

Finite-Time Convergence of Continuous-Time Optimization Algorithms via Differential Inclusions

Romero, Orlando; Benosman, Mouhacine

TR2020-001 January 08, 2020

Abstract

In this paper, we propose two discontinuous dynamical systems in continuous time with guaranteed prescribed finite-time local convergence to strict local minima of a given cost function. Our approach consists of exploiting a Lyapunov-based differential inequality for differential inclusions, which leads to finite-time stability and thus finite-time convergence with a provable bound on the settling time. In particular, for exact solutions to the aforementioned differential inequality, the settling-time bound is also exact, thus achieving prescribed finite-time convergence. We thus construct a class of discontinuous dynamical systems, of second order with respect to the cost function, that serve as continuous-time optimization algorithms with finite-time convergence and prescribed convergence time. Finally, we illustrate our results on the Rosenbrock function.

Advances in Neural Information Processing Systems (NIPS)- Workshop

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Finite-Time Convergence of Continuous-Time Optimization Algorithms via Differential Inclusions

Orlando Romero

Department of Industrial and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180, USA
rodrio2@rpi.edu

Mouhacine Benosman

Mitsubishi Electric Research Laboratories
Cambridge, MA 02139, USA
benosman@merl.com

Abstract

In this paper, we propose two discontinuous dynamical systems in continuous time with guaranteed prescribed finite-time local convergence to strict local minima of a given cost function. Our approach consists of exploiting a Lyapunov-based differential inequality for differential inclusions, which leads to finite-time stability and thus finite-time convergence with a provable bound on the settling time. In particular, for exact solutions to the aforementioned differential inequality, the settling-time bound is also exact, thus achieving prescribed finite-time convergence. We thus construct a class of discontinuous dynamical systems, of second order with respect to the cost function, that serve as continuous-time optimization algorithms with finite-time convergence and prescribed convergence time. Finally, we illustrate our results on the Rosenbrock function.

1 Introduction

In continuous-time optimization, an ordinary differential equation (ODE), partial differential equation (PDE), or differential inclusion is designed in terms of a given cost function, in such a way to lead the solutions to converge (forward in time) to an optimal value of the cost function. To achieve this, tools from Lyapunov stability theory are often employed, mainly because there already exists a rich body of work within the nonlinear systems and control theory community for this purpose. In particular, we seek *asymptotically* Lyapunov stable gradient-based systems with an equilibrium (stationary point) at an isolated extremum of the given cost function, thus certifying local convergence. Naturally, *global* asymptotic stability leads to global convergence, though such an analysis will typically require the cost function to be strongly convex everywhere.

For early work in this direction, see (Botsaris, 1978a,b), (Zghier, 1981), (Snyman, 1982, 1983), and (Brown, 1989). Brockett (1988) and, subsequently, Helmke and Moore (1994), studied relationships between linear programming, ODEs, and general matrix theory. Further, Schropp (1995) and Schropp and Singer (2000) explored several aspects linking nonlinear dynamical systems to gradient-based optimization, including nonlinear constraints. Cortés (2006) proposed two discontinuous normalized modifications of gradient flows to attain *finite-time convergence*. Later, Wang and Elia (2011) proposed a control-theoretic perspective on centralized and distributed convex optimization.

More recently, Su, Boyd, and Candes (2014) derived a second-order ODE as the limit of Nesterov’s accelerated gradient method, when the gradient step sizes vanish. This ODE is then used to study Nesterov’s scheme from a new perspective, particularly in an larger effort to better understand acceleration without substantially increasing computational burden. Expanding upon the aforementioned idea, França, Robinson, and Vidal (2018) derived a second-order ODE that models the continuous-time limit of the sequence of iterates generated by the alternating direction method of multipliers (ADMM). Then, the authors employ Lyapunov theory to analyze the stability at critical points of the dynamical systems and to obtain associated convergence rates.

Later, França, Robinson, and Vidal (2019) analyze general non-smooth and linearly constrained optimization problems by deriving equivalent (at the limit) non-smooth dynamical systems related to variants of the relaxed and accelerated ADMM. Then, the authors employ Lyapunov theory to analyze the stability at critical points of the dynamical systems and to obtain associated convergence rates. Later, França, Robinson, and Vidal (2019) analyze general non-smooth and linearly constrained optimization problems by deriving equivalent (at the limit) non-smooth dynamical systems related to variants of the relaxed and accelerated ADMM.

In the more traditional context of machine learning, only a few papers have adopted the approach of explicitly borrowing or connecting ideas from control and dynamical systems. For unsupervised learning, Plumbley (1995) proposes Lyapunov stability theory as an approach to establish convergence of principal component algorithms. Pequito et al. (2011) and Aquilanti et al. (2019) propose continuous-time generalized expectation-maximization (EM) algorithms, based on mean-field games, for clustering of finite mixture models. Romero, Chaterjee, and Pequito (2019) establish convergence of the EM algorithm, and a class of generalized EM algorithms denoted δ -EM, via discrete-time Lyapunov stability theory. For supervised learning, Liu and Theodorou (2019) provide a review of deep learning from the perspective of control and dynamical systems, with a focus in optimal control. Zhu (2018) and Rahnama, Nguyen, and Raff (2019) explore connections between control theory and adversarial machine learning.

Statement of Contribution

In this work, we provide a Lyapunov-based tool to both check and construct continuous-time dynamical systems that are finite-time stable and thus lead to finite-time convergence of the candidate Lyapunov function (intended as a surrogate to a given cost function) to its minimum value. In particular, we first extend one of the existing Lyapunov-based inequality condition for finite-time convergence of the usual Lipschitz continuous dynamical systems, to the case of arbitrary differential inclusions. We then use this condition to construct a family of *discontinuous, second-order flows*, which guarantee local convergence to a local minimum, in *prescribed finite time*. One of the proposed families of continuous-time optimization algorithms is tested on a well-known optimization testcase, namely, the *Rosenbrock function*.

2 Finite-Time Convergence in Optimization via Finite-Time Stability

Consider some objective cost function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ that we wish to minimize. In particular, let $x^* \in \mathbb{R}^n$ be an arbitrary local minimum of f that is unknown to us. In continuous-time optimization, we typically proceed by designing a nonlinear state-space dynamical system

$$\dot{x} = F(x), \tag{1}$$

or a time-varying one replacing $F(x)$ with $F(t, x)$, for which $F(x)$ can be computed without explicit knowledge of x^* and for which (1) is certifiably asymptotically stable at x^* . Ideally, computing $F(x)$ should be possible using only up to second-order information on f .

In this work, however, we seek dynamical systems for which (1) is certifiably *finite-time* stable at x^* . As will be clear later, such systems need to be possibly discontinuous or non-Lipschitz, based on differential inclusions instead of ODEs. Our approach to achieve this objective is largely based on exploiting the Lyapunov-like differential inequality

$$\dot{\mathcal{E}}(t) \leq -c\mathcal{E}(t)^\alpha, \quad \text{a.e. } t \geq 0, \tag{2}$$

with constants $c > 0$ and $\alpha < 1$, for absolutely continuous functions \mathcal{E} such that $\mathcal{E}(0) > 0$. Indeed, under the aforementioned conditions, $\mathcal{E}(t) \rightarrow 0$ will be reached in finite time $t \rightarrow t^* \leq \frac{\mathcal{E}(0)^{1-\alpha}}{c(1-\alpha)} < \infty$.

We now summarize the problem statement:

Problem 1. Given a sufficiently smooth cost function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ with a sufficiently regular local minimizer x^* , solve the following tasks:

1. Design a sufficiently smooth¹ candidate Lyapunov function V for which is defined and positive definite near and w.r.t. x^* ².
2. Design a (possibly discontinuous) system³ (1) such that $F(x)$ can be computed near $x = x^*$ without explicit knowledge of x^* , and, given any Filippov solution⁴ $x(\cdot)$ of (1) with $x(0) = x_0 \neq x^*$, the differential inequality (2) is satisfied for $\mathcal{E}(t) \triangleq V(x(t))$.

By following this strategy, we will therefore achieve (local and strong) *finite-time stability*, and thus finite-time convergence. Furthermore, if $V(x_0)$ can be upper bounded, then F can be readily tuned to achieve finite-time convergence under a *prescribed* range for the settling time, or even with *exact* prescribed settling time if $V(x_0)$ can be explicitly computed and (2) holds exactly.

3 A Family of Finite-Time Stable, Second-Order Optimization Flows

We now propose a family of second-order optimization methods with finite-time convergence constructed using two *gradient-based* Lyapunov functions, namely $V(x) = \|\nabla f(x)\|^2$ and $V(x) = \|\nabla f(x)\|_1$. First, we need to assume sufficient smoothness on the cost function.

Assumption 1. $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is twice continuously differentiable and strongly convex in an open neighborhood $\mathcal{D} \subseteq \mathbb{R}^n$ of a stationary point $x^* \in \mathbb{R}^n$.

Since $\nabla V(x) = 2\nabla^2 f(x)\nabla f(x)$ for $V(x) = \|\nabla f(x)\|^2$ and $\nabla V(x) = \nabla^2 f(x) \text{sign}(\nabla f(x))$ a.e. for $V(x) = \|\nabla f(x)\|_1$, we can readily design Filippov differential inclusions that are finite-time stable at x^* . In particular, we may design such differential inclusions to achieve an *exact* and *prescribed* finite settling time, at the trade-off of requiring second-order information on f .

Given a symmetric and positive definite matrix $P \in \mathbb{R}^{n \times n}$ with SVD decomposition $P = V\Sigma V^\top$, $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_1, \dots, \lambda_n > 0$, we define $P^r \triangleq V\Sigma^r V^\top$, where $\Sigma^r \triangleq \text{diag}(\lambda_1^r, \dots, \lambda_n^r)$.

Theorem 1. Let $c > 0$, $p \in [1, 2)$, and $r \in \mathbb{R}$. Under Assumption 1, any maximal Filippov solution to the discontinuous second-order generalized Newton-like flows

$$\dot{x} = -c\|\nabla f(x)\|^p \frac{[\nabla^2 f(x)]^r \nabla f(x)}{\nabla f(x)^\top [\nabla^2 f(x)]^{r+1} \nabla f(x)} \quad (3)$$

and

$$\dot{x} = -c\|\nabla f(x)\|_1^{p-1} \frac{[\nabla^2 f(x)]^r \text{sign}(\nabla f(x))}{\text{sign}(\nabla f(x))^\top [\nabla^2 f(x)]^{r+1} \text{sign}(\nabla f(x))} \quad (4)$$

with sufficiently small $\|x_0 - x^*\| > 0$ (where $x_0 = x(0)$) will converge in finite time to x^* . Furthermore, their convergence times are given exactly by

$$t^* = \frac{\|\nabla f(x_0)\|^{2-p}}{c(2-p)} \quad t^* = \frac{\|\nabla f(x_0)\|_1^p}{cp} \quad (5)$$

for (3)-(4), respectively, where $x_0 = x(0)$. In particular, given any compact and positively invariant subset $S \subset \mathcal{D}$, both flows converge in finite with the aforementioned settling time upper bounds (which can be tightened by replacing $\overline{\mathcal{D}}$ with S) for any $x_0 \in S$. Furthermore, if $\mathcal{D} = \mathbb{R}^n$, then we have global finite-time convergence, i.e. finite-time convergence to any maximal Filippov solution $x(\cdot)$ with arbitrary $x_0 = x(0) \in \mathbb{R}^n$.

Proof. See supplementary material, appendix C. ■

¹At least locally Lipschitz continuous and *regular* (see Definition 2 of the supplementary material, Appendix A).

²In other words, there exists some open neighborhood \mathcal{D} of x^* such that V is defined in \mathcal{D} and satisfies $V(x^*) = 0$ and $V(x) > 0$ for every $x \in \mathcal{D} \setminus \{x^*\}$.

³Right-hand side defined at least a.e., Lebesgue measurable, and locally essentially bounded.

⁴See supplementary material, Appendix A.

4 Numerical Experiment: Rosenbrock Function

We will now test one of our proposed flows on the Rosenbrock function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, given by

$$f(x_1, x_2) = (a - x_1)^2 + b(x_2 - x_1^2)^2, \quad (6)$$

with parameters $a, b \in \mathbb{R}$. This function is nonlinear and non-convex, but smooth. It possesses exactly one stationary point $(x_1^*, x_2^*) = (a, a^2)$ for $b \geq 0$, which is a strict global minimum for $b > 0$. If $b < 0$, then (x_1^*, x_2^*) is a saddle point. Finally, if $b = 0$, then $\{(a, x_2) : x_2 \in \mathbb{R}\}$ are the stationary points of f , and they are all non-strict global minima.

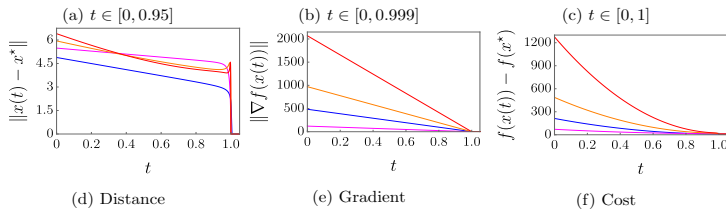


Figure 1: Trajectories of the proposed flow (3) with $(c, p, r) = (\|\nabla f(x_0)\|, 1, -1)$ and four different initial conditions $x_0 \in \mathbb{R}^2$, for the Rosenbrock function with parameters $(a, b) = (2, 50)$, which has a unique minimum $x^* = (a, a^2) = (2, 4)$.

As we can see in Figure 1, this flow converges correctly to the minimum $(a, a^2) = (2, 4)$ from all the tested initial conditions with an exact prescribed settling time $t^* = 1$. It should be noted that at any given point in the trajectory $x(\cdot)$, the functions $t \mapsto \|x(t) - x^*\|$ and $t \mapsto |f(x(t)) - f(x^*)|$ are not guaranteed to decrease or remain constant, indeed only $t \mapsto \|\nabla f(x(t))\|$ can be guaranteed to do so, which explains the increase in Figure 1-(d) that could never have occurred in Figure 1-(e).

5 Conclusions and Future Work

We have introduced a new family of second-order flows for continuous-time optimization. The main characteristic of the proposed flows is their *finite-time* convergence guarantees. Furthermore, they are designed in such a way that the finite convergence time can be prescribed by the user. To be able to analyze these discontinuous flows, we resorted to establishing finite-time stability. In order to do this, we first extended an existing sufficient condition for finite-time stability through a Lyapunov-based inequality, in the case of smooth dynamics, to the case of non-smooth dynamics modeled by differential inclusions. One of the proposed families was tested on a well-known optimization benchmark – the Rosenbrock function.

While the obtained results are encouraging, it should be clear that currently available numerical solvers for our proposed flows do not translate into competitive *iterative* optimization algorithms. Deciding how to best discretize a given continuous-time optimization algorithm (i.e. as given by an ODE, PDE, or differential inclusion), or how to compare two such continuous-time algorithms in terms of the performance of corresponding iterative schemes suitable for digital computers, largely remains an open problem, but nonetheless a very active topic of research as of right now.

For the aforementioned reasons, future work will be dedicated to studying discretization of our proposed flows (and other finite-time stable flows) that will hopefully lead to either accelerated schemes when compared to currently available methods, or to a better understanding of the intrinsic boundaries of acceleration strategies. Furthermore, we will extend our results to constrained optimization; gradient-free optimization; time-varying optimization; and extremum-seeking control where the derivative of the cost function is estimated from direct measurements of the cost.

References

- Aquilanti, L.; Cacace, S.; Camilli, F.; and De Maio, R. 2019. A mean field games approach to cluster analysis.
- Bacciotti, A., and Ceragioli, F. 1999. Stability and stabilization of discontinuous systems and nonsmooth lyapunov functions. *ESAIM: Control, Optimisation and Calculus of Variations* 4:361–376.
- Benosman, M.; Liao, F.; Lum, K.; and Wang, J. L. 2009. Nonlinear control allocation for non-minimum phase systems. *IEEE Transactions on Control Systems Technology* 17(2):394–404.
- Botsaris, C. 1978a. A class of methods for unconstrained minimization based on stable numerical integration techniques. *Journal of Mathematical Analysis and Applications* 63(3):729–749.
- Botsaris, C. 1978b. Differential gradient methods. *Journal of Mathematical Analysis and Applications* 63(1):177–198.
- Brockett, R. 1988. Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems. 799–803.
- Brown, A. 1989. Some effective methods for unconstrained optimization based on the solution of systems of ordinary differential equations. *Journal of Optimization Theory and Applications* 62(2):211–224.
- Clarke, F. H. 1981. Generalized gradients of lipschitz functionals. *Advances in Mathematics* 40(1):52–67.
- Cortés, J., and Bullo, F. 2005. Coordination and geometric optimization via distributed dynamical systems. *SIAM Journal on Control and Optimization* 44(5):1543–1574.
- Cortés, J. 2006. Finite-time convergent gradient flows with applications to network consensus. *Automatica* 42(11):1993–2000.
- Cortés, J. 2008. Discontinuous dynamical systems. *IEEE Control Systems Magazine* 28(3):36–73.
- Filippov, A. F., and Arscott, F. M. 1988. *Differential equations with discontinuous righthand sides*. Dordrecht, Netherlands: Kluwer Academic Publishers Group.
- França, G.; Robinson, D.; and Vidal, R. 2018. ADMM and accelerated ADMM as continuous dynamical systems.
- França, G.; Robinson, D.; and Vidal, R. 2019. A dynamical systems perspective on nonsmooth constrained optimization. *arXiv:1808.04048 [math.OC]*.
- Helmke, U., and Moore, J. B. 1994. *Optimization and Dynamical Systems*. Springer-Verlag.
- Hui, Q.; Haddad, W.; and Bhat, S. 2009. Semistability, finite-time stability, differential inclusions, and discontinuous dynamical systems having a continuum of equilibria. *IEEE Transactions on Automatic Control* 54:2465–2470.
- Liu, G.-H., and Theodorou, E. 2019. Deep learning theory review: An optimal control and dynamical systems perspective. *arXiv preprint 1908.10920*.
- Paden, B., and Sastry, S. 1987. A calculus for computing filippov’s differential inclusion with application to the variable structure control of robot manipulators. *IEEE Transactions on Circuits and Systems* 34:73–82.
- Pequito, S. D.; Aguiar, A. P.; Sinopoli, B.; and Gomes, D. A. 2011. Unsupervised learning of finite mixture models using mean field games. *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 321–328.
- Plumbley, M. D. 1995. Lyapunov functions for convergence of principal component algorithms. *Neural Networks* 8(1):11 – 23.

- Rahnama, A.; Nguyen, A. T.; and Raff, E. 2019. Connecting lyapunov control theory to adversarial attacks. *ArXiv abs/1907.07732*.
- Romero, O.; Chatterjee, S.; and Pequito, S. 2019. Convergence of the expectation-maximization algorithm through discrete-time lyapunov stability theory. *Proceedings of the American Control Conference (ACC)* 163–168.
- Schropp, J., and Singer, I. 2000. A dynamical systems approach to constrained minimization. *Numerical Functional Analysis and Optimization* 21:537–551.
- Schropp, J. 1995. Using dynamical systems methods to solve minimization problems. *Applied Numerical Mathematics* 18(1):321 – 335.
- Snyman, J. 1982. A new and dynamic method for unconstrained minimization. *Applied Mathematical Modelling* 6(6):448–462.
- Snyman, J. 1983. An improved version of the original leap-frog dynamic method for unconstrained minimization: LFOP1(b). *Applied Mathematical Modelling* 7(3):216–218.
- Su, W.; Boyd, S.; and Candes, E. J. 2014. A differential equation for modeling Nesterov’s accelerated gradient method: Theory and insights. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc. 2510–2518.
- Wang, J., and Elia, N. 2011. A control perspective for centralized and distributed convex optimization. In *IEEE Conference on Decision and Control and European Control Conference*, 3800–3805.
- Zghier, A. K. 1981. *The use of differential equations in optimization*. Ph.D. Dissertation, Loughborough University.
- Zhu, X. 2018. An optimal control view of adversarial machine learning. *ArXiv abs/1811.04422*.

Supplementary Material

A Discontinuous Systems and Filippov Differential Inclusions

Recall that for an initial value problem (IVP)

$$\dot{x}(t) = F(x(t)) \quad (7a)$$

$$x(0) = x_0 \quad (7b)$$

with $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$, the typical way to check for existence of solutions is by establishing continuity of F . Likewise, to establish unicity of solution, we typically seek Lipschitz continuity. When F is discontinuous, but nonetheless Lebesgue measurable and locally essentially uniformly bounded, we may understand (7a) as the Filippov differential inclusion

$$\dot{x}(t) \in \mathcal{K}[F](x(t)), \quad (8)$$

where $\mathcal{K}[F] : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ denotes the Filippov set-valued map given by

$$\mathcal{K}[F](x) \triangleq \bigcap_{\delta > 0} \bigcap_{\mu(S)=0} \overline{\text{co}} F(B_\delta(x) \setminus S), \quad (9)$$

where μ denotes the usual Lebesgue measure and $\overline{\text{co}}$ the convex closure, *i.e.* closure of the convex hull co . For more details, see (Paden and Sastry, 1987), from which this section is based on.

Assumption 2. $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined a.e. and is Lebesgue measurable in a non-empty open region $U \subset \mathbb{R}^n$. Further, F is locally essentially bounded, *i.e.*, for every point $x \in \mathbb{R}^n$, F is bounded a.e. in some bounded neighborhood of x .

Definition 1 (Filippov). Under Assumption 2, we say that $x : [0, \tau) \rightarrow \mathbb{R}^n$ with $0 < \tau \leq \infty$ is a *Filippov solution* to (7) if $x(\cdot)$ is absolutely continuous, $x(0) = x_0$, and (8) holds a.e. in every compact subset of $[0, \tau)$. Furthermore, we say that $x : [0, \tau) \rightarrow \mathbb{R}^n$ is a *maximal Filippov solution* if no other Filippov solution $x'(\cdot)$ exists with $x = x'|_{[0, \tau)}$.

For a comprehensive overview of discontinuous systems, including sufficient conditions for existence (Proposition 3) and uniqueness (Propositions 4 and 5) of Filippov solutions, see (Cortés, 2008). In particular, it can be established that Filippov solutions to (7) exist, provided that Assumption 2 holds.

Proposition 1 (Paden and Sastry (1987), Theorem 1). *Under Assumption 2, (9) can be computed as*

$$\mathcal{K}[F](x) = \left\{ \lim_{k \rightarrow \infty} F(x_k) : x_k \notin \mathcal{N}_F \cup S, x_k \rightarrow x \right\} \quad (10)$$

for some (Lebesgue) zero-measure set $\mathcal{N}_F \subset \mathbb{R}^n$ and any other zero-measure set $S \subset \mathbb{R}^n$. In particular, if F is continuous at a fixed x , then $\mathcal{K}[F](x) = \{F(x)\}$.

For instance, for a gradient flow, we have $\mathcal{K}[-\nabla f](x) = \{-\nabla f(x)\}$ for every $x \in \mathbb{R}^n$, provided that f is continuously differentiable. Furthermore, if f is only locally Lipschitz continuous and regular (see Definition 2 of Appendix B), then $\mathcal{K}[-\nabla f](x) = -\partial f(x)$, where

$$\partial f(x) \triangleq \left\{ \lim_{k \rightarrow \infty} \nabla f(x_k) : x_k \notin S \cup \mathcal{N}_f, x_k \rightarrow x \right\} \quad (11)$$

denotes Clarke's generalized gradient (Clarke, 1981) of f , with S being any zero-measure set and \mathcal{N}_f the zero-measure set over which f is not differentiable.

B Finite-Time Stability of Differential Inclusions

Consider a general differential inclusion (see (Bacciotti and Ceragioli, 1999) for more details)

$$\dot{x}(t) \in K(x(t)) \quad (12)$$

where $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a set-valued map.

Assumption 3. $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ has nonempty, compact, and convex values, and is *upper semi-continuous*.

Filippov and Arscott (1988) proved that, under Assumption 2, the Filippov set-valued map $K = \mathcal{K}[F]$ satisfies the conditions of Assumption 3.

Similarly to the previous case of Filippov solutions, we say that $x : [0, \tau) \rightarrow \mathbb{R}^n$ with $0 < \tau \leq \infty$ is a *Carathéodory solution* to (12) if $x(\cdot)$ is absolutely continuous and (12) is satisfied a.e. in every compact subset of $[0, \tau)$. Furthermore, we say that $x(\cdot)$ is a *maximal Carathéodory solution* if no other Carathéodory solution $x'(\cdot)$ exists with $x = x'|_{[0, \tau)}$.

We say that $x^* \in \mathbb{R}^n$ is an *equilibrium* of (12) if $x(t) \equiv x^*$ on some small enough non-degenerate interval is a Carathéodory solution to (12). In other words, if and only if $0 \in F(x^*)$. We say that (12) is (*Lyapunov stable*) at $x^* \in \mathbb{R}^n$ if, for every $\varepsilon > 0$, there exists some $\delta > 0$ such that, for every maximal Carathéodory solution $x(\cdot)$ of (12), we have $\|x_0 - x^*\| < \delta \implies \|x(t) - x^*\| < \varepsilon$ for every $t \geq 0$ in the interval where $x(\cdot)$ is defined. Note that, under Assumption 3, if (12) is stable at x^* , then x^* is an equilibrium of (12) (Bacciotti and Ceragioli, 1999). Furthermore, we say that (12) is (*locally and strongly*) *asymptotically stable* at $x^* \in \mathbb{R}^n$ if it is stable at x^* and there exists some $\delta > 0$ such that, for every maximal Carathéodory solution $x : [0, \tau) \rightarrow \mathbb{R}^n$ of (12), if $\|x_0 - x^*\| < \delta$ then $x(t) \rightarrow x^*$ as $t \rightarrow \tau$. Finally, (12) is (*locally and strongly*) *finite-time stable* at x^* if it is asymptotically stable and there exists some $\delta > 0$ and $T : B_\delta(x^*) \rightarrow [0, \infty)$ such that, for every maximal Carathéodory solution $x(\cdot)$ of (12) with $x_0 \in B_\delta(x^*)$, we have $\lim_{t \rightarrow T(x_0)} x(t) = x^*$.

We will now construct a Lyapunov-based criterion adapted from the literature of finite-time stability of Lipschitz continuous systems. To do this, we first adapt Lemma 1 in (Benosman et al., 2009) for absolutely continuous functions and non-positive exponents.

Lemma 1. *Let $\mathcal{E}(\cdot)$ be an absolutely continuous function satisfying the differential inequality*

$$\dot{\mathcal{E}}(t) \leq -c\mathcal{E}(t)^\alpha \quad (13)$$

a.e. in $t \geq 0$, with $c, \mathcal{E}(0) > 0$ and $\alpha < 1$. Then, there exists some $t^ > 0$ such that $\mathcal{E}(t) > 0$ for $t \in [0, t^*)$ and $\mathcal{E}(t^*) = 0$. Furthermore, $t^* > 0$ can be bounded by*

$$t^* \leq \frac{\mathcal{E}(0)^{1-\alpha}}{c(1-\alpha)}, \quad (14)$$

and this bound is exact provided (13) is exact as well. Finally, if (13) is exact and $\alpha \geq 1$, then $t^ = \infty$ with $\mathcal{E}(\infty) \triangleq \lim_{t \rightarrow \infty} \mathcal{E}(t)$.*

Proof. Suppose that $\mathcal{E}(t) > 0$ for every $t \in [0, T]$ with $T > 0$. Let t^* be the supremum of all such T 's, thus satisfying $\mathcal{E}(t) > 0$ for every $t \in [0, t^*)$. We will now investigate $\mathcal{E}(t^*)$. First, by continuity of \mathcal{E} , it follows that $\mathcal{E}(t^*) \geq 0$. Now, by rewriting

$$\dot{\mathcal{E}}(t) \leq -c\mathcal{E}(t)^\alpha \iff \frac{d}{dt} \left[\frac{\mathcal{E}(t)^{1-\alpha}}{1-\alpha} \right] \leq -c, \quad (15)$$

a.e. in $t \in [0, t^*)$, we can thus integrate to obtain

$$\frac{\mathcal{E}(t)^{1-\alpha}}{1-\alpha} - \frac{\mathcal{E}(0)^{1-\alpha}}{1-\alpha} \leq -ct, \quad (16)$$

everywhere in $t \in [0, t^*)$, which in turn leads to

$$\mathcal{E}(t) \leq [\mathcal{E}(0)^{1-\alpha} - c(1-\alpha)t]^{1-\alpha} \quad (17)$$

and

$$t \leq \frac{\mathcal{E}(0)^{1-\alpha} - \mathcal{E}(t)^{1-\alpha}}{c(1-\alpha)} \leq \frac{\mathcal{E}(0)^{1-\alpha}}{c(1-\alpha)}, \quad (18)$$

where the last inequality follows from $\mathcal{E}(t) > 0$ for every $t \in [0, t^*)$. Taking the supremum in (18) then leads to the upper bound (14). Finally, we conclude that $\mathcal{E}(t^*) = 0$, since $\mathcal{E}(t^*) > 0$ is impossible given that it would mean, due to continuity of \mathcal{E} , that there exists some $T > t^*$ such that $\mathcal{E}(t) > 0$ for every $t \in [0, T]$, thus contradicting the construction of t^* .

Finally, notice that if \mathcal{E} satisfies (13) exactly, then (17) and the first inequality in (18) are both exact as well. The exactness of the bound (14) thus follows immediately. Furthermore, notice that if $\alpha \geq 1$, and \mathcal{E} is an exact solution to (13), i.e. $\mathcal{E}(t) = [\mathcal{E}(0)^{1-\alpha} - c(1-\alpha)t]^{1-\alpha}$, then clearly $\mathcal{E}(t) > 0$ for every $t \geq 0$ and $\mathcal{E}(t) \rightarrow 0$ as $t \rightarrow \infty$. ■

In their Proposition 2.8, Cortés and Bullo (2005) proposed a Lyapunov-based criterion to establish finite-time stability of discontinuous systems, which fundamentally boils down to Lemma 1 with exponent $\alpha = 0$. This proposition was in turn based on Theorem 2 of Paden and Sastry (1987). Cortés (2006) later proposed a second-order Lyapunov criterion, which fundamentally boils down to $\mathcal{E}(t) = V(x(t))$ being strongly convex. Finally, Corollary 3.1 of Hui, Haddad, and Bhat (2009) generalized the aforementioned Proposition 2.8 of Cortés and Bullo (2005) to establish semistability. Indeed, these two results coincide for isolated equilibria.

We now present a novel result that generalizes the aforementioned first-order Lyapunov-based results, by exploiting our Lemma 1. More precisely, given a Lyapunov candidate function $V(\cdot)$, the objective is to set $\mathcal{E}(t) = V(x(t))$ and check that the conditions of Lemma 1 hold. To do this, and assuming V to be locally Lipschitz continuous, we first borrow and adapt from Bacciotti and Ceragioli (1999) the definition of *set-valued time derivative* of $V : \mathcal{D} \rightarrow \mathbb{R}$ w.r.t. the differential inclusion (12), given by

$$\dot{V}(x) \triangleq \{a \in \mathbb{R} : \exists v \in K(x) \text{ s.t. } a = p \cdot v, \forall p \in \partial V(x)\}, \quad (19)$$

for each $x \in \mathcal{D}$. Notice that, under Assumption 3 For Filippov differential inclusions, we have $K = \mathcal{K}[F]$, and the set-valued time derivative of V thus coincides with with the set-valued Lie derivative $\mathcal{L}_F V(\cdot)$. Indeed, more generally \dot{V} could be seen as a set-valued Lie derivative $\dot{V} = \mathcal{L}_K V$ w.r.t. the set-valued map K .

Definition 2. We say that $V(\cdot)$ is said to be *regular* if every directional derivative, given by

$$V'(x; v) \triangleq \lim_{h \rightarrow 0} \frac{V(x + h v) - V(x)}{h}, \quad (20)$$

exists and is equal to

$$V^\circ(x; v) \triangleq \limsup_{x' \rightarrow x, h \rightarrow 0^+} \frac{V(x' + h v) - V(x')}{h}, \quad (21)$$

known as *Clarke's upper generalized derivative* (Clarke, 1981).

Assumption 4. V is locally Lipschitz continuous and regular.

In practice, regularity is a fairly mild and easy to guarantee condition. For instance, it would suffice that V is convex or continuously differentiable to ensure that it is Lipschitz and regular.

We are now equipped to formally establish the correspondence between the set-valued time-derivative of V and the derivative of the energy function $\mathcal{E}(t) = V(x(t))$ associated with an arbitrary Carathéodory solution $x(\cdot)$ to the differential inclusion (12).

Lemma 2 (Lemma 1 of Bacciotti and Ceragioli (1999)). *Under Assumption 4, given any Carathéodory solution $x : [0, \tau) \rightarrow \mathbb{R}^n$ to (12), $\mathcal{E}(t) \triangleq V(x(t))$ is absolutely continuous and $\dot{\mathcal{E}}(t) = \frac{d}{dt} V(x(t)) \in \dot{V}(x(t))$ a.e. in $t \in [0, \tau)$.*

We are now ready to state and prove our Lyapunov-based sufficient condition for finite-time stability of differential inclusions.

Theorem 2. *Suppose that Assumptions 3 and 4 hold for some set-valued map $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and some function $V : \mathcal{D} \rightarrow \mathbb{R}$, where $\mathcal{D} \subseteq \mathbb{R}^n$ is an open and positively invariant neighborhood of a point $x^* \in \mathbb{R}^n$. Suppose that V is positive definite w.r.t. x^* and that there exist constants $c > 0$ and $\alpha < 1$ such that*

$$\sup \dot{V}(x) \leq -c V(x)^\alpha \quad (22)$$

a.e. in $x \in \mathcal{D}$. Then, (12) is finite-time stable at x^ , with settling time upper bounded by*

$$t^* \leq \frac{V(x_0)^{1-\alpha}}{c(1-\alpha)}, \quad (23)$$

where $x(0) = x_0$. In particular, any Carathéodory solution (in the sense of Filippov) $x(\cdot)$ with $x(0) = x_0 \in \mathcal{D}$ will converge in finite time to x^ under the upper bound (23). Furthermore, if $\mathcal{D} = \mathbb{R}^n$, then (12) is globally finite-time stable. Finally, if $\dot{V}(x)$ is a singleton a.e. in $x \in \mathcal{D}$ and (22) is exact, then so is (23).*

Proof. Note that, by Proposition 1 of Bacciotti and Ceragioli (1999), we know that (12) is Lyapunov stable at x^* . All that remains to show is local convergence towards x^* (which must be an equilibrium)

in finite time. Indeed, given any maximal solution $x : [0, t^*) \rightarrow \mathbb{R}^n$ to (12) with $x(0) = x_0 \neq x^*$, we know by Lemma 2, that $\mathcal{E}(t) = V(x(t))$ is absolutely continuous with $\dot{\mathcal{E}}(t) \in \dot{V}(x(t))$ a.e. in $t \in [0, t^*]$. Therefore, we have

$$\dot{\mathcal{E}}(t) \leq \sup \dot{V}(x(t)) \leq -cV(x(t))^\alpha = -c\mathcal{E}(t)^\alpha \quad (24)$$

a.e. in $t \in [0, t^*]$. Since $\mathcal{E}(0) = V(x_0) > 0$, given that $x_0 \neq x^*$, the result then follows by invoking Lemma 1 and noting that $\mathcal{E}(t^*) = 0 \iff V(t^*, x(t^*)) = 0 \iff x(t^*) = x^*$. \blacksquare

C Proof of the main result: Theorem 1

Let us focus first on (3), since the proof for (4) follows similar steps. The idea is to show that it is finite-time stable at x^* , with the inequality in Theorem 2 holding exactly for $V(x) = \|\nabla f(x)\|^2$. First, notice that $F(x) \triangleq -c\|\nabla f(x)\|^p \frac{[\nabla^2 f(x)]^r \nabla f(x)}{\nabla f(x)^\top [\nabla^2 f(x)]^{r+1} \nabla f(x)}$ is continuous near (but not at) $x = x^*$, and undefined at $x = x^*$ itself. Furthermore, we have

$$\|F(x)\| = c\|\nabla f(x)\|^p \frac{\|[\nabla^2 f(x)]^r \nabla f(x)\|}{\nabla f(x)^\top [\nabla^2 f(x)]^{r+1} \nabla f(x)} \quad (25a)$$

$$\leq c\|\nabla f(x)\|^p \frac{\lambda_{\max}(\nabla^2 f(x))^r \|\nabla f(x)\|}{\lambda_{\min}(\nabla^2 f(x))^{r+1} \|\nabla f(x)\|^2} \quad (25b)$$

$$\leq c \frac{\lambda_{\max}(\nabla^2 f(x))^r}{\lambda_{\min}(\nabla^2 f(x))^{r+1}} \|\nabla f(x)\|^{p-1}, \quad (25c)$$

with $p-1 \geq 0$ and $\lambda_{\min}[\nabla^2 f(x)] \geq m > 0$ everywhere near $x = x^*$ for some $m > 0$ (m -strong convexity). Therefore, F is Lebesgue integrable (and thus measurable) and locally essentially bounded, which means that Assumption 2 is satisfied.

Set $V(x) \triangleq \|\nabla f(x)\|^2$, defined over \mathcal{D} . If \mathcal{D} is not positively invariant w.r.t. (3), we can always replace it by a smaller open subset that is, e.g. a sufficiently small strict sublevel set contained within \mathcal{D} . Clearly, V is continuously differentiable, thus satisfying Assumption 4. Furthermore, it is positive definite w.r.t. x^* and, given $x \in \mathcal{D} \setminus \{x^*\}$, we have

$$\sup \dot{V}(x) = \sup \{a \in \mathbb{R} : \exists v \in \mathcal{K}[F](x) \text{ s.t. } a = p \cdot v, \forall p \in \partial V(x)\} \quad (26a)$$

$$= \sup \left\{ \left(2\nabla^2 f(x) \nabla f(x) \right) \cdot \left(-c\|\nabla f(x)\|^p \frac{[\nabla^2 f(x)]^r \nabla f(x)}{\nabla f(x)^\top [\nabla^2 f(x)]^{r+1} \nabla f(x)} \right) \right\} \quad (26b)$$

$$= -2c\|\nabla f(x)\|^p \quad (26c)$$

$$= -(2c)V(x)^{\frac{p}{2}} \quad (26d)$$

with $\frac{p}{2} < 1$. Furthermore, $\dot{V}(x^*) = \{\nabla V(x^*) \cdot v : v \in \mathcal{K}[F](x^*)\} = \{0\}$ since $\nabla V(x^*) = 2\nabla^2 f(x^*)\nabla f(x^*) = 0$. The result thus follows by invoking Theorem 2.

We now proceed to establish finite-time stability of (4) at x^* . Like before, we notice that $F(x) = -c\|\nabla f(x)\|_1^{p-1} x \frac{[\nabla^2 f(x)]^r \text{sign}(\nabla f(x))}{\text{sign}(\nabla f(x))^\top [\nabla^2 f(x)]^{r+1} \text{sign}(\nabla f(x))}$ is continuous near, but not at, $x = x^*$. Furthermore, notice that

$$\|F(x)\| = c\|\nabla f(x)\|_1^{p-1} \frac{\|[\nabla^2 f(x)]^r \text{sign}(\nabla f(x))\|}{\text{sign}(\nabla f(x))^\top [\nabla^2 f(x)]^{r+1} \text{sign}(\nabla f(x))} \quad (27a)$$

$$\leq c\|\nabla f(x)\|_1^{p-1} \frac{\lambda_{\max}(\nabla^2 f(x))^r \|\text{sign}(\nabla f(x))\|}{\lambda_{\min}(\nabla^2 f(x))^{r+1} \|\text{sign}(\nabla f(x))\|^2} \quad (27b)$$

$$\leq c \frac{\lambda_{\max}(\nabla^2 f(x))^r}{\lambda_{\min}(\nabla^2 f(x))^{r+1}} \frac{\|\nabla f(x)\|_1^{p-1}}{\|\text{sign}(\nabla f(x))\|} \quad (27c)$$

$$\leq c \frac{\lambda_{\max}(\nabla^2 f(x))^r}{\lambda_{\min}(\nabla^2 f(x))^{r+1}} \frac{\|\nabla f(x)\|_1^{p-1}}{\|\text{sign}(\nabla f(x))\|_1 / \sqrt{n}} \quad (27d)$$

$$\leq \frac{c}{\sqrt{n}} \frac{\lambda_{\max}(\nabla^2 f(x))^r}{\lambda_{\min}(\nabla^2 f(x))^{r+1}} \|\nabla f(x)\|_1^{p-1} \quad (27e)$$

$$(27f)$$

for every $x \in \mathcal{D} \setminus \mathcal{N}$, with $\mathcal{N} = \bigcup_{i=1}^n \left\{ x \in \mathcal{D} : \frac{\partial f}{\partial x_i}(x) = 0 \right\}$. Since \mathcal{N} is a zero-measure set due to being a finite union of hypersurfaces in \mathbb{R}^n ($n \geq 1$), and also recalling that f is strongly convex near x^* and $p - 1 \geq 0$, it follows that F is Lebesgue integrable and locally essentially bounded. Therefore, Assumption 2 is once again satisfied.

Now consider the candidate Lyapunov function $V(x) = \|\nabla f(x)\|_1$, defined over \mathcal{D} . Clearly, V is not continuously differentiable this time. However, it still satisfies Assumption 2 due to being a.e. differentiable. In particular, we have $\partial V(x) = \{\nabla^2 f(x) \text{sign}(\nabla f(x))\}$ for every $x \in \mathcal{D} \setminus \mathcal{N}$. In other words, we have $\partial V(x) = \{\nabla^2 f(x) \text{sign}(\nabla f(x))\}$ a.e. in $x \in \mathcal{D}$.

Given $x \in \mathcal{D} \setminus \mathcal{N}$, we thus have

$$\sup \dot{V}(x) = \sup\{a \in \mathbb{R} : \exists v \in \mathcal{K}[F](x) \text{ s.t. } a = p \cdot v, \forall p \in \partial V(x)\} \quad (28a)$$

$$= \left(\nabla^2 f(x) \text{sign}(\nabla f(x)) \right) \cdot \left(-\frac{c \|\nabla f(x)\|_1^{p-1} [\nabla^2 f(x)]^r \text{sign}(\nabla f(x))}{\text{sign}(\nabla f(x))^\top [\nabla^2 f(x)]^{r+1} \text{sign}(\nabla f(x))} \right) \quad (28b)$$

$$= -c \|\nabla f(x)\|_1^{p-1} \quad (28c)$$

$$= -c V(x)^{p-1}, \quad (28d)$$

with $p - 1 < 1$. The result once again follows by invoking Theorem 2.