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 two academic examples.

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. [Vil03, 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

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

numerical optimal control [LHPT08] as well as in several control-theoretic problems [BBFR19, Bro07, Bro12].

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 [DL89]. Continuity equations corresponding to one-sided Lipschitz vector fields have been studied in [BJ98, BJM05]. In [Amb08], 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 $f : \mathbb{R}^n \to \mathbb{R}^n$, we describe the evolution of constrained systems via the differential inclusion

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

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

$$\langle \dot{x}(t) - f(x(t)), y - x(t) \rangle \geqslant 0,$$

for all $y \in S$, $x(t) \in S$, $t \in [0, T]$. 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 [BT20], and a research monograph [Adl18], 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 [AF90, Mor06, RW98]. Results based on stability analysis with computational aspects have been addressed recently by the authors in [STH22].

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 [Mor77], see [ET06] 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 received very little attention in the literature. One can study such problems by considering stochastic versions of (1) by adding a diffusion term on the right-hand side. Such systems first came up in the study of variational inequalities arising in stochastic control [BL78], and in the literature, we can find results on existence and uniqueness of solutions in appropriate function space. In [C95́], this is done by considering Yosida approximations of the maximal monotone operator, whereas [Ber03] provides a proof based on time-discretization of system (1). These approaches have been generalized for prox-regular set S in [BV11], and the case where the drift term contains Young measures [CMMdF14, CMMdF16]. One could also, in principle, formulate a partial differential equation with set-valued elements and study the solutions of such equations under appropriate hypothesis, which is the case in [BF21] but it is not clear how to derive the corresponding set-valued partial differential equation for system (1) and whether the resulting inclusion would satisfy the necessary hypothesis for well-posedness. Different from these approaches, and inspired by the fact that the evolution of a probability measure for single-valued dynamical system is described by Liouville equation, it is natural to ask whether the evolution of a probability measure under the dynamics of system (1) can be studied using Liouville equation. To the best of authors' knowledge, this approach has only been adopted in [DMS16], where the authors consider system of form (1) without the drift term $f(\cdot)$. Since the right-hand side of (1) is set-valued, it is not immediately clear how the divergence term in the Liouville equation is to be interpreted. In [DMS16], the authors consider approximations to the solutions of Liouville equation associated with (1), which are similar to time-stepping algorithm. That is, a time-discretization technique is introduced which is based on projecting the density function on to the constraint set with respect to Wasserstein metric.

In this paper, we consider a different route for computing the approximate solution of system (1) in the space of probability measures. Inspired by the concepts presented in [Amb08], 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. We can then associate a single-valued Liouville equation to each of these approximants, and establish convergence of the resulting sequence of measures. Unlike [DMS16], our approach for numerically solving the Liouville equation does not depend upon discretization in time, or space for that matter. Instead, we use functional discretization: we choose a family of test functions (the monomials) on which the evolution measure and the associated moments are then approximated numerically by a hierarchy of semidefinite programs. 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. One possibility – that we do not explore here – could to use Monte-Carlo probabilistic algorithms. Instead, we investigate a purely deterministic approach: in order to get a quantitative measure of the distribution of state at any time instant, which involves building a hierarchy of moments defined by the action of a finite Borel measure on polynomial test functions, and encoding the positivity constraints on moment matrix by using sum-of-squares (SOS) decomposition. This technique, called moment-SOS hierarchy [HKL20] has been used in several engineering problems, and for our purposes, it 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 [LP19], 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 two academic examples 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.

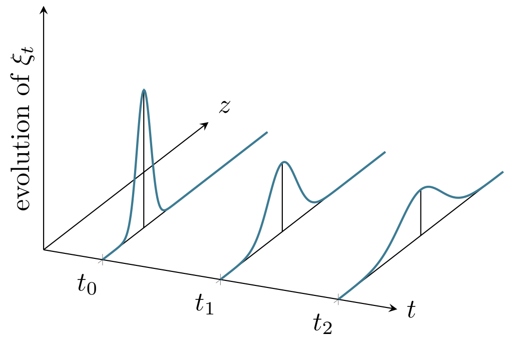


FIGURE 1. Evolution of probability measure ξ_t w.r.t. time and space.

2. Preliminaries and Problem formulation

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

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

over a given time interval $[0, T]$, where $g : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ 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 $G_t : \mathbb{R}^n \to \mathbb{R}^n$, 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

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

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 $t \geqslant 0$, denoted by $\xi_t \in \mathcal{P}(\mathbb{R}^n)$, we consider the pushforward or image measure of ξ_0 through the flow map $G_t(\cdot)$. That is, let

$$\xi_t := G_t \sharp \xi_0, \tag{4}$$

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 ξ_t is described by the following PDE, called the continuity or *Liouville* equation:

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

with the initial condition:

$$\xi|_{t=0} = \xi_0. \tag{6}$$

The Liouville equation (5) should be understood in the sense of distributions, i.e. the family of probability measures $t \mapsto \rho_t$ is a measure-valued solution of (5)-(6) if :

- it is continuous in the sense that for every compactly supported continuous function $\phi : \mathbb{R}^n \to \mathbb{R}$, the map $M_\phi : t \mapsto \int_{\mathbb{R}^n} \phi \, \mathrm{d}\xi_t$ is continuous on $[0, \infty)$ with $M_\phi(0) = \int_{\mathbb{R}^n} \phi \, \mathrm{d}\xi_0$.

- for every $r > 0$ and every $v \in \mathcal{C}^1([0,T] \times \mathbb{R}^n)$ such that $v(T,\cdot) = 0$ and $v(t,\cdot)$ is supported on a closed ball of radius r for every $t \in [0,T]$, one has
$$\int_0^T \int_{\mathbb{R}^n} (\partial_t v(t,z) + \partial_z v(t,z) \cdot g(t,z)) \, \mathrm{d}\xi_t(z) \, \mathrm{d}t = -\int_{\mathbb{R}^n} v(0,z) \, \mathrm{d}\xi_0(z).$$

The equivalence between the solutions of ODE (2) and PDE (5), is established in the following result, see e.g. [Vil03, Theorem 5.34]:

Theorem 2.1. *For each $t \in [0,T]$, let $G_t : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism so that (3) holds. Given $\xi_0 \in \mathcal{P}(\mathbb{R}^n)$, let ξ_t be defined as in (4). Then, ξ_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 ξ_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 2.1, 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

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

over an interval $[0,T]$ for some given $T > 0$, where $f : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ 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

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

If we consider a point $z \in \mathrm{int}(S)$, 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. The foregoing discussion motivates us to consider the following definition of a solution to (7) originating from a point mass in $S(0)$: An absolutely continuous function $z : [0,T] \to \mathbb{R}^n$ is called a solution to system (7) if there exists a selection $\eta(t) \in \mathcal{N}_{S(t)}(z(t))$ such that
$$\dot{z}(t) = f(t,z(t)) - \eta(t),$$
holds Lebesgue a.e. on $[0,T]$, and for each $t \in [0,T]$, we have $z(t) \in S(t)$. The reader may refer to [BT20] for different formalisms and methods for describing the selection rule η.

For this paper, 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 $F_t : S(t) \to S(t)$, 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. $z(t) \sim$

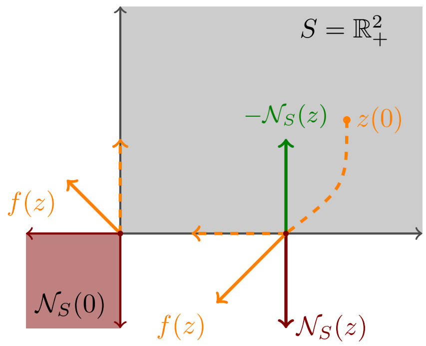


FIGURE 2. State trajectories in constrained system with $S = \mathbb{R}^2_+$.

$\xi_t \in \mathcal{P}(S(t))$ defined by $\xi_t := F_t \sharp \xi_0$. 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 $A = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ and $S = \mathbb{R}^2_+$ and let z_0 be a given initial condition, with angle θ_0. For $t \leqslant \theta_0$, we have $z(t) = F_t(z_0) = e^{At} z_0 = \left[\begin{smallmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{smallmatrix}\right] z_0$. And for $t \geqslant \theta_0$, 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 $t \geqslant \theta_0$, 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

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

Indeed, as we can observe, the flow map is not invertible since given a state $z(t)$ for a given time $t \geqslant \theta_0$, 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 2.1 are not satisfied in general for such systems. On the other hand, for each $t \in [0, T]$, the forward flow map F_t is well-defined and therefore the solution $\xi_t := F_t \sharp \xi_0$ exists and is uniquely defined. However, it is not possible to write down the evolution equation for ξ_t, like Liouville equation for smooth ODEs, due to nonsmooth set-valued dynamics in (7). Recent literature in this direction deals with such problems, either by studying partial differential equations with set-valued mappings [BF21] or by introducing an approximation based on time discretization [DMS16]. In this article, our goal is to find alternate methods based on functional discretization with monomial basis to approximate the measure ξ_t and propose computational algorithms to calculate such approximations numerically.

2.3. **Problem Formulation.** We consider the dynamical system (7) with flow map $F_t : \mathbb{R}^n \to \mathbb{R}^n$. For a given $\xi_0 \in \mathcal{P}(S(0))$, since there is no direct derivation of the PDE for characterizing the evolution of $\xi_t := F_t \sharp \xi_0$, we compute an approximation of ξ_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 $\xi_t^\lambda := F_t^\lambda \sharp \xi_0$.
- When λ tends to 0, prove that ξ_t^λ converges to $\xi_t := F_t \sharp \xi_0$ in the weak-star topology. In particular, all finite order moments of ξ_t^λ converge to the moments of ξ_t.
- When λ tends to 0, prove the convergence of the support of ξ_t^λ to the support of ξ_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 ξ_t^λ and the support of ξ_t^λ. This allows us to compute an approximation of the moments and support of ξ_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 [BT20]. 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 $\lambda > 0$:

$$\begin{aligned} \dot{z}_\lambda(t) &= f(t, z_\lambda(t)) - \frac{1}{\lambda}(z_\lambda(t) - \operatorname{proj}(z_\lambda(t), S(t))), \\ z_\lambda(0) &= z_0 \in S(0) \end{aligned} \tag{9}$$

over the interval $[0, T]$, where $\operatorname{proj}(z, S)$ 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 $z_\lambda : [0, T] \to \mathbb{R}^n$ 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 $t \in [0, T]$,

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

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 $L_S \geqslant 0$, such that

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

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

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

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

Theorem 3.1. *Under Assumptions 1–2, consider the sequence of solutions $\{z_\lambda\}_{\lambda>0}$ to parameterized ODE (9) on an interval $[0, T]$. Then, as $\lambda \to 0$, the sequence converges uniformly to a Lipschitz continuous function $z : [0, T] \to \mathbb{R}^n$, the unique solution to the differential inclusion (7).*

The proof of this theorem is described in the remainder of this section. Note that the right-hand side of (7) is typically seen as a Lipschitz perturbation of a time-varying maximal monotone operator and the existence of solutions for such systems is proved by constructing a sequence of solutions and studying their convergence. Such sequences are either obtained by time-discretization or Moreau-Yosida regularization. The simplest case of static maximal monotone operators is studied in [Bre73] but the situation is more complex when the domain of the multivalued mapping is time-dependent. Sweeping processes (defined by taking $f \equiv 0$ in (7) and deterministic initial condition) form a particular case of such systems and for the most part, the existence of solutions is proved using a discretization algorithm originating from [Mor77]; see [CIT22] for some recent developments. Our contribution in Theorem 3.1 provides a proof based on Moreau-Yosida regularization for differential inclusions with normal cones associated to time-varying sets and Lipschitz perturbation, which does not appear in the literature. One is obviously helped by the machinery developed for the proofs with no perturbations but the intermediate calculations are different and we find it instructive to provide a proof for self-contained exposition. For the proof that follows, certain calculations, leading to the intermediate lemmas used in the proof, have been included in the appendix.

Proof of Theorem 3.1. 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 $z_\lambda(.)$, let us begin by computing bounds on the norm of $\dot{z}_\lambda(\cdot)$ as stated in the following lemma, whose proof is given in A.

Lemma 3.2. *For each $\lambda > 0$, it holds*

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

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

Based on Lemma 3.2, let us now calculate $\frac{d}{dt}|z_\lambda(t)|^2$ for getting an estimate on $|z_\lambda(\cdot)|$. First, we observe that

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

Substituting (11) in (12) yields

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

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

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

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

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

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, $|z_\lambda(t)|^2 = y_\lambda(t) \leqslant \widehat{y}_\lambda(t) = v(t)^2$, and we obtain

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

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

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 3.3. *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 3.3 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 3.1, 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 A.

Lemma 3.4. *There exists a continuous function $\varphi : [0, T] \to \mathbb{R}^n$ that satisfies $\varphi(t_1) \in S(t_1)$ and $\varphi(s) + \int_{t_1}^{s} f(r, z_\lambda(r))\, dr \in S(s)$ for each $s \in [t_1, t_2]$, with $t_1, t_2 \in [0, T]$. Moreover, it holds that,*

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

We now complete the proof of Theorem 3.1 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, $\langle \xi - z(t), \dot{z}(t) - f(t, z(t))\rangle \geqslant 0$, for any $\xi \in S(t)$ and for almost every $t \geqslant 0$. This is indeed the case, since for every $\xi \in S(t)$, we can take a Lipschitz continuous function $\varphi : [t, T] \to \mathbb{R}^n$ with $\varphi(t) = \xi$ such that, due to Lemma 3.4, we get

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

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

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

which implies

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

From this, we get

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

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 ϵ, we get

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

for some constant $M>0$. Letting ϵ tend to zero, and recalling that $\xi = \varphi(t)\in S(t)$, we get

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

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

Remark 3.5. 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 [BT20, KM96, NT19]. The proof technique adopted here closely follows the outline given in [BT20], 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 $F_t^\lambda : \mathbb{R}^n \to \mathbb{R}^n$, 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

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

for each $t\in[0,T]$ and $\lambda>0$. From Theorem 2.1, it follows that ξ_t^λ satisfies the partial differential equation:

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

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

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

On the other hand, we do not know how to derive a meaningful PDE for ξ_t. However, in the sequel, we show that the probability measure ξ_t can be approximated by ξ_t^λ as $\lambda \to 0$. This way, a good numerical approximation of ξ_t^λ would also provide an approximation of ξ_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 ξ_t using the moments of ξ_t^λ. Given a measure ξ, we denote its support by $\operatorname{supp}(\xi)$, defined as the smallest closed set whose complement has zero measure with respect to ξ. Equivalently, it is the smallest closed set for which every point has a neighborhood of positive measure with respect to ξ.

Proposition 4.1. *Let $v : \mathbb{R}^n \to \mathbb{R}$ be a continuous function, and assume that ξ_0 has bounded support. Then,*

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

Proof. By definition of the pushforward measure ξ_t^λ, it holds

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

for all continuous functions v. From Theorem 3.1, for each $t \in [0, T]$, we have $\lim_{\lambda\to 0} z_\lambda(t) = z(t)$, which is equivalent to

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

Since v is any continuous function, this implies

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

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

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

Hence, (19) and (20) yield

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

Using again the change of variables formula, we obtain

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

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

Remark 4.2. In the proof of Proposition 4.1, the boundedness of $\operatorname{supp}(\xi_0)$ was used to invoke dominated convergence theorem. The result of Proposition 4.1 extends in some cases where $\operatorname{supp}(\xi_0)$ is unbounded. In particular, if it can be shown that there exists a function $g : [0, T] \times \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$ such that, for each $\lambda > 0$,

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

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 4.1 is that we can get a desired approximation of the moments of ξ_t by choosing appropriate test functions v. This amounts to computing the moments of ξ_t^λ. 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 recall the notion of *occupation measure* associated with the trajectories of a nonlinear ODE (2). In the following, we denote the indicator function of a set B by $\mathcal{I}_B(\cdot)$, that is, $\mathcal{I}_B(z) = 1$ when $z \in B$ and $\mathcal{I}_B(z) = 0$ when $z \notin B$.

Definition 1. Given an initial condition $z_0 \in \mathbb{R}^n$, the occupation measure of a trajectory $G_t(z_0)$ is defined by

$$\mu(A \times B | z_0) := \int_A \mathcal{I}_B(G_t(z_0))\, \mathrm{d}t$$

for every A, respectively B, contained in the Borel σ-algebra of $[0, T]$, respectively $\mathbb{R}^n$.

A geometric interpretation is that μ measures the time spent by the graph of the trajectory $(t, G_t(z_0))$ in a given subset $A \times B$ of $[0, T] \times \mathbb{R}^n$. An analytic interpretation is that integration with respect to μ is equivalent to time-integration along a system trajectory, that is,

$$\int_{[0,T]} v(t, G_t(z_0))\, \mathrm{d}t = \int_{[0,T]} \int_{\mathbb{R}^n} v(t, z)\, \mu(\mathrm{d}t, \mathrm{d}z | z_0)$$

for every test function $v \in \mathcal{C}([0, T] \times \mathbb{R}^n)$. As a consequence of this last interpretation, we observe that the Liouville equation (16) can be equivalently written as a linear PDE satisfied by the occupation measures

$$d\mu^\lambda := dt\; d\xi_t^\lambda, \quad \textit{with} \quad \mu_0^\lambda := \delta_0 \xi_0,\ \mu_T^\lambda := \delta_T \xi_T,$$

which is

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

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

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

for all continuously differentiable functions v over $[0, T] \times \mathbb{R}^d$ such that $v(t, \cdot)$ is supported on a compact set of $\mathbb{R}^n$ for every $t \in [0, T]$.

We compute approximate moments of μ^λ by applying the moment-SOS hierarchy. This method consists of minimizing a functional subject to the following constraints:

- The *Liouville* equation (21) expressed in the sense of distributions, as a linear constraints on the moments of μ^λ and μ_T^λ.
- 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

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

for all $a \in \mathbb{N}$ and $b \in \mathbb{N}^n$, with $a + b_1 + b_2 + \cdots + b_n \leqslant d$.

Besides, let us denote

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

and

$$y^T_{a,b} := \int_0^T \int_{\mathbb{R}^n} t^a z^b d\mu^\lambda_T(t, z), \tag{23}$$

$$y^0_{a,b} := \int_0^T \int_{\mathbb{R}^n} t^a z^b \; d\mu^\lambda_0(t, z). \tag{24}$$

We next provide an expression for the solution of (21). In what follows, we let $e_i \in \mathbb{R}^n$ denote the vector whose only non-zero entry is equal to one at position i.

Proposition 4.3. *The Liouville equation* (21) *is equivalently expressed as:*

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

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

$$\langle \partial_t \mu^\lambda, v\rangle + \langle \operatorname{div}(\mu^\lambda g), v\rangle + \langle \mu^\lambda_T, v\rangle = \langle \mu^\lambda_0, v\rangle,$$

which implies

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^n} (\partial_t v(t, z) + \partial_z v(t, z) \cdot g(z)) \; d\mu^\lambda(t, z) = \\ \int_0^T \int_{\mathbb{R}^n} v(t, z) \; (d\mu^\lambda_T(t, z) - d\mu^\lambda_0(t, z)). \end{aligned} \tag{26}$$

We have

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

and

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

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

$$\begin{aligned} &\int_0^T \int_{\mathbb{R}^n} (a t^{a-1} z^b + \sum_{i=1}^n b_i t^a z^{b-e_i} g_i(z)) \; d\mu^\lambda(t, z) \\ &= \int_0^T \int_{\mathbb{R}^n} t^a z^b \; d\mu^\lambda_T(t, z) - \int_0^T \int_{\mathbb{R}^n} t^a z^b \; d\mu^\lambda_0(t, z) \end{aligned}$$

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

4.3. **Numerical computation.** Based on the result of Proposition 4.3, we now describe a numerical method for computing $y^T_{a,b}$. It is assumed that the initial measure $d\mu_0$ is given, which allows us to compute $y^0_{a,b}$. We next describe the main steps involved in writing a semidefinite program for calculating $y^T_{a,b}$ corresponding to the measure $d\mu^\lambda$. Note that, for each $\lambda > 0$, the measure μ^λ is supported on a subset of $\mathbb{R}^{n+1}$. In what follows, we provide some elements of construction for our algorithm for a finite Borel measure μ supported on $\mathbb{R}^p$.

Given a Borel probability measure μ and $\alpha \in \mathbb{N}^p$, we let

$$y_\alpha(\mu) = \int_{\mathbb{R}^p} z^\alpha d\mu(z),$$

where we recall that $z^\alpha := z^{\alpha_1} z^{\alpha_2} \dots z_p^{\alpha_p}$. We consider the set $\{\alpha \in \mathbb{N}^p;\ \alpha_1 + \cdots + \alpha_p \leqslant d\}$ with graded lexicographic order, and denote it by $\mathbb{N}^p_d$; for example, with $p = 2$, $d = 2$, $\mathbb{N}^2_2 = \{(0,0),(1,0),(0,1),(2,0),(1,1),(0,2)\}$. The cardinality of $\mathbb{N}^p_d$ is $s(d) := \binom{p+d}{d}$, which is the number of monomials of degree at most d. The sequence $y = (y_\alpha(\mu))_{\alpha\in\mathbb{N}^p}$ therefore encodes the moments of the measure μ.

The moment matrix of degree d associated with a Borel measure μ, denoted by $M_d(\mu)$ is a matrix of dimension $s(d) \times s(d)$, whose rows and columns are indexed by monomials of degree at most d. For $\alpha, \beta \in \mathbb{N}^p_d$, the corresponding entry in $M_d(\mu)$ is defined by $(M_d(\mu))_{\alpha,\beta} := y_{\alpha+\beta}(\mu)$. As an example, once again with $p = 2$, $d = 2$, $M_2(\mu) \in \mathbb{R}^{6\times 6}$, and the element in second row ($\alpha = (1,0)$), third column ($\beta = (0,1)$), corresponds to $\int_{\mathbb{R}^2} z_1 z_2 \, d\mu(z)$.

To see an alternate representation of $M_d(\mu)$, let $b_d(z) := (z^\alpha)_{\alpha\in\mathbb{N}^p_d} \in \mathbb{R}[z]^{s(d)}_d$ denote the vector of monomials of degree less than or equal to d, with graded lexicographic order. If the sequence $\{y_\alpha\}_{\alpha\in\mathbb{N}^p}$ has a representing measure μ, i.e. $y_\alpha = \int_{\mathbb{R}^p} z^\alpha d\mu(z)$ for all $\alpha \in \mathbb{N}^p$, we can use the equivalent definition $M_d(\mu) := \int_{\mathbb{R}^p} b_d(z) b_d(z)^\top \, d\mu(z)$, where the integral is understood entrywise. We can also define the localizing matrix of degree d with respect to a given $q(z) \in \mathbb{R}[z]$ by

$$M_{d-\lceil \deg(q)/2\rceil}(q\mu) := \int_{\mathbb{R}^p} q(z) b_d(z) b_d(z)^\top \, d\mu(z)$$

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.

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

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

Proposition 4.4. *(Putinar's Theorem) The sequence of moments y has a representing measure supported on X if and only if $M_{d-\lceil \deg p_k/2\rceil}(p_k\mu)$, $k = 0, \dots, n_X$ are positive semidefinite for all $d \in \mathbb{N}$.*

The moment-SOS hierarchy, based on Proposition 4.4, allows us to compute approximate moments of the occupation measure and terminal measures. Recall that the moments of the initial measure are given since the initial measure is given. We fix a degree $d \in \mathbb{N}$ and we consider the linear constraint (25) linking moments of degree up to d, and subject to the constraints that the localizing matrices of the occupation measure and terminal measure, truncated to moments of degree up to d, are all positive semi-definite. This results in a finite-dimensional feasibility problem describe by linear matrix inequalities. The higher is the relaxation degree d, the better are the approximate moments, in the sense that when d tends to infinity, Proposition 4.4 and linear constraint (25) ensure that we have indeed moments of measures satisfying the Liouville equation. For a finite relaxation degree d, it may however happen that the approximate moments are not genuine moments of occupation measures. This is why the approximate moments are sometimes referred to as *pseudo-moments*.

The LMI constraints are automatically constructed by the `msdp` command in Gloptipoly for Matlab [HLL09]. For more details about the LMI constraints, we refer the reader to [Hen13, Section 3.3] or the two introductory chapters of [HKL20].

5. Convergence of Support of Measures

For several applications, it is important to approximate the support of the measure ξ_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 ξ_t by the support of ξ_t^λ where ξ_t^λ satisfies (16).

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

Proposition 5.1. *For each $t \in [0, T]$, it holds*

$$\lim_{\lambda \to 0} d_H(\operatorname{supp}(\xi_t^\lambda), \operatorname{supp}(\xi_t)) = 0. \tag{27}$$

Proof. First, let $A_t^\lambda := \operatorname{supp}(\xi_t^\lambda)$ and $A_t := \operatorname{supp}(\xi_t)$. For proving that

$$\lim_{\lambda \to 0} d_H(A_t^\lambda, A_t) = 0,$$

we need to prove the following two limits:

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

and

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

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 λ converges to zero. Since $y_\lambda \in A_t^\lambda$, there exists $z_0 \in \operatorname{supp}(\xi_0)$ such that $y_\lambda = F_t^\lambda(z_0)$. By choosing $x = F_t(z_0) \in A_t$, it follows from Theorem 3.1 that $\lim_{\lambda \to 0} F_t^\lambda(z_0) = F_t(z_0)$, or equivalently, $|x - y_\lambda|$ converges to 0 as $\lambda \to 0$.

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 $z_0 \in \operatorname{supp}(\xi_0)$ 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 3.1 that $|x - y_\lambda|$ converges to 0 as $\lambda \to 0$, and (29) is obtained. □

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 ξ_t^λ. By Proposition 5.1, by computing such an approximation for $\lambda > 0$ sufficiently small, we get an approximation of the support of the probability measure ξ_t for the original system.

The technique we present is based on approximating the support of a measure by the sublevel sets of a polynomial function. In particular, for a finite Borel measure μ, we consider the moment matrix $M_d(\mu)$ and introduce the mapping,

$$\mathbb{R}^n \ni x \mapsto \Lambda_{\xi,d}(x) := b_d(x)^\top M_d(\mu)^{-1} b_d(x) \in \mathbb{R},$$

which we call Christoffel-Darboux polynomial. Thus, the basic idea behind the construction of the support of the measure μ is to use the finite order moments, and show that the sublevel sets of the Christoffel-Darboux polynomial indeed converge to the actual support of μ. This technique has been proposed in [LP19] for stationary measures under certain hypothesis. Here, we show that by placing certain hypothesis on the initial measure ξ_0, the approximations ξ_t^λ obtained by the Liouville equation satisfy the required hypothesis, which allow us to approximate the support of ξ_t^λ by constructing the corresponding Christoffel-Darboux polynomial.

The following statement shows the existence of a sublevel set that approximates the support of the sequence of measures ξ_t^λ, when λ and $t \in [0, T]$ are fixed.

Proposition 5.2. *Let ξ_0 be absolutely continuous with respect to the Lebesgue measure and let us suppose that $\mathrm{supp}(\xi_0)$ is compact. For a fixed $\lambda > 0$, and $t \in [0, T]$, consider ξ_t^λ obtained by solving (16), and $M_{d,t}^\lambda(\xi_t^\lambda)$ the corresponding moment matrix of degree d. For every $\epsilon > 0$ (small enough), there exists $d \in \mathbb{N}$ (large enough) and $\gamma_d > 0$, such that the sublevel set*

$$S_{d,t}^\lambda := \{z \in \mathbb{R}^p \mid b_d(z)^\top M_{d,t}^\lambda(\xi_t^\lambda)^{-1} b_d(z) \leqslant \gamma_d\} \tag{30}$$

satisfies

$$d_H(S_{d,t}^\lambda, \mathrm{supp}(\xi_t^\lambda)) \leqslant \epsilon, \tag{31}$$

as $d \to +\infty$.

Proof. For each $\lambda > 0$ and $t \in [0, T]$, if we show that

- The set $\mathrm{supp}(\xi_t^\lambda)$ is compact and has nonempty interior.
- It holds that ξ_t^λ is absolutely continuous with respect to the Lebesgue measure.

then, the statement follows by applying [LP19, Theorem 3.11] to the measure ξ_t^λ.

The aforementioned properties basically follow from the fact that, for a fixed $t \in [0, T]$ and $\lambda > 0$, the mapping $F_t^\lambda : \mathbb{R}^n \to \mathbb{R}^n$ is a homeomorphism obtained from the solution of an ODE with Lipschitz continuous right-hand side (9). Let L^λ denote the (uniform with respect to time) Lipschitz constant for the mapping on the right-hand side of (9). One can readily show that for a pair of initial conditions y_0, z_0 and $y_t := F_t^\lambda(y_0)$, $z_t := F_t^\lambda(z_0)$, it holds that

$$|z_0 - y_0| \exp(-L^\lambda t) \leqslant |z_t - y_t| \leqslant |z_0 - y_0| \exp(L^\lambda t).$$

Using this estimate, and recalling that $\xi_0^\lambda = \xi_0$, it readily follows that $\mathrm{supp}(\xi_t^\lambda)$ is compact and has nonempty interior under the given hypothesis on ξ_0.

Absolute continuity of ξ_t^λ with respect to Lebesgue measure holds if ξ_t^λ is absolutely continuous with respect to ξ_0. The later indeed holds because for every

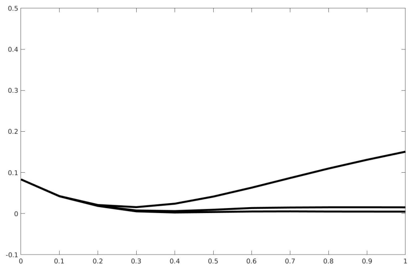


FIGURE 3. First order moment of the second state (vertical axis) of the occupation measure of the regularized system, as a function of time (horizontal axis), for different values of the regularization parameter (top curve $\lambda = 0.5$, middle curve $\lambda = 0.1$, bottom curve $\lambda = 0.05$)

measurable set A, Lipschitz continuity of F_t^λ implies that

$$\xi_0(A) = 0 \Rightarrow \xi_t^\lambda(A) = \xi_0((F_t^\lambda)^{-1}(A)) = 0, \tag{32}$$

whence the desired result follows. □

6. ILLUSTRATIVE EXAMPLES

In this section, we give two examples that illustrate the computation of the pseudo-moments associated with ξ_t^λ of the regularized system (9) in the case where $f : \mathbb{R}^2 \to \mathbb{R}^2$, by applying the methods discussed in this article.

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

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

or equivalently:

$$\begin{cases} \dot{r}(t) = 0, \\ \dot{\theta}_\lambda(t) = -1 - \frac{1}{\lambda}(\theta_\lambda(t) - \max(\theta_\lambda(t), 0)). \end{cases} \tag{33}$$

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

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

Then we apply the moment-SOS hierarchy [HKL20] 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 λ.
- 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 λ, which is illustrated on Figure 3.

Example 3. We now consider an example with moving set S, which is described as follows:

$$S(t) := \{x \in \mathbb{R}^2 \mid g_i(t,x) \leqslant 0, i = 1,2,3,4\} \tag{34}$$

where, for each $i \in \{1,2,3,4\}$, we take g_i to be an affine function with time-varying coefficients:

$$g_i(t,x) = a_i^\top(t)(x - c(t)) - b_i$$

with $a_i(t), c(t) \in \mathbb{R}^2$, $b_i \in \mathbb{R}^2_{>0}$ for each $t \in [0,T]$. Furthermore, the coefficients $a_i(\cdot)$ satisfy the following constraints

$$\begin{gathered}
a_2(t) = -a_1(t), \quad a_3^\top(t)a_1(t) = 0, \quad a_4(t) = -a_3(t), \\
a_i(t)^\top a_i(t) = 1, \qquad i \in \{1,2,3,4\},
\end{gathered}$$

for each $t \in [0,T]$, so that the resulting constraint is a rectangle moving in plane with center at $c(t)$. An illustration of the constraint set, for a fixed t, appears in Figure 4.

For the sake of simulations, we take

$$\begin{gathered}
a_1(t) = \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right)^\top, \quad a_3(t) = \left(\frac{-2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right)^\top \\
b_1 = b_2 = b_3 = b_4 = \frac{1}{4} \\
c(t) = \left(2t^2 - 1, 4t^3 - 3t\right).
\end{gathered}$$

The dynamical system under consideration in this example is

$$\dot{x}(t) \in -\mathcal{N}_{S(t)}(x(t)) \tag{35}$$

and the Moreau-Yosida regularization of this system is,

$$\dot{z}_\lambda = f_\lambda(t, z_\lambda) := -\frac{1}{\lambda}(z_\lambda(t) - \operatorname{proj}(z_\lambda(t), S(t))). \tag{36}$$

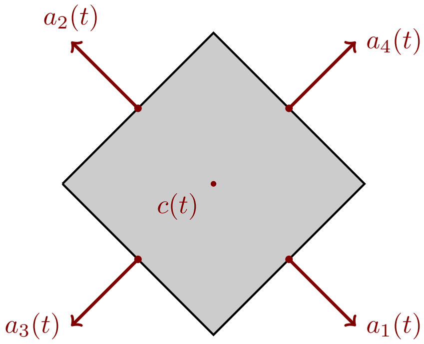


FIGURE 4. Box constraint formed by the intersection of four half-spaces where the coefficients a_i and the center c change with time.

For each $t \in [0,T]$, $S(t)$ has an affine representation. We let $S_j(t) := \{x \in \mathbb{R}^2 \,|\, g_j(t,x) \leqslant 0\}$, and the index of active constraints is defined as $\mathcal{I}_{\mathrm{act}} = \{j \in \{1,2,3,4\} \,|\, g_j(t,x) > 0\}$. Using the fact that the vectors $a_i(t)$ have norm equal to 1 for each $t \geqslant 0$ and are mutually orthogonal, we can write the projection vector as follows:

$$\mathrm{proj}(z, S(t)) = z - \sum_{j \in \mathcal{I}_{\mathrm{act}}} (a_j(t)^\top (z - c(t)) - b_j(t)) a_j(t).$$

Using the approach proposed in this paper, we consider the evolution of the uniformly distributed initial condition over the set $S(0)$ via the Liouville equation associated with (36). For numerical computation of the pseudo-moments associated to the occupation measure, we use `GloptiPoly`. The implementation requires us to provide data in the form of polynomials, and since the vector field in (36) is in the rational form, we have to define a new (scaled) occupation measure as explained next.

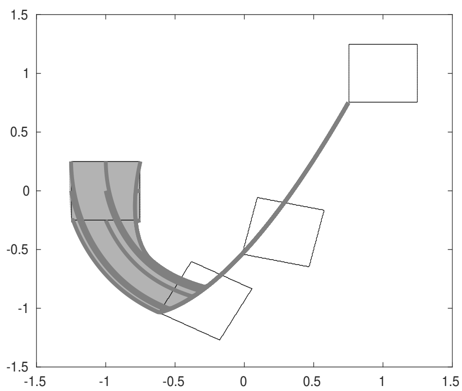
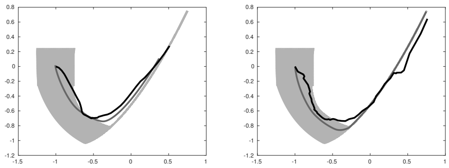


FIGURE 5. From left to right - thin black lines: box $S(t)$ for $t \in \{0, 0.6, 0.8, 1\}$; light gray: ensemble of points initially uniformly supported on $S(0)$ and transported by the flow; dark gray: 9 distinguished points transported by the flow. We observe that the points quickly concentrate on a corner of the box.



(A) With $\lambda = 10^{-1}$ (B) With $\lambda = 10^{-3}$

FIGURE 6. For two values of regularization parameter λ, we represent in dark gray the empirical first moments of the trajectory of (36) (obtained by sampling initial conditions) and in black the first pseudo-moments of the terminal measure computed with GloptiPoly and MOSEK for the relaxation of order 4 (moment-SOS of degree 8). In light gray we represent for reference the ensemble trajectories for a uniform distribution of the initial condition in the box S(0).

For the approach discussed in Section 4.2, we solve the following equation

$$\int_{[0,T]}\int_{\mathbb{R}^2}\left(\partial_t v(t,z)+\partial_z v(t,z)\cdot\frac{h_\lambda(t,z)}{(1+t^2)^2}\right)d\mu^\lambda(t,z) = \int_{[0,T]}\int_{\mathbb{R}^2} v(t,z)(d\mu_T^\lambda(t,z)-d\mu_0^\lambda(t,z)),$$

for a given test function $v(t,z)$, with the measure $\mu^\lambda(t,z)$ as the unknown, and $h_\lambda(t,z)=(1+t^2)^2 f_\lambda(t,z)$ being a polynomial function. To rewrite the foregoing term purely in terms of polynomial functions, instead of rationals, we now consider a new measure of the form

$$\mathrm{d}\nu^\lambda(t,x)=\frac{\mathrm{d}\mu^\lambda(t,x)}{(1+t^2)^2}=\frac{\mathrm{d}t\,\mathrm{d}\xi^\lambda(x|t)}{(1+t^2)^2},$$

so that the modified transport equation takes the following form:

$$\int_{[0,T]}\int_{\mathbb{R}^2}\left(\partial_t v(t,z)(1+t^2)^2+\partial_z v(t,z)\cdot h_\lambda(t,z)\right)d\nu^\lambda(t,z) = \int_{[0,T]}\int_{\mathbb{R}^2} v(t,z)(d\mu_T^\lambda(t,z)-d\mu_0^\lambda(t,z))$$

with $\nu^\lambda,\mu_T^\lambda$ and μ_0^λ as the unknowns. By choosing $v(t,z)=t^a z^b$, all the entries are in polynomial form. It is therefore possible to compute the moments associated to the measure $\mathrm{d}\mu^\lambda$ of the following form:

$$\int_{[0,T]}\int_{\mathbb{R}^2} t^a x^b \mathrm{d}\mu_T^\lambda(t,x)=\int_{[0,T]}\int_{\mathbb{R}^2}(t^a x^b)(1+t^2)^2\ \mathrm{d}\nu_T^\lambda(t,x), \tag{37}$$

where the term on the right-hand sides are computed by our solver.

For illustration of our approach, our objective is to compute the empirical moments associated with the particles which evolve according to the equation (36), and then compare them with the moments which are obtained by the distribution satisfying the Liouville equation with polynomial coefficients associated to system (36). The simulation results are plotted in Figure 6 where we plot the value of (37) by taking $a=0$, $b=(1\ 0)$ and $b=(0\ 1)$ for $T\in[0,1]$ and different values of λ. Along with the ensemble of the trajectories of the differential inclusion (35) with different initial conditions, we plot the empirical moments using the time and space discretization of the regularized ordinary differential equation (36), and the pseudo-moments of the terminal measure computed with `GloptiPoly` and `MOSEK` for certain relaxation order. Solving a moment LMI relaxation of order 4 takes a few seconds on a standard desktop PC. We observe experimentally that increasing the relaxation order does not really improve the accuracy of the results: the optimization problem becomes larger and worse conditioned. Similarly, decreasing the regularization parameter generate large coefficients in the vector field and hence in the coefficients of the LMI problem, and as a result, the optimization problem becomes worse conditioned. On the other hand, the accuracy of the empirical moments computed by time and space discretization of the ODE (36) improves as we decrease the regularization parameter λ.

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, it is not obvious 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. While evolution of probability measure for a class of constrained systems has been studied in [DMS16], here we adopted a different approach for computing the approximations of differential equations associated with the evolution of probability measure. To seek generality in the class of systems studied in our work, one could, just like [DMS16], consider a congestion constraint formulated as a uniform bound on the density of the measure. With the method proposed in this article, we can readily deal with such constraints, which would translate into a linear semidefinite constraint on the moments.

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 examples, we implemented the projection map onto a set 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 [CK20] 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 [LHPT08], 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 [CCMN21, CHHM16, VBP20]. 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 3.1

A.1. **Proof of Lemma 3.2.** For each $\lambda > 0$, the dynamics for z_λ in (9) yield

$$\begin{aligned}|\dot z_\lambda(t)| &= |f(t,z_\lambda(t)) - \frac{1}{\lambda}(z_\lambda(t) - \mathrm{proj}(z_\lambda(t),S(t)))| \\ &\leqslant |f(t,z_\lambda(t))| + \frac{1}{\lambda}|z_\lambda(t) - \mathrm{proj}(z_\lambda(t),S(t))|. \end{aligned} \tag{38}$$

For the first term on the right-hand side of (38), we have that

$$|f(t,z_\lambda(t))| \leqslant L_f(1+|z_\lambda(t)|). \tag{39}$$

For the second term on the right-hand side of (38), 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 $d_\lambda(t) = |z_\lambda(t) - \mathrm{proj}(z_\lambda(t),S(t))|$. So $\frac{1}{\lambda}|z_\lambda(t) - \mathrm{proj}(z_\lambda(t),S(t))| = \frac{1}{\lambda}d_\lambda(t)$.
To obtain a bound on $d_\lambda(t)$, we compute the derivative of $d_\lambda^2(t)$:

$$\begin{aligned}\frac{d}{dt}d_\lambda^2(t) &= \frac{d}{dt}d^2_{S(t)}(z_\lambda(t)) \\ &= \lim_{\epsilon\to 0}\frac{d^2_{S(t+\epsilon)}(z_\lambda(t+\epsilon)) - d^2_{S(t)}(z_\lambda(t))}{\epsilon} \\ &= \lim_{\epsilon\to 0}\frac{d^2_{S(t+\epsilon)}(z_\lambda(t+\epsilon)) - d^2_{S(t)}(z_\lambda(t+\epsilon))}{\epsilon} \\ &\qquad + \frac{d^2_{S(t)}(z_\lambda(t+\epsilon)) - d^2_{S(t)}(z_\lambda(t))}{\epsilon}.\end{aligned} \tag{40}$$

For the first term in the limit, we use that

$$\begin{aligned}&d^2_{S(t+\epsilon)}(z_\lambda(t+\epsilon)) - d^2_{S(t)}(z_\lambda(t+\epsilon)) \\ &\leqslant d_H(S(t+\epsilon),S(t))\big(d_{S(t+\epsilon)}(z_\lambda(t+\epsilon)) + d_{S(t)}(z_\lambda(t+\epsilon))\big) \\ &\leqslant |\epsilon| L_S\big(d_{S(t+\epsilon)}(z_\lambda(t+\epsilon)) + d_{S(t)}(z_\lambda(t+\epsilon))\big).\end{aligned} \tag{41}$$

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

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

Since $z_\lambda(.)$ 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

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

And,

$$\begin{aligned}&\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[d^2_{S(t)}(z_\lambda(t+\epsilon)) - d^2_{S(t)}(z_\lambda(t))\right] \\ &= \langle \nabla d^2_{S(t)}(z_\lambda(t)), \dot z_\lambda(t)\rangle \\ &= 2\langle z_\lambda(t) - \mathrm{proj}(z_\lambda(t),S(t)), \dot z_\lambda(t)\rangle \\ &= -\frac{2}{\lambda}d_\lambda^2(t) + 2\langle z_\lambda(t) - \mathrm{proj}(z_\lambda(t),S(t)), f(t,z_\lambda(t))\rangle.\end{aligned} \tag{42}$$

By substitution of (41) and (42) in equation (40), we obtain

$$2d_\lambda(t)\dot{d}_\lambda(t) = \frac{d}{dt}d_\lambda^2(t) \leqslant -\frac{2}{\lambda}d_\lambda^2(t) + 2d_\lambda(t)\big|f(t,z_\lambda(t))\big| + 2L_S d_\lambda(t).$$

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

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

which implies that,

$$d_\lambda(t) \leqslant e^{-t/\lambda}d_\lambda(0) + \int_0^t e^{-(t-s)/\lambda}\left(\big|f(s,z_\lambda(s))\big| + L_S\right)\,ds.$$

Or, $d_\lambda(0) = |z_0 - \operatorname{proj}(z_0, S(0))| = 0$ since $z_0 \in S(0)$ and we have that f satisfies (39), then it follows

$$\frac{1}{\lambda}d_\lambda(t) \leqslant \frac{1}{\lambda}\int_0^t e^{-(t-s)/\lambda}(L_f + L_f|z_\lambda(s)| + L_S)\,ds. \tag{43}$$

And therefore, substituting (39) and (43) in (38), we get

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

We have

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

Similarly,

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

Besides, we have

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

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

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

A.2. **Proof of Lemma 3.4.** For fixed $t_1, t_2 \in [0,T]$, let

$$\widetilde{S}_\lambda(s) := S(s) - \int_{t_1}^s f(r, z_\lambda(r))\,dr,$$

with $s \in [t_1, t_2]$. Under the assumptions imposed on $S(\cdot)$ and the mapping $f(\cdot,\cdot)$, the set-valued mapping $\widetilde{S}_\lambda(\cdot)$ is Hausdorff continuous. Thus, by a theorem on continuous selections [Mic56], there exists a continuous function $\varphi_\lambda : [t_1, t_2] \to \mathbb{R}^n$ such that $\varphi_\lambda(s) \in \widetilde{S}_\lambda(s)$ for each $s \in [t_1, t_2]$ and by construction, we have $\varphi_\lambda(t_1) \in \widetilde{S}_\lambda(t_1) = S(t_1)$. By taking $\lambda \to 0$, we see that $z_\lambda \to z$; and letting

$\varphi(s) := \lim_{\lambda\to 0}\varphi_\lambda(s)$, we observe that $\varphi(\cdot)$ is continuous and $\varphi(s) \in \widetilde{S}(s) := S(s) - \int_{t_1}^s f(r, z(r))\ dr$.

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

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

Then

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

which implies that,

$$\begin{aligned}&\int_{t_1}^{t_2} \Big\langle \varphi_\lambda(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s))\Big\rangle\ ds \geqslant \\ &\qquad\qquad \int_{t_1}^{t_2} \Big\langle \bar{z}_\lambda(s) - \int_{t_1}^s f(r, z_\lambda(r))\ dr, \dot{z}_\lambda(s) - f(s, z_\lambda(s))\Big\rangle\ ds.\end{aligned}$$

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

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

it follows that,

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

and,

$$\begin{aligned}&\int_{t_1}^{t_2} \Big\langle \varphi_\lambda(s), \dot{z}_\lambda(s) - f(s, z_\lambda(s))\Big\rangle\ ds \geqslant \\ &\qquad\qquad \int_{t_1}^{t_2} \Big\langle z_\lambda(s) - \int_{t_1}^s f(r, z_\lambda(r))\ dr, \dot{z}_\lambda(s) - f(s, z_\lambda(s))\Big\rangle\ ds.\end{aligned}$$

We have

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

hence, we obtain that

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

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

Therefore, this yields to

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

and Lemma 3.4 is then proved.

References

[Adl18] S. Adly. *A variational approach to nonsmooth dynamics*. Springer Briefs in Mathematics. Springer, 2018.

[AF90] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Birkhäuser, Berlin, 1990.

[Amb08] L. Ambrosio. Transport equation and cauchy problem for non-smooth vector fields. In B. Dacorogna and P. Marcellini, editors, *Calculus of Variations and Nonlinear Partial Differential Equations*, pages 1–41. Springer Berlin Heidelberg, 2008.

[BBFR19] J. Bartsch, A. Borzì, 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.

[Ber03] F. Bernardin. Multivalued stochastic differential equations: Convergence of a numerical scheme. *Set-Valued Analysis*, 11:393–415, 2003.

[BF21] B. Bonnet and H. Frankowska. Differential inclusions in Wasserstein spaces: The Cauchy-Lipschitz framework. *Journal of Differential Equations*, 271:594–637, 2021.

[BJ98] F. Bouchut and F. James. One dimensional transport equation with discontinuous coefficients. *Nonlinear Analysis*, 32:891–933, 1998.

[BJM05] 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.

[BL78] A. Bensoussan and J.L. Lions. *Applications des inégalités variationnelles en contrôle stochastique*. Dunod, Paris, 1978.

[Bre73] H. Brezis. *Opérateurs Maximaux Monotones et Semi-Groupes des Contractions dans les Espaces de Hilbert*. North-Holland, Mathematics Studies, 1973.

[Bro07] R.W. Brockett. Optimal control of the Liouville equation. *AMS/IP Studies in Advanced Mathematics*, 39:23–35, 2007.

[Bro12] R.W. Brockett. Notes on the control of the liouville equation. In P. Cannarsa and J.-M Coron, editors, *Control of Partial Differential Equations: Cetraro, Italy 2010*, pages 101–129. Springer Berlin Heidelberg, 2012.

[BT20] 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.

[BV11] F. Bernicot and J. Venel. Stochastic perturbation of sweeping process and a convergence result for an associated numerical scheme. *Journal of Differential Equations*, 251(4-5):1195–1224, 2011.

[Ć95] E. Cépa. Équations différentielles stochastiques multivoques. In *Séminaire de Probabalités*, volume 29, pages 86–107, 1995.

[CCMN21] 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.

[CHHM16] 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.

[CIT22] M.-K. Camlibel, L. Iannelli, and A. Tanwani. Convergence of proximal solutions for evolution inclusions with time-dependent maximal monotone operators. *Mathematical Programming*, 194(1-2):1017–1059, 2022.

[CK20] 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.

[CMMdF14] C. Castaing, M.-D.-P Monteiro-Marques, and P. Raynaud de Fitte. Some problems in optimal control governed by the sweeping process. *Journal of Nonlinear and Convex Analysis*, 15(5):1043–1070, 2014.

[CMMdF16] C. Castaing, M.-D.-P Monteiro-Marques, and P. Raynaud de Fitte. A Skorokhod problem governed by a closed convex moving set. *Journal of Convex Analysis*, 23(2):387–423, 2016.

[DL89] R.-J. DiPerna and P.L. Lions. Ordinary differential equations, transport theory and sobolev spaces. *Invent. Math.*, 98:511–547, 1989.

[DMS16] S. Di Marino, B. Maury, and F. Santambrogio. Measure sweeping processes. *Journal of Convex Analysis*, 23(2):567–601, 2016.

[ET06] J.-F. Edmond and L. Thibault. BV solutions of nonconvex sweeping process differential inclusion with perturbation. *Journal of Differential Equations*, 226:135–179, 2006.

[Hen13] D. Henrion. Optimization on linear matrix inequalities for polynomial systems control. In *Les cours du C.I.R.M.*, volume 3(1), pages 1–44. 2013.

[HKL20] D. Henrion, M. Korda, and J. B. Lasserre. *The Moment-SOS Hierarchy.* World Scientific, 2020.

[HLL09] D. Henrion, J.B. Lasserre, and J. Löfberg. Gloptipoly 3: moments, optimization and semidefinite programming. *Optimization Methods & Software*, 24(4-5):761–779, 2009.

[KM96] M. Kunze and M.D.P. Monteiro Marques. Yosida-Moreau regularization of sweeping processes with unbounded variation. *J. Differential Equations*, 130:292–306, 1996.

[LHPT08] 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.

[LP19] J.B. Lasserre and E. Pauwels. The empirical Christoffel function with applications in data analysis. *Advances in Computational Mathematics*, 45(4):1439–1468, 2019.

[Mic56] E. Michael. Continuous selections. I. *Annals of Mathematics*, 63(2):361–382, 1956.

[Mor77] 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.

[Mor06] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation. Part I: Basic Theory.* Springer, Berlin, 2006.

[NT19] F. Nacry and L. Thibault. Regularization of sweeping process: old and new. *Pure Applied Functional Analysis*, 4:59–117, 2019.

[RW98] R.-T. Rockafellar and R.-J.-B. Wets. *Variational Analysis.* Springer Verlag, Berlin Heidelberg, 1998.

[STH22] M. Souaiby, A. Tanwani, and D. Henrion. Cone-copositive Lyapunov functions for complementarity systems: Converse result and polynomial approximation. *IEEE Transactions on Automatic Control*, 67(3):1253–1268, 2022.

[VBP20] 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.

[Vil03] C. Villani. *Topics in optimal transportation.* American Mathematical Society, Providence, 2003.