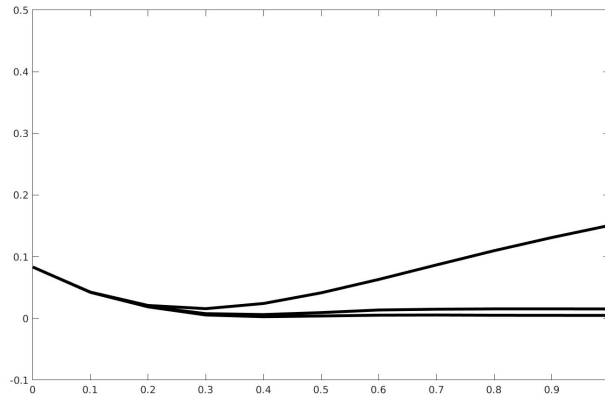


FIGURE 3. First order moment of the second state (vertical axis) of the occupation measure of the regularized system, as a function of time (horizontal axis), for different values of the regularization parameter (top curve  $\lambda = 0.5$ , middle curve  $\lambda = 0.1$ , bottom curve  $\lambda = 0.05$ )

To seek improvements in the approach adopted in this paper, it is observed that the proposed Lipschitz approximations are difficult to simulate numerically. In particular, for the illustrated example, we implemented the projection map onto a cone by splitting the Liouville equation in different parts, where each of them corresponds to the region where the approximating ODE is continuous. One could use some recent work on approximating ODEs with twice differentiable right-hand side [14] to see if the resulting implementation is easier to simulate for a broader class of constraint sets.

Another potential direction of research that comes out from this work is the possibility of using the proposed tools for optimal control problems. As was done for ODEs [24, 19], it is possible to use the formalism of Liouville equation for optimal control problems. The optimal control for the class of nonsmooth systems studied in this paper is a challenging problem, and it has been addressed recently in [11, 13, 32]. It would be interesting to see if the methods proposed in this paper provide a numerically constructive solution to such challenging problems.

## APPENDIX A. PROOFS OF LEMMAS IN THEOREM 2

A.1. **Proof of Lemma 3.** For each  $\lambda > 0$ , the dynamics for  $z_\lambda$  in (9) yield

$$(33) \quad \begin{aligned} |\dot{z}_\lambda(t)| &= |f(t, z_\lambda(t)) - \frac{1}{\lambda}(z_\lambda(t) - \text{proj}(z_\lambda(t), S(t)))| \\ &\leq |f(t, z_\lambda(t))| + \frac{1}{\lambda}|z_\lambda(t) - \text{proj}(z_\lambda(t), S(t))|. \end{aligned}$$

For the first term in the right-hand side of (33), we have that

$$(34) \quad |f(t, z_\lambda(t))| \leq L_f(1 + |z_\lambda(t)|).$$

For the second term in the right-hand side of (33), we introduce the function  $d_\lambda(t) = \inf_{y \in S(t)} |y - z_\lambda(t)|$ , so that  $d_\lambda(t) = d_{S(t)}(z_\lambda(t))$ . It is seen that  $d_\lambda(t) = |z_\lambda(t) - \text{proj}(z_\lambda(t), S(t))|$ . So  $\frac{1}{\lambda}|z_\lambda(t) - \text{proj}(z_\lambda(t), S(t))| = \frac{1}{\lambda}d_\lambda(t)$ . To obtain a bound on  $d_\lambda(t)$ , we compute the derivative of  $d_\lambda^2(t)$ :

$$(35) \quad \begin{aligned} \frac{d}{dt}d_\lambda^2(t) &= \frac{d}{dt}d_{S(t)}^2(z_\lambda(t)) \\ &= \lim_{\epsilon \rightarrow 0} \frac{d_{S(t+\epsilon)}^2(z_\lambda(t+\epsilon)) - d_{S(t)}^2(z_\lambda(t))}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{d_{S(t+\epsilon)}^2(z_\lambda(t+\epsilon)) - d_{S(t)}^2(z_\lambda(t+\epsilon))}{\epsilon} + \frac{d_{S(t)}^2(z_\lambda(t+\epsilon)) - d_{S(t)}^2(z_\lambda(t))}{\epsilon}. \end{aligned}$$

For the first term in the limit, we use that

$$(36) \quad \begin{aligned} &d_{S(t+\epsilon)}^2(z_\lambda(t+\epsilon)) - d_{S(t)}^2(z_\lambda(t+\epsilon)) \\ &\leq d_H(S(t+\epsilon), S(t))(d_{S(t+\epsilon)}(z_\lambda(t+\epsilon)) + d_{S(t)}(z_\lambda(t+\epsilon))) \\ &\leq |\epsilon|L_S(d_{S(t+\epsilon)}(z_\lambda(t+\epsilon)) + d_{S(t)}(z_\lambda(t+\epsilon))). \end{aligned}$$

For the second term in the limit, we first notice that

$$\begin{aligned} d_{S(t)}^2(z_\lambda(t+\epsilon)) - d_{S(t)}^2(z_\lambda(t)) &= d_{S(t)}^2(z_\lambda(t) + \epsilon \dot{z}_\lambda(t)) - d_{S(t)}^2(z_\lambda(t)) \\ &\quad + (d_{S(t)}(z_\lambda(t+\epsilon)) - d_{S(t)}(z_\lambda(t) + \epsilon \dot{z}_\lambda(t))) \\ &\quad (d_{S(t)}(z_\lambda(t+\epsilon)) + d_{S(t)}(z_\lambda(t) + \epsilon \dot{z}_\lambda(t))). \end{aligned}$$

Since  $z_\lambda(\cdot)$  is differentiable,  $z_\lambda(t+\epsilon) = z_\lambda(t) + \epsilon \dot{z}_\lambda(t) + \mathcal{O}(\epsilon)$  and hence  $d_{S(t)}(z_\lambda(t+\epsilon)) - d_{S(t)}(z_\lambda(t) + \epsilon \dot{z}_\lambda(t)) = \mathcal{O}(\epsilon)$ . This implies that

$$d_{S(t)}^2(z_\lambda(t+\epsilon)) - d_{S(t)}^2(z_\lambda(t)) = d_{S(t)}^2(z_\lambda(t) + \epsilon \dot{z}_\lambda(t)) - d_{S(t)}^2(z_\lambda(t)).$$



And,

$$(37) \quad \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left[ d_{S(t)}^2(z_\lambda(t+\epsilon)) - d_{S(t)}^2(z_\lambda(t)) \right] = \langle \nabla d_{S(t)}^2(z_\lambda(t)), \dot{z}_\lambda(t) \rangle \\ = 2\langle z_\lambda(t) - \text{proj}(z_\lambda(t), S(t)), \dot{z}_\lambda(t) \rangle.$$

By substitution of (36) and (37) in equation (35), we obtain

$$\begin{aligned} \frac{d}{dt} d_\lambda^2(t) &= 2d_\lambda(t) \dot{d}_\lambda(t) \leq 2d_\lambda(t) \dot{z}_\lambda(t) + 2L_S d_\lambda(t) \\ &\leq 2d_\lambda(t) \left( f(t, z_\lambda(t)) - \frac{1}{\lambda} d_\lambda(t) \right) + 2L_S d_\lambda(t) \\ &\leq -\frac{2}{\lambda} d_\lambda^2(t) + 2d_\lambda(t) f(t, z_\lambda(t)) + 2L_S d_\lambda(t). \end{aligned}$$

Dividing by  $2d_\lambda(t)$ , we get

$$\frac{d}{dt} d_\lambda(t) \leq -\frac{1}{\lambda} d_\lambda(t) + f(t, z_\lambda(t)) + L_S,$$

which implies that,

$$d_\lambda(t) \leq e^{-t/\lambda} d_\lambda(0) + \int_0^t e^{-(t-s)/\lambda} (f(s, z_\lambda(s)) + L_S) ds.$$

Or,  $d_\lambda(0) = |z_0 - \text{proj}(z_0, S(0))| = 0$  since  $z_0 \in S(0)$  and we have that  $f$  satisfies (34), then it follows

$$(38) \quad \frac{1}{\lambda} d_\lambda(t) \leq \frac{1}{\lambda} \int_0^t e^{-(t-s)/\lambda} (L_f + L_f |z_\lambda(s)| + L_S) ds.$$

And therefore, substituting (34) and (38) in (33), we get

$$\begin{aligned} |\dot{z}_\lambda(t)| &\leq L_f + L_f |z_\lambda(t)| + \frac{1}{\lambda} \int_0^t e^{-(t-s)/\lambda} (L_f + L_f |z_\lambda(s)| + L_S) ds \\ &\leq L_f + L_f |z_\lambda(t)| + \frac{L_f}{\lambda} \int_0^t e^{-(t-s)/\lambda} ds + \frac{L_f}{\lambda} \int_0^t e^{-(t-s)/\lambda} |z_\lambda(s)| ds \\ &\quad + \frac{L_S}{\lambda} \int_0^t e^{-(t-s)/\lambda} ds. \end{aligned}$$

We have

$$\begin{aligned} \frac{L_f}{\lambda} \int_0^t e^{-(t-s)/\lambda} ds &= \frac{L_f}{\lambda} e^{-t/\lambda} \left[ \lambda e^{s/\lambda} \right]_0^t = \frac{L_f}{\lambda} e^{-t/\lambda} (\lambda e^{t/\lambda} - \lambda) \\ &= L_f (1 - e^{-t/\lambda}) \leq L_f. \end{aligned}$$

Similarly,

$$\frac{L_S}{\lambda} \int_0^t e^{-(t-s)/\lambda} ds \leq L_S.$$

Besides, we have

$$\begin{aligned} \frac{L_f}{\lambda} \int_0^t e^{-(t-s)/\lambda} |z_\lambda(s)| ds &\leq \underbrace{\frac{L_f}{\lambda} \int_0^t e^{-(t-s)/\lambda} ds}_{\leq L_f} \cdot \max_{0 \leq s \leq t} |z_\lambda(s)| \\ &\leq L_f \max_{0 \leq s \leq t} |z_\lambda(s)|. \end{aligned}$$

The bound of  $|\dot{z}_\lambda(t)|$  is then expressed as

$$|\dot{z}_\lambda(t)| \leq 2L_f + L_f |z_\lambda(t)| + L_f \max_{0 \leq s \leq t} |z_\lambda(s)| + L_S.$$

**A.2. Proof of Lemma 5.** Let  $\bar{z}_\lambda(s) := \text{proj}(z_\lambda(s), S(s))$ ; then  $s \mapsto \bar{z}_\lambda(s)$  is a continuous mapping. Since  $\varphi(s) + \int_{t_1}^s f(r, z_\lambda(r)) dr \in S(s)$  and  $\lambda$  is positive, it follows from the definition of the projections that

$$\begin{aligned} & \langle \varphi(s) + \int_{t_1}^s f(r, z_\lambda(r)) dr - \bar{z}_\lambda(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle \\ &= \frac{1}{\lambda} \langle \varphi(s) + \int_{t_1}^s f(r, z_\lambda(r)) dr - \bar{z}_\lambda(s), \bar{z}_\lambda(s) - z_\lambda(s) \rangle \geq 0. \end{aligned}$$

Then

$$\langle \varphi(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle \geq \langle \bar{z}_\lambda(s) - \int_{t_1}^s f(r, z_\lambda(r)) dr, \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle,$$

which implies that,

$$\int_{t_1}^{t_2} \langle \varphi(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle ds \geq \int_{t_1}^{t_2} \langle \bar{z}_\lambda(s) - \int_{t_1}^s f(r, z_\lambda(r)) dr, \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle ds.$$

Since at the points where  $z_\lambda(\cdot)$  is differentiable, we have

$$\begin{aligned} \langle \bar{z}_\lambda(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle &= \langle \bar{z}_\lambda(s) - z_\lambda(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle \\ &\quad + \langle z_\lambda(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle \\ &= \underbrace{\frac{1}{\lambda} |\bar{z}_\lambda(s) - z_\lambda(s)|^2}_{\geq 0} + \langle z_\lambda(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle, \end{aligned}$$

it follows that,

$$\langle \bar{z}_\lambda(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle \geq \langle z_\lambda(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle,$$

and,

$$\int_{t_1}^{t_2} \langle \varphi(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle ds \geq \int_{t_1}^{t_2} \langle z_\lambda(s) - \int_{t_1}^s f(r, z_\lambda(r)) dr, \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle ds.$$

We have

$$\begin{aligned} & \int_{t_1}^{t_2} \langle z_\lambda(s) - \int_{t_1}^s f(r, z_\lambda(r)) dr, \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle ds \\ &= \frac{1}{2} \left[ \|z_\lambda(s) - \int_{t_1}^s f(r, z_\lambda(r)) dr\|^2 \right]_{t_1}^{t_2} \\ &= \frac{1}{2} \left( \|z_\lambda(t_2) - \int_{t_1}^{t_2} f(r, z_\lambda(r)) dr\|^2 - \|z_\lambda(t_1)\|^2 \right), \end{aligned}$$

hence, we obtain that

$$\int_{t_1}^{t_2} \langle \varphi(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle ds \geq \frac{1}{2} \left( \|z_\lambda(t_2) - \int_{t_1}^{t_2} f(r, z_\lambda(r)) dr\|^2 - \|z_\lambda(t_1)\|^2 \right).$$

We take limits with respect to  $\lambda \rightarrow 0$ . Since  $z_\lambda(\cdot)$  converges pointwise to  $z(\cdot)$ , we have  $\langle \varphi(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s)) \rangle \rightarrow \langle \varphi(s), \dot{z}(s) - f(s, z(s)) \rangle$  for each  $s \in [t_1, t_2]$ , and  $\|z_\lambda(t_2) - \int_{t_1}^{t_2} f(r, z_\lambda(r)) dr\|^2 \rightarrow \|z(t_2) - \int_{t_1}^{t_2} f(r, z(r)) dr\|^2$ , and  $\|z_\lambda(t_1)\|^2 \rightarrow \|z(t_1)\|^2$ .

Therefore, this yields to

$$\int_{t_1}^{t_2} \langle \varphi(s), \dot{z}(s) - f(s, z(s)) \rangle ds \geq \frac{1}{2} \left( \|z(t_2) - \int_{t_1}^{t_2} f(r, z(r)) dr\|^2 - \|z(t_1)\|^2 \right),$$

and Lemma 5 is then proved.

## REFERENCES

- [1] S. Adly. *A variational approach to nonsmooth dynamics*. Springer Briefs in Mathematics. Springer, 2020.
- [2] L. Ambrosio. Transport equation and Cauchy problem for non-smooth vector fields. *Lecture Notes in Mathematics*, Springer, 2005.
- [3] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*, Birkhäuser, Boston Basel Berlin, 1990.
- [4] J. Bartsch, A. Borzi, F. Fanelli and S. Roy. A theoretical investigation of Brockett’s ensemble optimal control problems. *Calculus of Variations and Partial Differential Equations*, 58(162), 2019.
- [5] B. Bonnet and H. Frankowska. Differential inclusions in Wasserstein spaces: The Cauchy-Lipschitz framework. *Journal of Differential Equations*, 271:594–637, 2021.
- [6] F. Bouchut and F. James. One dimensional transport equation with discontinuous coefficients. *Nonlinear Analysis*, 32:891–933, 1998.
- [7] F. Bouchut, F. James and S. Mancini. Uniqueness and weak stability for multi-dimensional transport equations with one-sided Lipschitz coefficients. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze*, Série 5, 4(1):1–25, 2005.
- [8] R. W. Brockett. Optimal control of the Liouville equation. *AMS/IP Studies in Advanced Mathematics*, 39:23–35, 2007.
- [9] R. W. Brockett. Notes on the control of the Liouville equation. *In Control of Partial Differential Equations. Lecture Notes in Mathematics*, Springer, Heidelberg, 2048:101–129, 2012.
- [10] B. Brogliato and A. Tanwani. Dynamical systems coupled with monotone set-valued operators: Formalisms, applications, well-posedness, and stability. *SIAM Review*, 62(1):3–129, 2020.
- [11] T.H. Cao, G. Colombo, B.S. Mordukhovich and D. Nguyen. Optimization and discrete approximation of sweeping processes with controlled moving sets and perturbations. *Journal of Differential Equations*, 274(461-509), 2021.
- [12] G. Carlier. Lecture Notes on Optimal Transportation and Economic Applications. *IMA Workshop on New Mathematical Models in Economics and Finance*, June 2010.
- [13] G. Colombo, R. Henrion, N.D. Hoang and B.S. Mordukhovich. Optimal control of the sweeping process over polyhedral controlled sets. *J. Differential Equations*, 260: 3397–3447, 2016.
- [14] G. Colombo and C. Kozaily. Existence and uniqueness of solutions for an integral perturbation of Moreau’s sweeping process. *Journal of Convex Analysis*, 27(1):227–236, 2020.
- [15] R. J. Di Perna and P.L. Lions. Ordinary differential equations, transport theory and Sobolev spaces. *Invent. Math.*, 98: 511–547, 1989.
- [16] J.-F. Edmond and L. Thibault. *BV solutions of nonconvex sweeping process differential inclusion with perturbation*, *Journal of Differential Equations*, 226:135–179, 2006.
- [17] L.C. Evans. *Partial differential equations. 2nd edition*, Amer. Math. Society, Providence, NJ, 2010.
- [18] D. Henrion. Optimization on linear matrix inequalities for polynomial systems control. *Les cours du C.I.R.M.*, 3(1):1–44, 2013.
- [19] D. Henrion and E. Pauwels. Linear conic optimization for nonlinear optimal control *In S. Ahmed, M. Anjos, T. Terlaky (Editors). Advances and Trends in Optimization with Engineering Applications*, SIAM, Chap. 10, pp. 121-134, 2017.
- [20] D. Henrion, M. Korda and J. B. Lasserre. *The Moment-SOS Hierarchy*. World Scientific, 2020.
- [21] D. Henrion, J.B. Lasserre and J. Löfberg. GloptiPoly 3: moments, optimization and semidefinite programming. *Optimization Methods & Software*, 24(4-5): 761-779, 2009.
- [22] M. Kunze and M.D.P. Monteiro Marques. Yosida-Moreau regularization of sweeping processes with unbounded variation. *J. Differential Equations*, 130:292–306, 1996.
- [23] J.B. Lasserre and E. Pauwels. The empirical Christoffel function with applications in data analysis. *Advances in Computational Mathematics*, 45(4): 1439-1468, 2019.
- [24] J. B. Lasserre, D. Henrion, C. Prieur and E. Trélat. Nonlinear optimal control via occupation measures and LMI relaxations. *SIAM Journal on Control and Optimization*, 47(4):1643-1666, 2008.
- [25] D.G. Luenberger. *Optimization by vector space methods*. John Wiley & Sons, 1969.
- [26] M.-D.-P. Monteiro-Marques. *Differential Inclusions in Nonsmooth Mechanical Problems. Shocks and Dry Friction*, vol. 9, Birkhäuser, Basel, 1993.
- [27] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation. Part I: Basic Theory*, Springer, Berlin, 2006.
- [28] J.-J. Moreau. Evolution problem associated with a moving convex set in a Hilbert space. *Journal of Differential Equations*, 26:347–374, 1977. Preliminary version in: *Problème d’évolution*

- associé à un convexe mobile d'un espace Hilbertien. *Comptes Rendus de l'Académie des Sciences. Série Iib, Mécanique*, Elsevier, 276: 791-794, 1973.
- [29] F. Nacry and L. Thibault. Regularization of sweeping process: old and new. *Pure Applied Functional Analysis*, 4: 59–117, 2019.
- [30] R.-T. Rockafellar and R.-J.-B. Wets. *Variational Analysis*, vol. 317 of Comprehensive Studies in Mathematics, Springer Verlag, Berlin Heidelberg, 1998.
- [31] M. Souaiby, A. Tanwani and D. Henion. Cone-copositive Lyapunov functions for complementarity systems: Converse result and polynomial approximation. *IEEE Transactions on Automatic Control*, in press, 2021. DOI: 10.1109/TAC.2021.3061557
- [32] A. Vieira, B. Brogliato and C. Prieur. Quadratic optimal control of linear complementarity systems: First-order necessary conditions and numerical analysis. *IEEE Transactions on Automatic Control*, 65(6): 2743-2750, 2020.
- [33] C. Villani. Topics in optimal transportation. *American Mathematical Society, Providence*, 2003.