MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Time-Varying Continuous-Time Optimization with Pre-Defined Finite-Time Stability

Romero, Orlando; Benosman, Mouhacine

TR2020-088 July 01, 2020

Abstract

In this paper we propose a new family of continuous-time optimization algorithms based on discontinuous second order gradient optimization flows, with finite-time convergence guarantees to local optima, for locally strongly convex (time-varying) cost functions. To analyze our flows, we first extend a well-know Lyapunov inequality condition for finite-time stability, to the case of (time-varying) differential inclusions. We then prove the convergence of these second-order flows in finite-time. In some particular cases, we can show that the finite-time convergence can be pre-defined by the user. We propose a robustification of the flows to bounded additive uncertainties, and extend some of the results to the case of constrained optimization. We show the performance of these flows on well-know optimization benchmarks, namely, the Rosenbrock function, and the Rastringin function.

International Journal of Control

Time-Varying Continuous-Time Optimization with Pre-Defined Finite-Time Stability

Orlando Romero† Mouhacine Benosman‡

ARTICLE HISTORY
Compiled April 13, 2020

ABSTRACT
In this paper we propose a new family of continuous-time optimization algorithms based on discontinuous second order gradient optimization flows, with finite-time convergence guarantees to local optima, for locally strongly convex (time-varying) cost functions. To analyze our flows, we first extend a well-know Lyapunov inequality condition for finite-time stability, to the case of (time-varying) differential inclusions. We then prove the convergence of these second-order flows in finite-time. In some particular cases, we can show that the finite-time convergence can be *pre-defined by the user*. We propose a robustification of the flows to bounded additive uncertainties, and extend some of the results to the case of constrained optimization. We show the performance of these flows on well-know optimization benchmarks, namely, the Rosenbrock function, and the Rastringin function.

1. Introduction

In continuous-time optimization, an ordinary differential equation (ODE) or a partial differential equation (PDE) is designed in such a way that its solution convergences over time to an optimal value of the cost function. There has been a recent surge in research papers in this direction, arguably starting with the pioneer work by Brockett, in Brockett (1988), e.g., Ariyur & Krstić (2003); (2015); Attouch et al. (2015, 2018); Cortes (2006); Faybusovich (1991); Franca et al. (2019a,b); Franka et al. (2018); Guay& Zhang (2003); Grushkovskaya et al. (2018); Helmke & Moore (1996); Krstić (2000); Poveda & Teel (2017); Scieur et al. (2017); Scheinker & Krstić (2016); Su et al. (2016); Wang & Elia (2011); Wang & Lu (2017); Wilson et al. (2016); Wang & Elia (2010); Zhang & Ordónez (2012); Zhang et al. (2018), Grune & Karafyllis (2013); Karafyllis & Krstic (2017); Karafyllis (2014)

An important class of continuous optimization algorithms are the so-called extremum seeking (ES) controllers, which deal with static cost functions, as well as dynamic cost functions, modeled as the output of a dynamical system. Most importantly, ES algorithms are often based only on the cost function measurements, i.e., zero-order optimization methods, whereas the higher order derivatives of the cost function, e.g., gradient and Hessian, are estimated from the cost function measurements using feed-

†Department of Industrial and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA, E-mail: rodrio2@rpi.edu.
‡Mitsubishi Electrics Research Laboratories, Cambridge, MA 02139, USA, E-mail: m_benosman@ieee.org.

back filters, e.g., Ariyur & Krstić (2003); Guay& Zhang (2003); Grushkovskaya et al. (2018); Krstić (2000); Poveda & Teel (2017); Scheinker & Krstić (2016); Zhang & Ordónez (2012). Since we are not considering zero-order methods in this work, we will not discuss specifically ES results, and we will focus on the more general class of continuous-optimization algorithms, including higher order methods.

For instance, in Su et al. (2016), the authors derive a second-order ODE as the limit of Nesterov's accelerated gradient method, when the gradient step sizes go to zero. This ODE is then used to attempt to analyze Nesterov's scheme, particularly in an larger effort to better understand acceleration without substantially increasing computational burden. Thanks to the ODE continuous-time approximation of the algorithm, the authors also obtain a family of schemes with similar convergence rates as Nesterov's algorithm.

In Franka et al. (2018), The differential equations that model the continuous-time limit of the sequence of iterates generated by the alternating direction method of multipliers (ADMM), are derived. Then, the authors employ Lyapunov theory to analyze the stability of critical points of the dynamical systems and to obtain associated convergence rates.

In Franca et al. (2019a), non-smooth and linearly constrained optimization problems are analyzed by deriving equivalent (at the limit) non-smooth dynamical systems related to variants of the relaxed and accelerated ADMM. In particular, two new ADMM-like algorithms are proposed, one based on Nesterov's acceleration and the other inspired by Polyak's heavy ball method, and derive differential inclusions modeling these algorithms in the continuous-time limit. Using a non-smooth Lyapunov analysis, results on rate-of-convergence are obtained for these dynamical systems in the convex and strongly convex setting.

In Franca et al. (2019b), the authors study the crucial problem of structure-preserving discretizations of continuous-time optimization flows. More specifically, the authors focus on two classes of conformal Hamiltonian systems whose trajectories lie on a symplectic manifold, namely a classical mechanical system with linear dissipation and its relativistic extension. One of the most noticeable claims in this paper is that conformal symplectic integrators can preserve convergence rates of the continuous-time system up to a negligible error. As a by product of this, the authors show that the classical momentum method is a symplectic integrator. Finally, a relativistic generalization of classical momentum called *relativistic gradient descent* is introduced, and it is argued that it may result in more stable/faster optimization for some optimization problems.

In Cortes (2006), two normalized first-order gradient flows are proposed. Their convergence is rigorously analyzed using tools from non-smooth dynamics theory, and conditions guaranteeing finite-time convergence are derived. Finally, the proposed non-smooth flows are applied to problems in multi-agent systems and it is shown they achieve consensus in a finite-time. The finite convergence time's upper bound is given as function of the gradient value at the initial point as well as the minimum eigenvalue of Hessian at the initial point.

More recently, in Poveda & Li (2019), the authors establish uniform asymptotic stability and robustness properties for the continuous-time limit of the Nesterov's accelerated gradient method, by using resetting mechanisms that are modeled by well-posed hybrid dynamical systems.

In Karafyllis & Krstic (2017); Karafyllis (2014), the authors propose a new family of dynamical systems to solve several nonlinear programing (NLP) problems. Feedback stabilization methods are used for the explicit construction of interior-point dynamical NLP solvers in Karafyllis (2014), and exterior-point dynamical NLP solvers in Karafyllis & Krstic (2017). These dynamical systems are derived from an extension of the control Lyapunov function methodology, based on new extensions of LaSalle's theorem. The cases of equality as well as inequality constraints are treated, and the proposed flows are proven to lead to asymptotic, and in some cases exponential, convergence results to strict local minima.

In this work, we want to focus on the specific class of continuous-time optimization algorithms with finite-time convergence, for static as well as *time-varying cost functions*. We propose a new family of *discontinuous second-order flows*, which guarantee local convergence to an optimum, in a *desired pre-defined finite-time*. We use some ideas from Lyapunov-based finite-time state control to an invariant set, proposed by one of the current authors in an early paper Benosman & Lum (2009), in the context of aerospace applications, to design a new family of discontinuous flows, which ensure a desired finite-time convergence to the invariant set containing a unique local optima. Furthermore, due to the discontinuous nature of the proposed flows, we propose to extend one of the existing Lyapunov-based inequality condition for finite-time convergence of continuous-time dynamical systems, to the case of differential inclusions. *We also propose a robustification of these flows w.r.t. time-varying bounded additive uncertainties.* Finally, we extend part of the results to the case of constrained optimization, by using some recent results from barrier Lyapunov functions control theory, e.g., Liu & Tong (2016); Yang et al. (2019). The proposed continuous-time optimization algorithms are tested on well-known optimization testbeds, namely, the Rosenbrock function, and the Rastringin function.

This paper is organized as follows: Section 2 is dedicated to recalling some preliminaries about continuous-time optimization and finite-time stability in the context of differential inclusions. Our main results are presented in Section 3, where we first establish an extension to (time-varying) differential inclusions of a well-know Lyapunov-based inequality condition for finite-time stability. We then propose and analyze our second-order discontinuous flows, including a flow for time-varying cost functions, and its robustification w.r.t. additive uncertainties. Finally, we extend these results to the case of constrained optimization, using ideas from barrier Lyapunov function control theory. In Section 4, we show the efficiency of this continuous-time optimization flow on some well established optimization benchmarks. The paper ends with a summarizing conclusion and a discussion of our ongoing investigations, in Section 5.

2. Preliminaries

Consider an unconstrained nonlinear optimization problem of the form

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the nonlinear objective function. One of the most popular numerical schemes to solve (1) is through the *gradient descent* algorithm, given by

$$x_{k+1} = x_k + \eta_k \nabla f(x_k), \tag{2}$$

for $k \in \mathbb{Z}_+ = \{0, 1, 2, \ldots\}$, where $\nabla(\cdot)$ denotes the gradient operator and $\eta_k > 0$ denote the *step sizes* (also known as the *learning rate*), which are usually chosen as small values. It is well-known that, provided that the objective function is sufficiently regular (*e.g.* twice continuously differentiable), the initial approximation $x_0 \in \mathbb{R}^n$ is sufficiently close to a sufficiently regular local minimizer (*e.g.* strict local minimum and isolated stationary point), and the step sizes are sufficiently small (*e.g.* smaller than the inverse of a Lipschitz constant of the gradient of f), then the sequence $\{x_k\}$ given by (2) will converge to that local minimum under a sublinear convergence rate.

The autonomous (time-invariant) state-space dynamical system known as the *gradient flow*, given by

$$\dot{x}(t) = -\nabla f(x(t)) \tag{3a}$$

$$x(0) = x_0, \tag{3b}$$

serves as a "smoothed" continuous-time variant of the gradient descent algorithm (2). From now on, the independent time variable t will be left implicit in $x(t)$, except when it could lead to ambiguity. Convergence of (3a) can be established by proving that sufficiently regular local minima are locally assymptotically stable equilibria of the dynamical system. This, in turn, can be readily done through a suitable Lyapunov function such as $V(x) = \frac{1}{2}\|x - x^\star\|^2$, $V(x) = f(x) - f(x^\star)$ or $V(x) = \frac{1}{2}\|\nabla f(x)\|^2$, where $x^\star = 0$ denotes one of the aforementioned local minima, e.g., Benosman & Lum (2009); Cortes (2006).

2.1. Filippov Differential Inclusion for Time-Invariant Systems

First recall that a solution to an initial value problem

$$\dot{x}(t) = F(x(t)) \tag{4a}$$

$$x(0) = x_0 \tag{4b}$$

with $F : \mathbb{R}^n \to \mathbb{R}^n$ can only be garanteed to exist and be unique if $F(\cdot)$ is Lipschitz continuous. When $F(\cdot)$ is not Lipschitz continuous (*e.g.* due to singularities or discontinuities), but nevertheless assuming it to be Lebesgue measurable and locally essentially bounded, we understand (4a) as the differential inclusion

$$\dot{x}(t) \in \mathcal{K}[F](x(t)), \tag{5}$$

almost everywhere (a.e.) in $t \geq 0$, with $x(\cdot)$ absolutely continuous. More precisely, $\mathcal{K}[F](\cdot)$ denotes the Filippov set-valued map Paden & Sastry (1987) given by

$$\mathcal{K}[F](x) \stackrel{\text{def}}{=} \bigcap_{\delta>0} \bigcap_{\tilde{\mu}(S)=0} \overline{\text{co}}(F(B_\delta(x) \setminus S)), \tag{6}$$

where $\tilde{\mu}(\cdot)$ denotes the Lebesgue measure and $\overline{\text{co}}(\cdot)$ the convex closure (*i.e.* closure of the convex hull). In Theorem 1 of Paden & Sastry (1987), the authors proved that, if F is locally bounded, then (11) can be computed as

$$\mathcal{K}[F](x) = \left\{ \lim_{k\to\infty} F(x_k) : x_k \notin \mathcal{N}_F \cup S, x_k \to x \right\} \tag{7}$$

for some set $\mathcal{N}_F \subset \mathbb{R}^n$ of measure zero and any other set $S \subset \mathbb{R}^n$ of measure zero. In particular, if F is continuous at a fixed x, then $\mathcal{K}[F](x) = \{F(x)\}$. For instance, for the gradient flow (3a) 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 Lipschitz continuous, then $\mathcal{K}[-\nabla f](x) = -\partial f(x)$, where $\partial f(\cdot)$ denotes Clarke's generalized gradient; see Clarke (2001).

2.1.1. Finite-Time Stability for Time-Invariant Differential Inclusions

Consider a general differential inclusion Bacciotti & Ceragioli (1999)

$$\dot{x}(t) \in K(x(t)) \tag{8a}$$

$$x(0) = x_0 \tag{8b}$$

where $K : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is a set-valued map [1], assumed to be upper semi-continuous with compact and convex values. In Filippov & Arscott (1988), the authors proved that $K(\cdot) = \mathcal{K}[F](\cdot)$ is indeed upper semi-continuous, with nonempty, compact, and convex values.

We say that $x : [0, \tau] \to \mathbb{R}^n$ with $\tau > 0$ is a solution to (8) if $x(\cdot)$ is absolutely continuous on any closed subinterval of $[0, \tau]$, (8a) is satisfied a.e. in $t \in [0, \tau]$, and $x(0) = x_0$. We say that $x^\star \in \mathbb{R}^n$ is an *equilibrium* of (8) if $x(t) = x^\star$ on some small enough non-degenerate interval is a solution to (8). In other words, if and only if $0 \in K(x^\star)$.

We say that an equilibrium point $x^\star \in \mathbb{R}^n$ of (8) is *Lyapunov stable* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that, for every solution $x(\cdot)$ of (8), we have $\|x_0 - x^\star\| < \delta \implies \|x(t) - x^\star\| < \varepsilon$ for every $t \geq 0$ in the interval where $x(\cdot)$ is defined. Furthermore, we say that $x^\star \in \mathbb{R}^n$ is *(locally) asymptotically stsble* if it is Lyapunov stable and there exists some $\delta > 0$ such that, for every solution $x(\cdot)$ of (8), if $\|x_0 - x^\star\| < \delta$ then $x(t)$ will converge to $x^\star$. Finally, $x^\star \in \mathbb{R}^n$ is said to be *(locally) finite-time stable* if it is asymptotically stable and there exists some $\delta > 0$ and $T : B_\delta(x^\star) \setminus \{x^\star\} \to (0, \infty)$ such that, for every solution $x(\cdot)$ of (8) with $x_0 \in B_\delta(x^\star) \setminus \{x^\star\}$, we have $x(t) \in B_\delta(x^\star) \setminus \{x^\star\}$ for every $t \in [0, T(x_0))$ and $x(t) \to x^\star$ as $t \to T(x_0)$.

2.2. Filippov Differential Inclusion for Time-Variant Systems

Similarly to the time-invariant case, a solution to an initial value problem

$$\dot{x}(t) = F(t, x(t)) \tag{9a}$$

$$x(0) = x_0 \tag{9b}$$

with $F : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ is typically guaranteed to exist and be unique by ensuring that $F(\cdot, x)$ is continuous near $x = x^\star$ and $F(t, \cdot)$ is Lipschitz continuous near $t = 0$. When $F(t, \cdot)$ is not Lipschitz continuous (*e.g.* due to singularities or discontinuities), we understand solutions to (9a) in the sense of Filippov. More precisely, $x : [0, \tau) \to \mathbb{R}^n$ with $0 < \tau \leq \infty$ is a *Filippov solution* to (9) if it is absolutely continuous, $x(0) = x_0$, and

$$\dot{x}(t) \in \mathcal{K}[F](t, x(t)) \tag{10}$$

[1] 2^X denotes the power set of a set X

holds almost everywhere (a.e.) within every compact subinterval of $[0, \tau)$, where $\mathcal{K}[F]$ denotes the Filippov set-valued map Cortes (2008); Paden & Sastry (1987) given by

$$\mathcal{K}[F](t, x) \triangleq \bigcap_{\delta>0} \bigcap_{\mu(S)=0} \overline{\text{co}}\, F(t, B_\delta(x) \setminus S), \tag{11}$$

where μ denotes the Lebesgue measure and $\overline{\text{co}}$ the convex closure. Furthermore, $x(\cdot) : [0, \tau) \to \mathbb{R}^n$ is a *maximal* Filippov solution if it cannot be extended, *i.e.* if no Filippov solution exists over an interval $[0, \tau')$ with $\tau' > \tau$.

Assumption 1. F is Lebesgue measurable and locally essentially bounded, *i.e.* given any (t, x), F is bounded a.e. on every bounded neighborhood of (t, x).

Under Assumption 1, at least one Filippov solution to (9) must exist Cortes (2008); Paden & Sastry (1987). Furthermore, the Filippov set-valued map (11) can be computed as

$$\mathcal{K}[F](t, x) = \overline{\text{co}} \left\{ \lim_{k\to\infty} F(t, x_k) : \mathcal{N}_F \cup S \not\ni x_k \to x \right\} \tag{12}$$

for some set $\mathcal{N}_F \subset \mathbb{R}^n$ of measure zero and any other set $S \subset \mathbb{R}^n$ of measure zero. In particular, if $F(t, \cdot)$ is continuous at a fixed point x, then $\mathcal{K}[F](t, x) = \{F(t, x)\}$. For instance, for the gradient flow, we have $\mathcal{K}[-\nabla f](t, x) = \{-\nabla f(x)\}$ for every $x \in \mathbb{R}^n$, provided that f is continuously differentiable. Furthermore, if f is only Lipschitz continuous, then $\mathcal{K}[-\nabla f](t, x) = -\partial f(x)$, where ∂f denotes Clarke's generalized gradient Clarke (2001).

2.2.1. Finite-Time Stability for Time-Variant Differential Inclusions

Consider a general time-varying differential inclusion Bacciotti & Ceragioli (1999)

$$\dot{x}(t) \in K(t, x(t)) \tag{13a}$$
$$x(0) = x_0 \tag{13b}$$

where $K : \mathbb{R}_+ \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is an arbitrary set-valued map.

Assumption 2. $K : \mathbb{R}_+ \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is an upper semi-continuous set-valued map, with nonempty, compact, and convex values.

For instance, in Filippov & Arscott (1988) the authors proved that, under Assumption 1, $K = \mathcal{K}[F]$ satisfies Assumption 2.

We say that $x : [0, \tau) \to \mathbb{R}^n$ with $0 < \tau \leq \infty$ is a *Carathéodory solution* to (13) if $x(\cdot)$ is absolutely continuous on any closed subinterval of $[0, \tau)$, (13a) is satisfied a.e. within every compact subinterval of $[0, \tau)$, and $x(0) = x_0$.

Proposition 1. *Under Assumption 2, at least one Carathéodory solution to* (13) *must exist. In particular, under Assumption 1, at least one Filippov solution to* (9) *must exist.*

We say that $x : [0, \tau) \to \mathbb{R}^n$ is a *maximal* Carathéodory solution of (13) if it cannot be extended, *i.e.* if no solution exists over an interval $[0, \tau')$ with $\tau' > \tau$. In particular, (maximal) Filippov solutions to (9) are nothing but (maximal) Carathéodory solutions to the Filippov differential inclusion (10) with initial condition $x(0) = x_0$.

Furthermore, we say that $x^\star \in \mathbb{R}^n$ is an *equilibrium* of (13) if $x(t) \equiv x^\star$ over $(0, \infty)$ is a Carathéodory solution to (13). In other words, if $0 \in K(t, x^\star)$ holds a.e. in $t \geq 0$. We say that (13) is *(strongly) Lyapunov stable* at $x^\star \in \mathbb{R}^n$ if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that, for every Carathéodory solution $x(\cdot)$ of (13), we have $\|x_0 - x^\star\| < \delta \implies \|x(t) - x^\star\| < \varepsilon$ for every $t \geq 0$ in the interval where $x(\cdot)$ is defined. Furthermore, we say that (13) is *(locally and strongly) asymptotically stable* at $x^\star \in \mathbb{R}^n$ if it is Lyapunov stable at $x^\star$ and there exists some $\delta > 0$ such that every maximal Carathéodory solution $x(\cdot)$ to (13) is defined over $[0, \infty)$ and, if $\|x_0 - x^\star\| < \delta$ then $x(t) \to x^\star$ as $t \to \infty$. Finally, we say that (13) is *(locally and strongly) finite-time stable* at $x^\star \in \mathbb{R}^n$ if it is asymptotically stable at $x^\star$ and there exist some $\delta > 0$ and positive definite function (w.r.t. $x^\star$) $T : B_\delta(x^\star) \to \mathbb{R}_+$ (called the *settling time*) such that, for every Carathéodory solution $x(\cdot)$ of (13) with $x_0 \in B_\delta(x^\star) \setminus \{x^\star\}$, we have $x(t) \in B_\delta(x^\star) \setminus \{x^\star\}$ for every $t \in [0, T(x_0))$ and $x(t) \to x^\star$ as $t \to T(x_0)$

3. Main results

To establish finite-time stability, first we will propose an extension to the case of (time-variant) differential inclusions of a well-know Lyapunov-based result for the case of systems of the form (9), with Lipschitz continuous flow $F(.)$, e.g., see (Lemma 1 in Benosman & Lum (2009)). Next, we will use these results to analyze the stability of our discontinuous gradient-like flows for continuous-time optimization.

Theorem 1 (Finite-time stability condition for time-invariant differential inclusions)**.** *Let $x^\star \in \mathbb{R}^n$ be an equilibrium point of* (8) *and let $V : \mathcal{D} \to \mathbb{R}$ be a continuously differentiable and positive definite function w.r.t. $x^\star$, where $\mathcal{D} \subset \mathbb{R}^n$ is an open and positively invariant neighborhood of $x^\star$. Suppose that $K(x)$ is nonempty a.e. in $x \in \mathcal{D}$. Let*

$$\dot{V}(x) = \{\nabla V(x) \cdot v : v \in K(x)\},\ \forall t > 0,\ \forall x \in \mathcal{D} \tag{14}$$

If there exist constants $c > 0$ and $\alpha \in (0, 1)$ such that

$$\sup \dot{V}(x) \leq -c[V(x)]^\alpha \tag{15}$$

a.e. in $x \in \mathcal{D}$, then $x(t) \to x^\star$ in finite-time for every solution $x(\cdot)$ of (8) *with $x_0 \in \mathcal{D}$, and the settling time $t^\star$ is upper bounded by*

$$t^\star \leq \frac{V(x_0)^{1-\alpha}}{c(1-\alpha)}. \tag{16}$$

Furthermore, if $\dot{V}(x)$ contains a single point a.e. in $x \in \mathcal{D}$ and (15) *is exact, then so is* (16)*.*

Proof 1. *Since V is continuously differentiable, then $\dot{V}(x) = \{\nabla V(x) \cdot v : v \in K(x)\}$. Therefore, a.e. in $x \in \mathcal{D}$, we have*

$$\nabla V(x) \cdot v \leq -c\, V(x)^\alpha \tag{17}$$

for every $v \in K(x)$. In particular, given a solution $x(\cdot)$ of (8) with $x(0) = x_0 \in \mathcal{D}$, we have $x(t) \in \mathcal{D}$, and thus

$$\nabla V(x(t)) \cdot \dot{x}(t) \leq -c\, V(x(t))^{\alpha} \tag{18}$$

a.e. in $t \geq 0$. Notice that, since $x(\cdot)$ is absolutely continuous and $V(\cdot)$ is continuously differentiable (and thus Lipschitz continuous), then the composition $t \mapsto V(x(t))$ is absolutely continuous as well. Therefore, (18) can be rewritten as

$$\frac{d}{dt}\left[\frac{V(x(t))^{1-\alpha}}{1-\alpha}\right] \leq -c \tag{19}$$

a.e. in $t \geq 0$. Therefore, integrating (19), we find that

$$\frac{V(x(t))^{1-\alpha}}{1-\alpha} - \frac{V(x_0)^{1-\alpha}}{1-\alpha} \leq -ct \tag{20}$$

everywhere in $t \geq 0$ (not just a.e.). The result follows by setting $x(t^\star) = x^\star$ and rearranging the terms. Finally, in the case of $\dot{V}(x)$ containing a single point a.e. in $x \in \mathcal{D}$ and (15) being exact, then (17) through (20) are exact as well, which leads to (16) being exact.

We will now consider time-varying differential inclusions (13).

Theorem 2 (Finite-time stability condition for time-variant differential inclusions)**.** *Let $x^\star \in \mathbb{R}^n$ be an equilibrium point of (13) and let $V : \mathbb{R}_+ \times \mathcal{D} \to \mathbb{R}$ be a continuously differentiable and positive definite function w.r.t. $x^\star$, where $\mathcal{D} \subset \mathbb{R}^n$ is an open and positively invariant neighborhood of $x^\star$. Suppose that $K(t,x) = \mathcal{K}[F(t,\cdot)](x)$ is nonempty for every $x \in \mathcal{D}$. Let*

$$\dot{V}(t,x) \stackrel{\text{def}}{=} \left\{\frac{\partial V}{\partial t}(t,x) + \nabla V(t,x) \cdot v : v \in K(t,x)\right\} \tag{21}$$

for $t \geq 0$ and $x \in \mathcal{D}$, where $\nabla V(t,x)$ denotes the gradient of $V(t,x)$ w.r.t x. If there exist constants $c > 0$ and $\alpha \in (0,1)$ such that

$$\sup \dot{V}(t,x) \leq -c[V(t,x)]^{\alpha} \tag{22}$$

a.e. in $t \geq 0$ and $x \in \mathcal{D}$, then $x(t) \to x^\star$ in finite-time for every solution $x(\cdot)$ of (13) with $x_0 \in \mathcal{D}$, and the settling time $t^\star$ is upper bounded by

$$t^\star \leq \frac{V(0,x_0)^{1-\alpha}}{c(1-\alpha)}. \tag{23}$$

Furthermore, if $\dot{V}(t,x)$ contains a single point a.e. in $x \in \mathcal{D}$ and (22) is exact, then so is (23).

Proof 2. *The proof follows the same basic reasoning of the proof conducted for Theorem 1. Indeed, since $t \mapsto V(t, x(t))$ is absolutely continuous (Appendix, Lemma 3) due* *to $V(\cdot)$ being continuously differentiable, from* (22) *we note that*

$$\frac{d}{dt}V(t,x(t)) \leq -cV(t,x(t))^{\alpha}, \tag{24}$$

a.e. in $t \geq 0$, for every solution $x(\cdot)$ of (13)*. The rest of the proof follows by integrating and setting $x(t^\star) = x^\star$.*

We will now present a general inequality condition for time-varying flows as in (13), to design finite-time optimization dynamics. Subsequently, we will introduce several flows that satisfy this condition.

First, let us state a basic assumption on the cost function.

Assumption 3. $f : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable in both variables, with $f(t,.)$ strongly convex (respectively, strongly concave) in a convex open set $\mathcal{D} \subset \mathbb{R}^n$, and $x^{opt} : \mathbb{R}_+ \to \mathcal{D}$ s.t., for each t, $x^{opt}(t)$ is a strict local minimizer (respectively, maximizer) and isolated stationary point of $f(t,\cdot)$.

Proposition 2. *Under Assumption 3, any Filippov solution $x(\cdot)$ of (13), where F satisfies the condition*

$$\frac{\partial \|\nabla f(t,x)\|^2}{\partial t} + 2\nabla f(t,x)^T[\nabla^2 f(t,x)]F(t,x) \leq -c\|\nabla f(t,x)\|^{2\alpha}, \tag{25}$$

with $c > 0$, $\alpha \in (0,1)$, for all $t \geq 0$ and $x \in \mathcal{D}$, with $x(0) = x_0$ sufficiently close to $x^{opt}(t)$ for a given $t \geq 0$, will converge in finite-time to $x^{opt}(.)$ with a settling time $t^\star \leq \frac{\|\nabla f(0,x_0)\|^{2(1-\alpha)}}{c(1-\alpha)}$.

Proof 3. *The proof relies on using the results of Theorem 2 with the Lyapunov function*

$$V(t,e) = \|\nabla f(t, e + x^{opt}(t))\|^2, \tag{26}$$

where $e = x - x^{opt}(t)$ defines the tracking error. Indeed, if we compute the time derivative of (26) along the solutions of (13), we can write[2]

$$\begin{aligned}
\sup\dot{V}(t,e) &= sup\{\tfrac{\partial V(t,e)}{\partial t} + \tfrac{\partial V(t,e)}{\partial e}v, \,: v \in K(t,e)\} \\
&= \tfrac{\partial \|\nabla f(t,e+x^{opt}(t))\|^2}{\partial t} + 2\nabla f(t,e+x^{opt}(t))^T[\nabla^2 f(t,e+x^{opt}(t))]\dot{e} \\
&= \tfrac{\partial \|\nabla f(t,e+x^{opt}(t))\|^2}{\partial t} + 2\nabla f(t,e+x^{opt}(t))^T[\nabla^2 f(t,e+x^{opt}(t))](\dot{x} - \dot{x}^{opt}) \\
&= \tfrac{\partial \|\nabla f(t,x)\|^2}{\partial t} + 2\nabla f(t,e+x^{opt}(t))^T[\nabla^2 f(t,e+x^{opt}(t))](\dot{x}^{opt}) \\
&\qquad + 2\nabla f(t,e+x^{opt}(t))^T[\nabla^2 f(t,e+x^{opt}(t))](\dot{x} - \dot{x}^{opt}) \\
&= \tfrac{\partial \|\nabla f(t,x)\|^2}{\partial t} + 2\nabla f(t,x)^T[\nabla^2 f(t,x)]F(t,x),
\end{aligned} \tag{27}$$

which together with (22) leads to the condition (25), and the finite-time convergence result follows from the statement of Theorem 2.

[2] In the remaining of this paper, for simplicity, we will not repeat the exact set definition of $\dot{V}(t,x)$ as introduced in (21). Indeed, in this paper the sets $K(t,e)$ are all defined as singletons $K(t,e) = \{F(t, e + x^{opt}) - \dot{x}^{opt}\}$ due to the continuity of the flows F, a.,e., except at the optimal point, i.e., $e = 0$.

Remark 1. Condition (25) is in the form of a PDE inequality, and hence is difficult to solve numerically. However, we intend to use this condition to design a famlily of flows which satisfy it in closed-form, avoiding any need for numerical integration of the PDE inequality itself.

We are ready to propose some optimization flows which satisfy the general condition of Proposition 2. The first family of flows is in the form of Newton-like discontinuous flows with *pre-defined finite* settling time, for static cost function optimization. These flows are then extended to the case of time-varying cost function. Finally, a robustification of the flows is proposed, for the case of bounded additive uncertainties.

Assumption 4. $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, with $f(.)$ strongly convex (respectively, strongly concave) in a convex open set $\mathcal{D} \subset \mathbb{R}^n$, and $x^{opt} \in \mathcal{D}$ is a strict local minimizer (respectively, maximizer) and isolated stationary point of $f(\cdot)$.

Proposition 3 (Family of flows for pre-defined finite-time optimization)**.** *Under Assumption 4, any Filippov solution $x(\cdot)$ of*

$$\dot{x} = -\frac{\|\nabla f(x_0)\|^{2(1-\alpha)}}{2T(1-\alpha)} \frac{\|\nabla f(x)\|^{2\alpha}}{\nabla^T f(x)[\nabla^2 f(x)]^{r+1}\nabla f(x)}[\nabla^2 f(x)]^r \nabla f(x),\ \alpha \in [0.5, 1),\ r \in \mathbb{R} \tag{28}$$

with $x(0) = x_0$ sufficiently close to x^{opt} will converge in finite-time to x^{opt} with an exact settling time $t^\star = T$.

Proof 4. *Let $\mathcal{D}$ be an open neighborhood of x^{opt}, where x^{opt} is the only stationary point of f and $\nabla^2 f(x)$ is positive definite for every $x \in \mathcal{D}$ (respectively, negative defnite for every $x \in \mathcal{D}$). Let $V : \mathcal{D} \to \mathbb{R}$ be given by $V(x) = \|\nabla f(x)\|^2$. Clearly, V is continuously differentiable and positive definite w.r.t. x^{opt}. Furthermore, if $x \in \mathcal{D} \setminus \{x^{opt}\}$, then*

$$\begin{aligned} &\sup \dot{V}(x) = \nabla V(x) \cdot \left(-\tfrac{\|\nabla f(x_0)\|^{2(1-\alpha)}}{2T(1-\alpha)} \tfrac{\|\nabla f(x)\|^{2\alpha}}{\nabla^T f(x)[\nabla^2 f(x)]^{r+1}\nabla f(x)}[\nabla^2 f(x)]^r \nabla f(x) \right) \\ &= -\tfrac{\|\nabla f(x_0)\|^{2(1-\alpha)}}{T(1-\alpha)} \tfrac{\|\nabla f(x)\|^{2\alpha}}{\nabla^T f(x)[\nabla^2 f(x)]^{r+1}\nabla f(x)} \nabla^T f(x)[\nabla^2 f(x)]^{r+1}\nabla f(x) \\ &= -\tfrac{\|\nabla f(x_0)\|^{2(1-\alpha)}}{T(1-\alpha)} V(x)^\alpha. \end{aligned} \tag{29}$$

The result follows by invoking Theorem 1.

Remark 2. Note that, in Proposition 3 we selected the parameter $\alpha \in [0.5, 1) \subset (0, 1)$, the reason for this choice is to ensure that the flow remains bounded for all $x \in \mathbb{R}^n$, including at $x = x^{opt}$. To see the boundedness of the right hand side of the flow (28) with $c = -\frac{\|\nabla f(x_0)\|^{2(1-\alpha)}}{2T(1-\alpha)}$, we can write the following:

$$\begin{aligned} \|F(x)\| &= c\|\nabla f(x)\|^{2\alpha} \frac{\|[\nabla^2 f(x)]^r \nabla f(x)\|}{\nabla f(x)^\top [\nabla^2 f(x)]^{r+1}\nabla f(x)} \\ &\leq c\|\nabla f(x)\|^{2\alpha} \frac{\lambda_{\max}(\nabla^2 f(x))^r \|\nabla f(x)\|}{\lambda_{\min}(\nabla^2 f(x))^{r+1}\|\nabla f(x)\|^2} \\ &\leq c\frac{\lambda_{\max}(\nabla^2 f(x))^r}{\lambda_{\min}(\nabla^2 f(x))^{r+1}} \|\nabla f(x)\|^{2\alpha-1}, \end{aligned} \tag{30}$$

which is bounded for $2\alpha - 1 \geq 0$, the upper-bound on α, i.e., $\alpha < 1$ is still needed to ensure the finite-time convergence result.

In practical implementation, if we want to keep all the remaining range of α, i.e., $\alpha \in (0, 0.5)$, we could simply add a regularization term in the flow (28) , as $\dot{x} = -\frac{\|\nabla f(x_0)\|^{2(1-\alpha)}}{2T(1-\alpha)} \frac{\|\nabla f(x)\|^{2\alpha}}{\epsilon + \nabla^T f(x)[\nabla^2 f(x)]^{r+1} \nabla f(x)} [\nabla^2 f(x)]^r \nabla f(x)$, where $\epsilon \in \mathbb{R}$ is a very small non-zero scalar term, used to regularize the flow when $x \to x^*$.Theoretically, this implementation 'fix' will simply change the finite-time convergence, to a finite-time practical convergence, i.e., convergence to an ϵ-neighborhood of x^*.

Next, we extend this results to the case of time-varying cost function, i.e., a functional optimization problem.

We first introduce the following assumption.

Assumption 5. Let $f : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, in both variables, let $x^{opt} : \mathbb{R}_+ \to \mathcal{D}$ s.t., for each t, $x^{opt}(t)$ be a strict local optima and isolated stationary point of $f(t, \cdot)$, where $\mathcal{D} \subset \mathbb{R}^n$ is an open set s.t. $x^{opt}(t) \in \mathcal{D}$, $\forall t \geq 0$. Then, there exists a continuous function $l : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$, such that

$$\|\frac{\partial}{\partial t}[\nabla f(t, x)]\| \leq l(t, x), \forall t \geq 0, \forall x \in \mathcal{D}. \tag{31}$$

Proposition 4 (Discontinuous flow for finite-time time-varying optimization)**.** *Let $f : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, in both variables, let $x^{opt} : \mathbb{R}_+ \to \mathcal{D}$ s.t., for each t, $x^{opt}(t)$ be a strict local minimizer (respectively, maximizer) and isolated stationary point of $f(t, \cdot)$, where $\mathcal{D} \subset \mathbb{R}^n$ is an open set s.t. $x^{opt}(t) \in \mathcal{D}$, $\forall t \geq 0$. Consider the flow given by*

$$\dot{x} = -\frac{1}{2} \frac{[\nabla^2 f(t, x)]^r \nabla f(t, x)}{\nabla f(t, x)^T [\nabla^2 f(t, x)]^{r+1} \nabla f(t, x)} \big(2l(t, x)\|\nabla f(t, x)\| + c\|\nabla f(t, x)\|^{2\alpha}\big), \tag{32}$$

with $c > 0$, $\alpha \in [0.5, 1)$, $r \in \mathbb{R}$, and where $l : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ satisfies Assumption 5. Then, under Assumption 3, any Filippov solution $x(\cdot)$ of (32), with $x(0) = x_0$ sufficiently close to $x^{opt}(t)$ for a given $t \geq 0$, will converge in finite-time to $x^{opt}(.)$ with a settling time $t^\star \leq \frac{\|\nabla f(0, x_0)\|^{2(1-\alpha)}}{c(1-\alpha)}$.

Proof 5. *Let us define the tracking error as $e = x - x^{opt}(t)$, we then consider the Lyapunov function $V(t, e) = \|\nabla f(t, e + x^{opt}(t))\|^2$, and write its derivative as follows, for $e \in \{x - x^{opt} : x \in \mathcal{D}\} \setminus \{0\}$:*

$$\begin{aligned}
\sup \dot{V}(t, e) &= \tfrac{\partial}{\partial t}[\nabla f(t, e + x^{opt}(t))^T \nabla f(t, e + x^{opt}(t))] \\
&\quad + \tfrac{\partial}{\partial e}[\nabla f(t, e + x^{opt}(t))^T \nabla f(t, e + x^{opt}(t))]\dot{e}, \\
&= \tfrac{\partial}{\partial t}[\nabla f(t, e + x^{opt}(t))^T \nabla f(t, e + x^{opt}(t))] \\
&\quad + 2\nabla f(t, x)^T [\nabla^2 f(t, x)](\dot{x} - \dot{x}^*(t)), \\
&= \tfrac{\partial}{\partial t}[\nabla f(t, x)^T \nabla f(t, x)] + \tfrac{\partial}{\partial x}[\nabla f(t, x)^T \nabla f(t, x)]\dot{x}^*(t) \\
&\quad + 2\nabla f(t, x)^T [\nabla^2 f(t, x)](\dot{x} - \dot{x}^*(t)) \\
&= \tfrac{\partial}{\partial t}[\nabla f(t, x)^T \nabla f(t, x)] + 2\nabla f(t, x)^T [\nabla^2 f(t, x)]\dot{x},
\end{aligned} \tag{33}$$

next, by using (32), we can write

$$\begin{aligned}\sup \dot{V}(t,e) &= \tfrac{\partial}{\partial t}[\nabla f(t,x)^T \nabla f(t,x)] - 2l(t,x)\|\nabla f(t,x)\| - c\|\nabla f(t,x)\|^{2\alpha} \\ &\leq -c\|\nabla f(t, e + x^{opt}(t)\|^{2\alpha} = -cV(t,e)^{\alpha},\end{aligned} \tag{34}$$

which, by Theorem 2, leads to the desired finite-time convergence result.

Remark 3. Condition (31) might seem restrictive, however, it is an assumed upper-bound on the norm of the partial derivative of the cost's gradient function, which could be precisely computed if the closed-form of the cost is known. Alternatively, it can simply be selected as an arbitrarily large positive defined function, which enforces this condition, e.g., a large positive constant can be used if no a-priori knowledge of the cost function is available. This point will be demonstrated via numerical examples in Section 4.

Remark 4. It is clear from equation (34) that if instead of using an upper-bound $l(t,x)$ of the norm of $\frac{\partial}{\partial t}[\nabla f(t,x)]$ in the flow (32), one uses the exact term $-\frac{\partial}{\partial t}[\nabla f(t,x)^T \nabla f(t,x)]$, we can obtain an exact value of the finite-time convergence, i.e., $t^* = \frac{\|\nabla f(0,x_0)\|^{2(1-\alpha)}}{c(1-\alpha)}$. However, this will not be very practical, since it is difficult to be able to obtain the term $\frac{\partial}{\partial t}[\nabla f(t,x)^T \nabla f(t,x)]$ in closed-form is any meaningful application, and its numerical approximation will induce numerical errors, implying a lack of robustness of this solution, since it is based on an exact cancellation of this time-varying term.

3.1. Robustification of the flow w.r.t. additive time-varying uncertainties

First, we assume that residual computational errors can appear in the flow as bounded additive uncertainties. We then propose to modify the flow, based on robust nonlinear control, i.e., Lyapunov reconstruction technique, to reject the effect of these bounded uncertainties, and regain the nominal finite-time stability.

Indeed, consider the flow (32) with additive uncertainties, as follows:

$$\dot{x} = -\frac{1}{2}\frac{[\nabla^2 f(t,x)]^r \nabla f(t,x)}{\nabla f(t,x)^T [\nabla^2 f(t,x)]^{r+1} \nabla f(t,x)} \left(2l(t,x)\|\nabla f(t,x)\| + c\|\nabla f(t,x)\|^{2\alpha}\right) + \epsilon_1(t,x), \tag{35}$$

with $c > 0$, $\alpha \in [0.5, 1)$, $r \in \mathbb{R}$, and where the function $\epsilon_1 : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ represents bounded multiplicative uncertainties, which satisfies the following assumption.

Assumption 6. $\epsilon_1 : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$, satisfies $\|\epsilon_1(t,x)\| \leq \bar{\epsilon}_1,\ \forall t > 0,\ \forall x \in \mathbb{R}^n$.

To compensate for the effect of ϵ_1 in (35), we introduce an extra robustifying term $v_{rob}(x)$ as follows

$$\begin{aligned} x^{opt}\ \dot{x} = -\tfrac{1}{2}\tfrac{[\nabla^2 f(t,x)]^r \nabla f(t,x)}{\nabla f(t,x)^T [\nabla^2 f(t,x)]^{r+1} \nabla f(t,x)} &\left(2l(t,x)\|\nabla f(t,x)\| + c\|\nabla f(t,x)\|^{2\alpha} + v_{rob}(x)\right) \\ &+\epsilon_1(t,x),\ r \in \mathbb{R},\ c > 0,\ \alpha \in [0.5, 1)\end{aligned} \tag{36}$$

We then use Lyapunov reconstruction theory, e.g., Benosman & Lum (2010) to design a robustifying term v_{rob} which cancels the effect of the uncertainty ϵ_1 on the finite-time stability. This result is formalized in the following proposition.

Proposition 5 (Flow robustification w.r.t. bounded additive uncertainties)**.** *Let $f : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, in both variables, let $x^{opt} : \mathbb{R}_+ \to \mathcal{D}$ s.t., for each t, $x^{opt}(t)$ be a strict local minimizer (respectively, maximizer) and isolated stationary point of $f(t,\cdot)$, where $\mathcal{D} \subset \mathbb{R}^n$ is an open set s.t. $x^{opt}(t) \in \mathcal{D}$, $\forall t \geq 0$. Consider the flow given by (36), where l satisfies Assumption 5, ϵ_1 satisfies Assumption 6, and where v_{rob} is given by*

$$v_{rob}(x) = -\|\nabla f(t,x)^T[\nabla^2 f(t,x)]\|k,\ k \geq \bar{\epsilon_1}. \tag{37}$$

Then, under Assumption 3, any Filippov solution $x(\cdot)$ of (36), with $x(0) = x_0$ sufficiently close to $x^{opt}(t)$ for a given $t \geq 0$, will converge in finite-time to $x^{opt}(.)$ with a settling time $t^\star \leq \frac{\|\nabla f(0,x_0)\|^{2(1-\alpha)}}{c(1-\alpha)}$.

Proof 6. *Let us define the tracking error as $e = x - x^{opt}(t)$, we then consider the Lyapunov function $V(t,e) = \|\nabla f(t, e + x^{opt}(t))\|^2$, and write its derivative as follows, for $e \in \{x - x^{opt} :\ x \in \mathcal{D}\} \setminus \{0\}$:*

$$\begin{aligned} \sup \dot{V}(t,e) = & \tfrac{\partial}{\partial t}[\nabla f(t, e + x^{opt}(t))^T \nabla f(t, e + x^{opt}(t))] \\ & + \tfrac{\partial}{\partial e}[\nabla f(t, e + x^{opt}(t))^T \nabla f(t, e + x^{opt}(t))]\dot{e}, \\ = & \tfrac{\partial}{\partial t}[\nabla f(t, e + x^{opt}(t))^T \nabla f(t, e + x^{opt}(t))] \\ & + 2\nabla f(t,x)^T[\nabla^2 f(t,x)](\dot{x} - \dot{x}^*(t)), \\ = & \tfrac{\partial}{\partial t}[\nabla f(t,x)^T \nabla f(t,x)] + \tfrac{\partial}{\partial x}[\nabla f(t,x)^T \nabla f(t,x)]\dot{x}^*(t) \\ & + 2\nabla f(t,x)^T[\nabla^2 f(t,x)](\dot{x} - \dot{x}^*(t)) \\ = & \tfrac{\partial}{\partial t}[\nabla f(t,x)^T \nabla f(t,x)] + +2\nabla f(t,x)^T[\nabla^2 f(t,x)]\dot{x}, \end{aligned} \tag{38}$$

Next, by using (36) and (37), we write

$$\begin{aligned} & \sup \dot{V}(t,e) = \tfrac{\partial}{\partial t}[\nabla f^T \nabla f] - 2l(t,x)\|\nabla f(t,x)\| - c\|\nabla f(t,x)\|^{2\alpha} + v_{rob}(x) \\ & \qquad\qquad + \nabla f(t,x)^T[\nabla^2 f(t,x)]\epsilon_1(t,x), \\ & \leq -c\|\nabla f(t,x)\|^{2\alpha} + v_{rob}(x) + \|\nabla f(t,x)^T[\nabla^2 f(t,x)]\|\bar{\epsilon_1} \\ & \leq -c\|\nabla f(t,x)\|^{2\alpha} - \|\nabla f(t,x)^T[\nabla^2 f(t,x)]\|k + \|\nabla f(t,x)^T[\nabla^2 f(t,x)]\|\bar{\epsilon_1} \\ & \leq -c\|\nabla f(t,x)\|^{2\alpha} + \|\nabla f(t,x)^T[\nabla^2 f(t,x)]\|(-k + \epsilon_1) \\ & \leq -c\|\nabla f(t, e + x^{opt}(t))\|^{2\alpha} = -cV(t,e)^\alpha, \end{aligned} \tag{39}$$

which, by Theorem 2, leads to the desired finite-time convergence result.

3.2. Extension to Constrained Optimization

Consider a general constrained nonlinear optimization problem

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h_j(x) \geq 0, \quad j = 1, \ldots, p \\ & g_i(x) = 0, \quad i = 1, \ldots, e, \end{aligned} \tag{40}$$

with $f, h_1, \ldots, h_p, g_1, \ldots, g_e : \mathbb{R}^n \to \mathbb{R}$.

3.2.1. Static penalty barrier function formulation

First, a straightforward approach is to reformulate the constrained problem as an approximate unconstrained problem using a static[3] penalty formulation with barrier function. For instance, we can define the auxiliary penalized cost function

$$f_\mu(x) \stackrel{\text{def}}{=} f(x) - \mu \sum_{j=1}^{p} \log h_j(x) + \frac{1}{2\mu} \sum_{i=1}^{i=e} g_i^2(x) \tag{41}$$

where $\mu > 0$ denotes a penalty parameter. We then optimize $f_\mu(x)$ using algorithms of the form (28) for $f_\mu(x)$. This formulation is well known to converge to a neighborhood of a local minima for a small enough penalty parameter μ, e.g., Theorem 17.3, in Nocedal & Wright (1999).

We formalize this result in the following proposition.

Proposition 6. *Consider the optimization problem given by (40), where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and let $x^{opt} \in \mathbb{R}^n$ be a strict local solution of (40). Then, under Assumption 4, any Filippov solution $x(\cdot)$ of (28), where f_μ in (41) is substituted for f, with $x(0) = x_0$ sufficiently close to x^{opt} will converge[4] in finite-time to x^{opt} with a settling time $t^\star = T$, if $\mu \to 0$.*

Proof 7. *The proof follows from the arguments of Theorem 17.3, in Nocedal & Wright (1999), and the results of Proposition 3.*

3.2.2. Exact barrier function formulation

In the case of strict inequality constraints only, a more exact solution[5], follows ideas form barrier Lyapunov functions, e.g., Liu & Tong (2016); Yang et al. (2019), where we are interested in transforming, using an exact one-to-one mapping, the problem[6]

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h_j(x) > 0, \quad j = 1, \dots, p, \end{aligned} \tag{42}$$

into an equivalent unconstrained optimization one, of the form

$$\underset{\tilde{x} \in \mathbb{R}^n}{\text{minimize}} \quad \tilde{f}(\tilde{x}) \tag{43}$$

via a change of variables $x = \varphi(\tilde{x})$ with $\varphi : \mathbb{R}^m \to \Omega$ and $\tilde{f} = f \circ \varphi$, where $\circ$ denotes composition. Notice that, if f and T are twice continuously differentiable, then

$$\nabla \tilde{f}(\tilde{x}) = \nabla (f \circ \varphi)(\tilde{x}) = \mathbf{J}_\varphi(\tilde{x})^\top \nabla f(\varphi(\tilde{x})) \tag{44}$$

[3]By static we mean using a constant penalty coefficient.

[4]In the sense of pointwise convergence w.r.t. μ.

[5]In contrast with an approximate static penalty formulation.

[6]Notice that we are only considering here strict inequality constraints, due to the exact change of variables, which is not defined at the constraints' boundaries.

and

$$\nabla^2 \tilde{f} = \mathbf{J}_\varphi^\top (\nabla^2 f \circ \varphi) \mathbf{J}_\varphi + [\mathbb{I}_m \otimes (\nabla f \circ \varphi)^\top] \boldsymbol{\nabla}^2 \varphi, \tag{45}$$

where $\mathbf{J}_\varphi$ denotes the Jacobian matrix of T, $\mathbb{I}_m$ denotes the $m \times m$ identity matrix, $\otimes$ denotes the Kronecker product, and $\boldsymbol{\nabla}^2 \varphi \overset{\text{def}}{=} \left[\frac{\partial^2 \varphi}{\partial x'_i \partial x'_j} \right] \in \mathbb{R}^{mn \times m}$. Suppose now $\ker \mathbf{J}_\varphi(\tilde{x}) = \{0\}$, which clearly requires $m \leq n$. This is precisely the definition of a *immersion* in differential geometry[7].

Assumption 7. $\varphi : \mathbb{R}^m \to \Omega$ is a twice continuously differentiable immersion.

Remark 5. Every (possibly local) diffeomorphism[8] is an immersion.

Under Assumption 7, $x^{opt} \in \Omega$ is an isolated stationary point of f if and only if every $\tilde{x}^{opt} \in \varphi^{-1}(x^{opt})$ is an isolated stationary point of $\tilde{f}$, provided that φ is surjective. To show this, we first state the following lemma.

Lemma 1. *If $\Omega \subseteq \mathbb{R}^n$ and $\varphi : \mathbb{R}^m \to \Omega$ is an immersion, then φ is locally injective*[9].

Proof 8. *Refer to the Appendix Section.*

Proposition 7. *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable and $\varphi : \mathbb{R}^m \to \Omega$ ($\Omega \subseteq \mathbb{R}^n$) a continuously differentiable immersion. Then, $\tilde{x}^{opt} \in \mathbb{R}^m$ is an isolated stionary point of $\tilde{f}$ if and only if $\varphi(\tilde{x}^{opt})$ is an isolated stationary point of f.*

Proof 9. *We can choose $\varepsilon > 0$ small enough to ensure that $\tilde{x}^{opt}$ is the only stationary point of $\tilde{f}$ in $B_\varepsilon(\tilde{x}^{opt})$, and also that $\varphi|_{B_\varepsilon(\tilde{x}^{opt})} : B_\varepsilon(\tilde{x}^{opt}) \to \varphi(B_\varepsilon(\tilde{x}^{opt}))$ is injective. Since $\ker \mathbf{J}_\varphi(\tilde{x}) = \{0\}$ for every $\tilde{x} \in \mathbb{R}^m$ and $\nabla \tilde{f}(\tilde{x}) = \mathbf{J}_\varphi(\tilde{x})^\top \nabla f(\varphi(\tilde{x}))$, then $\tilde{x} \in \mathbb{R}^m$ is a stationary point of $\tilde{f}$ if and only if $\varphi(\tilde{x})$ is a stationary point of f. Therefore, $\varphi(\tilde{x}^{opt})$ is a stationary point of f. Furthermore, given any $x \in \varphi(B_\varepsilon(\tilde{x}^{opt})) \setminus \{\varphi(\tilde{x}^{opt})\}$, we'll have $\tilde{x} \neq \tilde{x}^{opt}$, where $\tilde{x} = \varphi|^{-1}_{B_\varepsilon(\tilde{x}^{opt})}(x)$, and thus x is not a stationary point of f. Therefore, $\varphi(\tilde{x}^{opt})$ is the only stationary point of f in the open set $\varphi(B_\varepsilon(\tilde{x}^{opt}))$, and thus it is an isolated stationary point.*

Notice that if x^{opt} is a strict local minimizer of f, then any $\tilde{x}^{opt} \in \varphi^{-1}(x^{opt})$ is also strict local minimizer of $\tilde{f}$, since the first term in (45) is positive definite at $\tilde{x}^{opt}$ and the second term vanishes, thus forcing $\nabla^2 \tilde{f}$ to be positive definite in an open neighborhood of $\tilde{x}^{opt}$.

The construction of a suitable barrier function transformation $\varphi(\cdot)$, in order to leverage our proposed Newton-like flows for unconstrained optimization problems, needs to be done on a case by case basis, since $\varphi(\cdot)$ will be fundamentally dictated by the constraints in (42).

---

[7]Let M and N be differentiable manifolds. Then, $\varphi : M \to N$ is an *immersion* if it is differentiable and its differential (the Jacobian matrix for Euclidean spaces) is everywhere injective.

[8]Let M and N be differentiable manifolds. Then, $\varphi : M \to N$ is a *local diffeomorphism* if for every $x \in M$, there exists an open neighborhood $V \subseteq M$ such that $\varphi|_V : V \to \varphi(V)$ is differentiable, invertible, and has differentiable inverse. In particular, $\varphi : M \to N$ is a *diffeomorphism* if it is differentiable, invertible, and has differentiable inverse.

[9]Let M be a topological space. Then, $\varphi : M \to N$ is *locally injective* at $x \in M$ if there exists an open neighborhood $V \subseteq M$ of x such that $\varphi|_V : V \to \varphi(V)$ is injective. Furthermore, $\varphi : M \to N$ is locally injective if it is locally injective at every $x \in M$.

Example: The case of box constraints

Consider the constrained optimization problem

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{minimize}} && f(x) \\ &\text{subject to} && \underline{x}_i < x_i < \overline{x}_i, \quad i = 1,\dots,n \end{aligned} \tag{46}$$

with $\underline{x}_i, \overline{x}_i \in \mathbb{R}$ such that $\underline{x}_i < \overline{x}_i$ for every $i = 1,\dots,n$. To construct a suitable transformation $\varphi(\cdot)$, we first recall the prototypical bijection between a bounded interval and the entire real line: the tangent barrier function and its inverse, the arctangent function. More precisely, recall that the tangent function defines a bijection $\tan : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$ with inverse $\tan^{-1} = \arctan$.

Using the arctangent function as a starting point, we can readily construct a bijection between $\mathbb{R}^n$ and $(\underline{x}, \overline{x})$. To do this, the proposed mapping will consist of applying the arctangent function component-wise on $\tilde{x} \in \mathbb{R}^m$ with $m = n$, and subsequently forcing that value, currently in $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)^n$, into $(\underline{x}, \overline{x})$, by applying the mapping $\psi_i(s) \overset{\text{def}}{=} \left(\frac{s}{\pi} + \frac{1}{2}\right)(\overline{x}_i - \underline{x}_i) + \underline{x}_i$ to its i-th component for $i = 1,\dots,n$. The affine maps $\psi_1(\cdot),\dots,\psi_n(\cdot)$ are clearly invertible, as they have non-zero slope, and their inverses are given by $\psi_i^{-1}(s) = \left(\frac{s-\underline{x}_i}{\overline{x}_i-\underline{x}_i}\right)\pi - \frac{\pi}{2}$ for $i = 1,\dots,n$. Putting everything together, we propose the barrier function $\varphi : \mathbb{R}^n \to (\underline{x}, \overline{x})$ given by $\varphi(\tilde{x}) = (\varphi_1(\tilde{x}_1),\dots,\varphi_n(\tilde{x}_n))$ with $\varphi_i \overset{\text{def}}{=} \psi_i \circ \arctan : \mathbb{R} \to (\underline{x}_i, \overline{x}_i)$ for $i = 1,\dots,n$. Clearly, each φ_i is a smooth diffeomorphism, and therefore so is φ. In other words, φ is smooth, invertible, and has a smooth inverse given by $\varphi^{-1}(x) = (\varphi_1^{-1}(x_1),\dots,\varphi_n^{-1}(x_n))$ with $\varphi_i^{-1} = \tan \circ \psi_i^{-1} : (\underline{x}_i, \overline{x}_n) \to \mathbb{R}$ for $i = 1,\dots,n$.

4. Numerical Examples

We will now test the proposed flows on some nonlinear cost functions.

We start by considering the Rosenbrock function $f : \mathbb{R}^2 \to \mathbb{R}$, which is given by

$$f(x_1, x_2) = (a - x_1)^2 + b(x_2 - x_1^2)^2, \tag{47}$$

with parameters $a, b \in \mathbb{R}$. This function nonlinear and non-convex, but smooth. It possesses exactly one stationary point $(x_1^{opt}, x_2^{opt}) = (a, a^2)$ for $b \geq 0$, which is a strict global minimum for $b > 0$. If $b < 0$, then (x_1^{opt}, x_2^{opt}) 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.

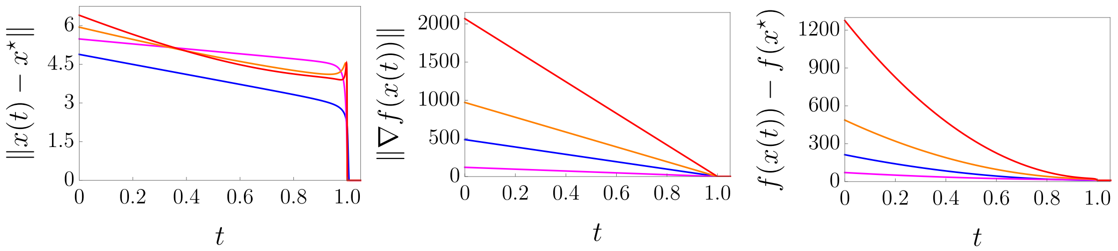
We apply the proposed flow (28), with the coefficients $\alpha = 0.5,\ T = 1,\ r = -1$.

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. Furthermore, we see that it does so in finite-time, with pre-defined settling time $T = 1$. It should be noted that at any given point in the trajectory $x(t)$, the functions $t \mapsto \|x(t) - x^{opt}\|$ and $t \mapsto |f(x(t)) - f(x^{opt})|$ are not guaranteed to decrease or remain constant, indeed only $t \mapsto \|\nabla f(x(t))\|$ can be guaranteed to do so, as we can see in Figure 1-(e).

Next, we consider the Rastringin function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = 10n + \sum_{i=1}^{n} [x_i^2 - 10\cos(2\pi x_i)], \tag{48}$$

(a) $t \in [0, 0.95]$ (b) $t \in [0, 0.999]$ (c) $t \in [0, 1]$



(d) Distance (e) Gradient (f) Cost

Figure 1.: Trajectories of the proposed flow (28) for the Rosenbrock function with parameters $(a, b) = (2, 50)$, i.e., a unique minimum $x^{opt} = (a, a^2) = (2, 4)$. Tests of four different initial conditions with the same settling time $T = 1$.

which is also nonlinear, non-convex, and smooth. It possesses a unique global minimum at the origin, and a countably infinite number of strict local minima, strict local maxima, and saddle points. In Figure 2 we can see the contour of the function in two dimensions ($n = 2$), and how its stationary points are distributed over the domain.

We implemented the same flow as int he first test. The corresponding results are reported in Figure 3, where we see clear convergence in $T = 1$ to the minimum at $(0, 0)$ when starting from 4 different initial conditions close to this minima. We also note that if we start close to another local extrema[10] (blue line in Figure (3, (a)-(b)-(c)), we converge to it in finite-time as well, this is due to the local nature of our convergence results.

Next, we report some examples for the constrained case. First, the exact barrier function formulation from Section 3.2.2, is tested on the Rosenbrock function under box constraints. The results are reported in Figure 4, where we see that the solution of the flow stays within the box constraints during the convergence time, i.e., for $t \in [0, 1]$.

Second, we test the barrier function penalty formulation from Section 3.2.1, on the Rosenbrock function, under linear inequality constraint, and where the optimum point has been chosen on the boundary of the constraint set. We test the flow convergence for several values of the penalty parameter μ. The results are reported in Figure 5, where the convergence in the desired finite-time $T = 1\,\text{sec}$ is achieved when μ is chosen

[10]This is a well know characteristic of second order methods, which can converge to local minima or maxima depending on the eigenvalues of the Hessian matrix. This problem has an 'implementation fix' as reported for example in (Nocedal & Wright (1999), Chapter 6), which could easily be implemented with our method.

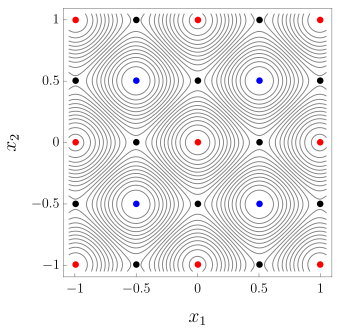
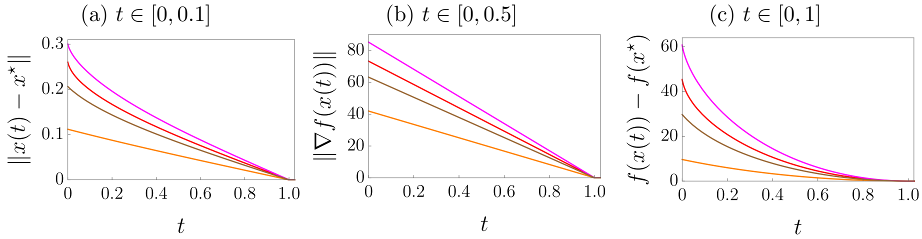


Figure 2.: The Rastringin function in two dimensions. The blue, red, and black points represent, respectively, strict local minima, strict local maxima, and saddle points. This pattern repeats throughout the entire domain $\mathbb{R}^2$, and together exhaust all stationary points.

(a) $t \in [0, 0.1]$ (b) $t \in [0, 0.5]$ (c) $t \in [0, 1]$



(d) Distance (e) Gradient (f) Cost

Figure 3.: Trajectory of proposed flow (28) for the the 2-dimensional Rastringin function with four different initial conditions and with the same settling time $T = 1$.

to be small enough, i.e. for $\mu = 0.01$.

(a) $t \in [0, 0.3]$ (b) $t \in [0, 0.9]$ (c) $t \in [0, 1]$

Figure 4.: Rosenbrock function with $(a, b) = (2, 100)$, $x_0 = (0.75, 3.8)$ and box constraint $x \in [\underline{x}, \overline{x}]$ with $\underline{x} = (1, 3.5)$ and $\overline{x} = (3, 4.5)$.

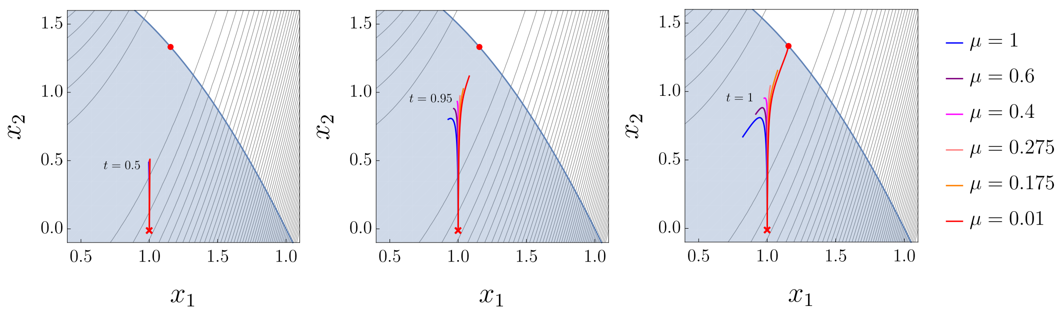


Figure 5.: Constrained penalization $\mu = 1$ (blue), $\mu = 0.65$ (purple), $\mu = 0.4$ (orange), $\mu = 0.01$ (red) over Rosenbrock function with $(a, b) = (2, 100)$, $x_0 = (1.25, 0.25)$ and linear constraint $2x_1 + x_2 \leq 4$.

Before moving to the time-varying cases, we want to report a comparative test between one of the finite-time convergent flows proposed here and one of the state of the art flows proposed in Karafyllis & Krstic (2017), with exponential convergence guarantees. Indeed, we consider the following example[11]

$$\begin{array}{l} \text{Minimize } f(t, x_1, x_2) = x_1^2 + ax_2^2, \; a > 0 \\ \text{s.t. } h(x) = x_1 - b = 0, \; b > 0. \end{array} \tag{49}$$

To solve this constrained problem, we first use the barrier function penalty formulation from Section 3.2.1, with the flow (28), and the coefficients $\alpha = 0.5$, $T = 0.005$, $r = -1$, $\mu = 1e - 6$. Then, we compare the results with the fast exponentially convergent flow, proposed in (Karafyllis & Krstic (2017), equation (7.9)), defined by

$$\dot{x} = -\sigma(x) \begin{bmatrix} \psi(x_1)(x_1 - b) \\ 2ax_2(1 + (x_1 - b)^2) \end{bmatrix}, \tag{50}$$

where, the choice $\sigma(x) \equiv 1$, and $\psi(x_1) = c > 0$ is shown (Karafyllis & Krstic (2017), P. 1323) to lead to exponential convergence of the solution of (50) to the minimum $(b,\ 0)'$.

The numerical solutions are reported in Figure 6, where we labelled the solution of the finite-time convergent flow by FF, and the solution of the exponentially convergent flow by EF. One can see that although the exponential convergent flow can have a very fast convergence rate, which we tuned by increasing σ to 20 and ψ to 100, it is still relatively 'slower' than the finite-time convergent flow, for which we can select the exact convergence time T to be even smaller than the convergence time obtained by a well-tuned exponentially convergent flow. The zoom on the flows' solutions reported in Figure 6- bottom plots, show that the finite-time flow converges exactly at time T (see Proposition 3), whereas the exponentially convergent flow reaches rapidly, but smoothly, i.e., not in finite-time, the minimum point. This difference is even clearer if we examine the phase plot reported in Figure 7-top, where we can observe that the exponentially convergent flow reaches, following a smooth trajectory, the constraint line, and then slides on it towards the optimal point, whereas the finite-time flow follows a more direct trajectory towards the optimal point. Finally, the gradient of the penalized cost f_μ in (41) is shown in 7-bottom, to display the convergence of the

[11] Example 7.2 in Karafyllis & Krstic (2017).

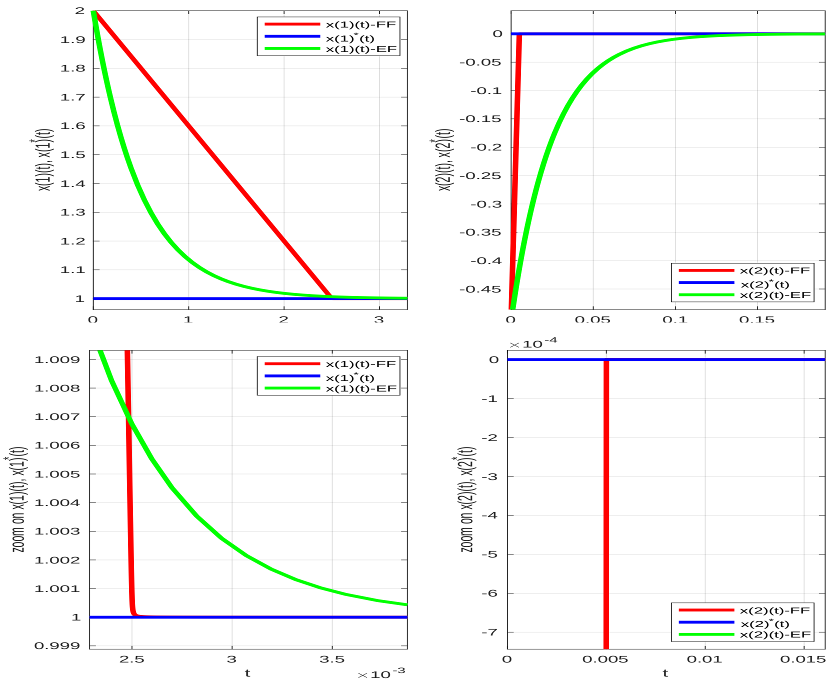


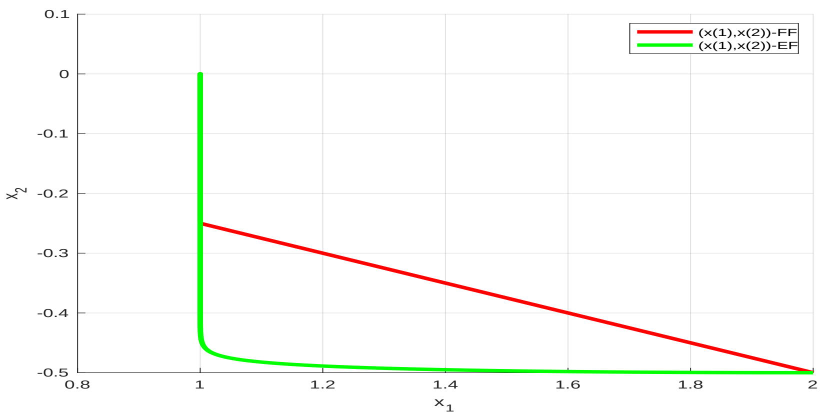
Figure 6.: A numerical comparison between a flow convergent in finite-time and an exponentially convergent flow: flows' solutions (FF- finite-time flow, EF-exponential flow)

gradient of the cost to zero at the predefined time $T = 5.10^{-3}$ for the finite-time convergent flow.

Finally, we report some numerical results for the time-varying case. We consider the time-varying Rosenbrock cost function

$$\begin{aligned} &f(t, x_1, x_2) = (a(t) - x_1)^2 + b(t)(x_2 - x_1^2)^2, \\ &a(t) = 2 + sin(t), \\ &b(t) = 50(1 + sin(t)), \end{aligned} \tag{51}$$

We apply the flow given by (32), with different upper-bound functions l. We choose the constants to be $r = -1,\ \alpha = 2/3,\ c = 100$, and the initial condition $x(0) = (4,\ 15)'$. We report below the numerical results for each upper-bound function l. In the first case, we assume exact knowledge of the cost function and compute the upper-bound in (31) in closed-form as $l(x) = \sqrt{(2 + 200(x_1 x_2 - x_1^3))^2 + 200(x_2 + x_1^2)^2}$. The corresponding results are reported in Figure 8. It is clear that the flow manages to minimize the time-varying cost function in finite-time within the expected time convergence limit $t^* = 2.6$ $[sec]$. This can be seen in Figure 8- bottom, where we display the norm of the gradient, which is decreasing monotonically and reaching zero in a finite-time less



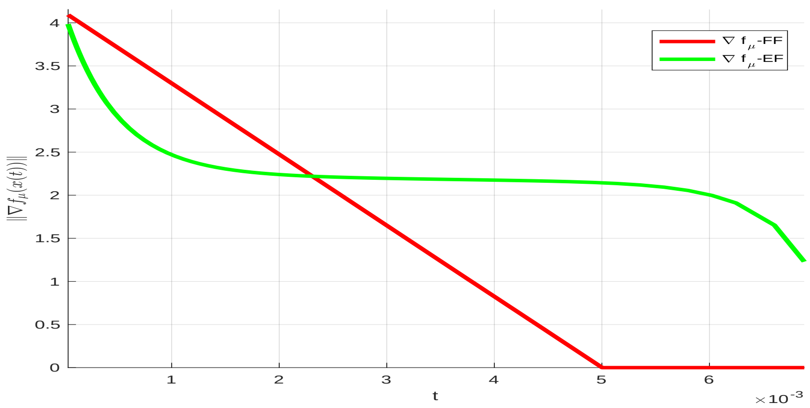


Figure 7.: A numerical comparison between a flow convergent in finite-time and an exponentially convergent flow: phase plots, and gradient plot (FF- finite-time flow, EF-exponential flow)

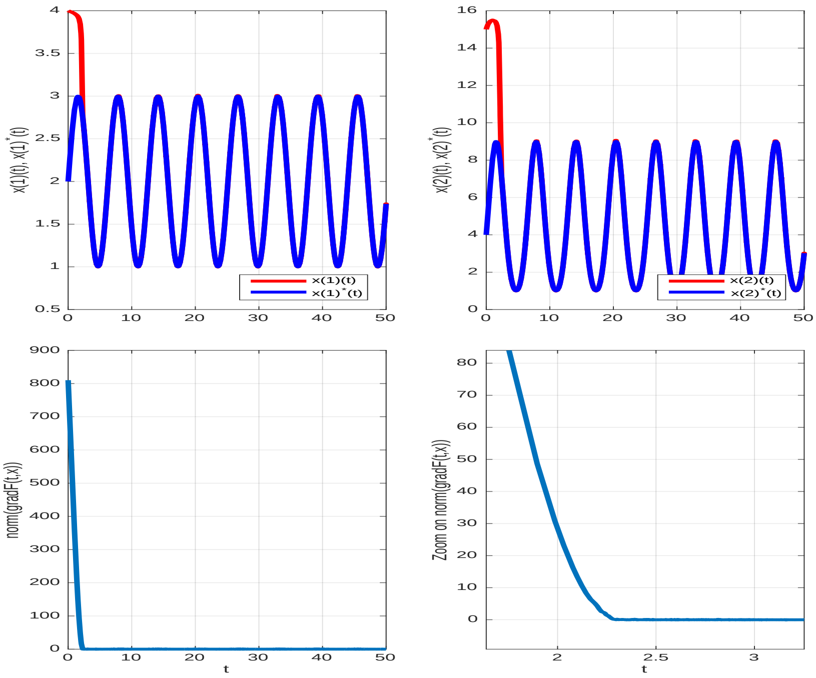


Figure 8.: Time-varying cost function: Flow (32) with $l(x) = \sqrt{(2 + 200(x_1x_2 - x_1^3))^2 + 200(x_2 + x_1^2)^2}$.

than t^*. To appreciate this result, we also report in Figure 9 the results when we force l to zero. One can see that without the upper-bound term l the norm of the gradient is not decreasing at all time, and although the flow solutions finally reach the true optimal solutions, they do so after the expected finite-time convergence limit t^*, which underlines the necessity of this term in ensuring convergence in desired finite-time, as seen in the analysis of the results of Proposition 4.

In the previous result we used an exact upper-bound l derived from closed-from manipulation of the cost function. However, this is not always possible in real applications, and thus we show next that any loose upper-bound suffice to ensure the finite-time stability result. Indeed, we first report in Figure 10 the results obtained with the positive definite function $l(t, x) = 500 + t^2$. Then we report in Figure 11 the results corresponding to the case $l(t, x) = 500$, which is the simplest upper-bound one can choose. In both cases the flow achieves the expected finite-time convergence. This shows that the proposed flow (32) is not very sensitive to the choice of the function l, as long as it is a valid upper-bound, as defined in (31).

We now report some results related to the performance of the flow under additive bounded uncertainties, as discussed in Section 3.1.

Let us first show how the previous results in the time-varying case deteriorates when we add a small uncertainty on the flow (32). Indeed, we apply the flow (32),

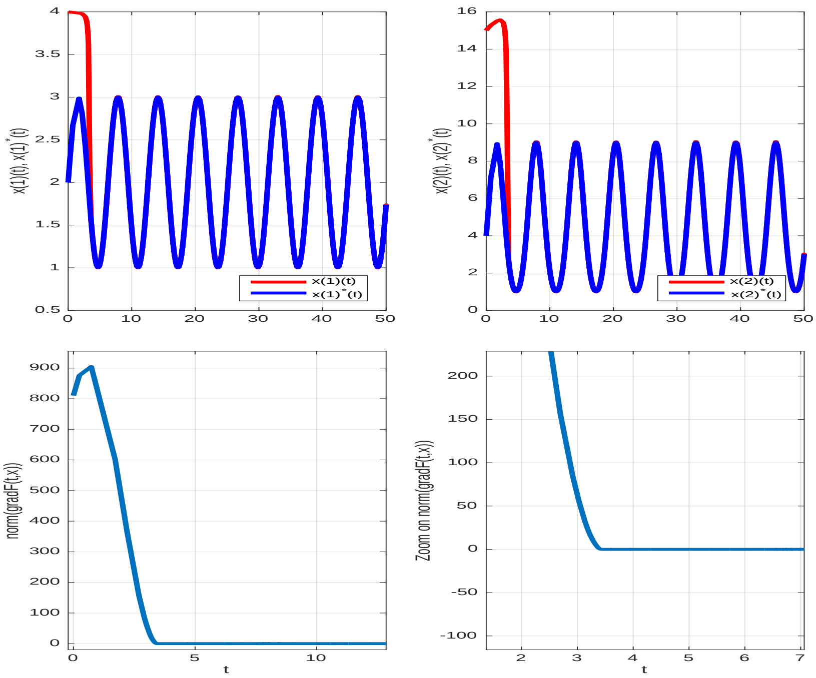


Figure 9.: Time-varying cost function: Flow (32) with $l(x) = 0$.

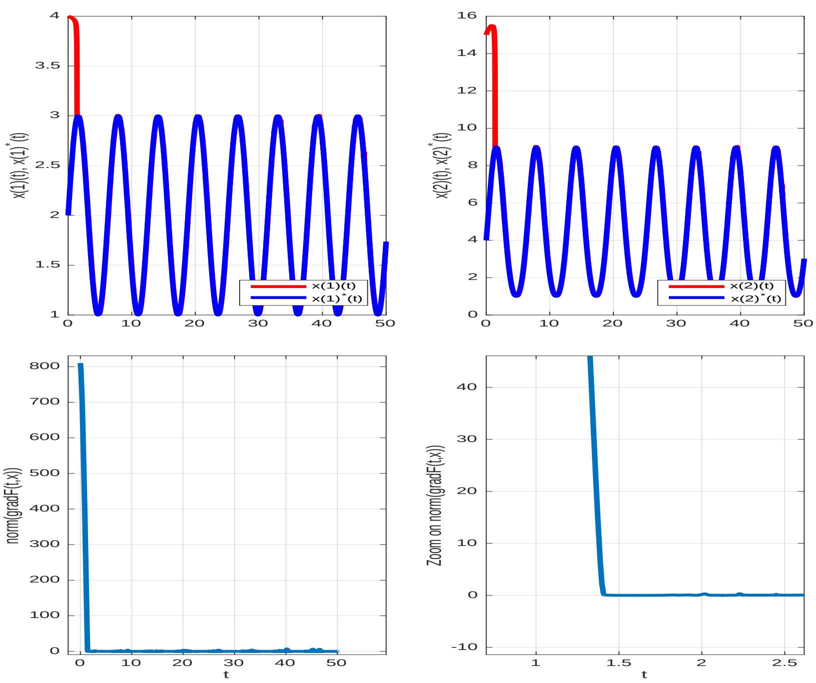


Figure 10.: Time-varying cost function: Flow (32) with $l(t, x) = 500 + t^2$.

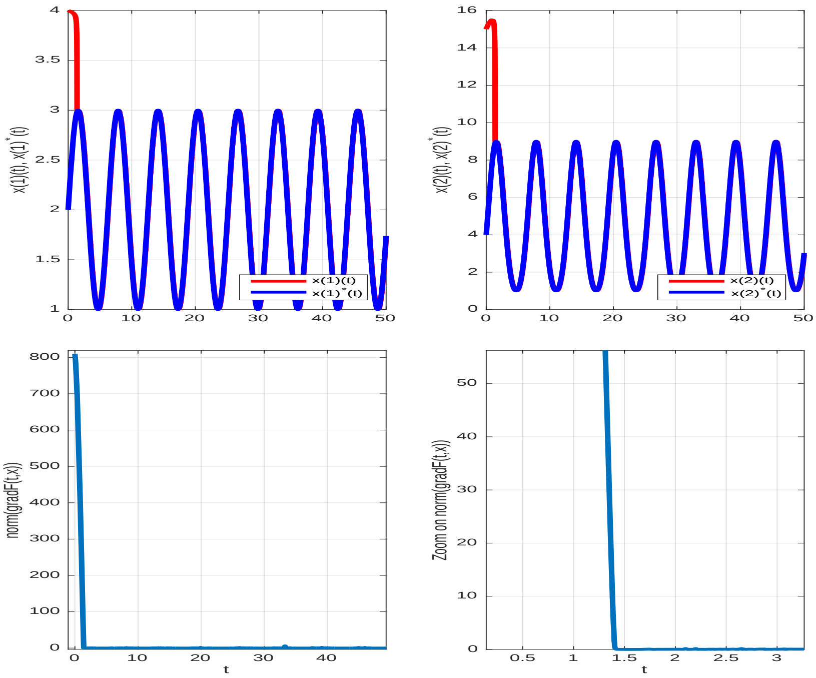


Figure 11.: Time-varying cost function: Flow (32) with $l(t, x) = 500$.

with $l(t, x) = 500$ with the same constants as the previous tests, but add the additive uncertainty $\epsilon_1(t) = -10^{-6}(1 + sin(t))$, as described in (35). The results are reported in Figure 12, where we clearly see that this uncertainty makes the flow unstable. We now use the robustified flow (36) and (37), with $k = 10^2$. The corresponding results are shown in Figure 13. One can see that the robustification term compensates for this bounded uncertainty, as expected from the results of Proposition 5.

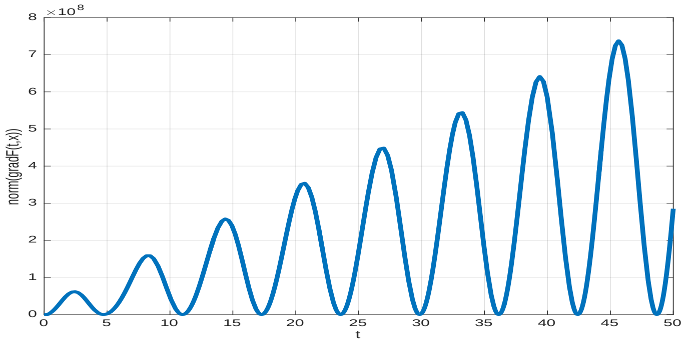


Figure 12.: Time-varying cost function: Nominal flow (32) with additive uncertainty $\epsilon_1(t) = -10^{-6}(1 + sin(t))$.

5. Conclusions and Future Work

We have introduced a new family of second-order flows for continuous-time optimization of time-varying cost functions. The main characteristic of the proposed flows is their finite-time convergence guarantees. Furthermore, some of the flows are designed in such a way that the finite convergence time can be pre-defined by the user. To be able to analyze these discontinuous flows, we had to first extend an exiting Lyapunov-based inequality condition for finite-time stability in the case of smooth dynamics to the case of non-smooth dynamics modeled by (time-varying) differential inclusions. We have proposed a robustification of the flows w.r.t. bounded additive uncertainties, and extended some of the results to the constrained case. These flows were tested on two well known optimization benchmarks.

Although the obtained results are encouraging, we underline that we have used available numerical solvers to validate the optimization performance of these flows. For these results to be more useful to the optimization and machine learning communities, it is necessary to find appropriate discretization schemes, which lead to actual optimization algorithms, with similar finite-time (or accelerated) convergence rates. This is the subject of our ongoing investigations. Other points which we will investigate in the future are the extension of these results to a fully extremum seeking version, where the cost function derivatives are estimated from direct measurements of the cost, and the application of these flows (or their best discretization) to actual large scale optimization problems, e.g., deep learning optimization problems.

Ackowledgements

The research for this work was largely written and conducted while at a summer research internship at Mitsubishi Electrics Research laboratories (MERL), in Cambridge,

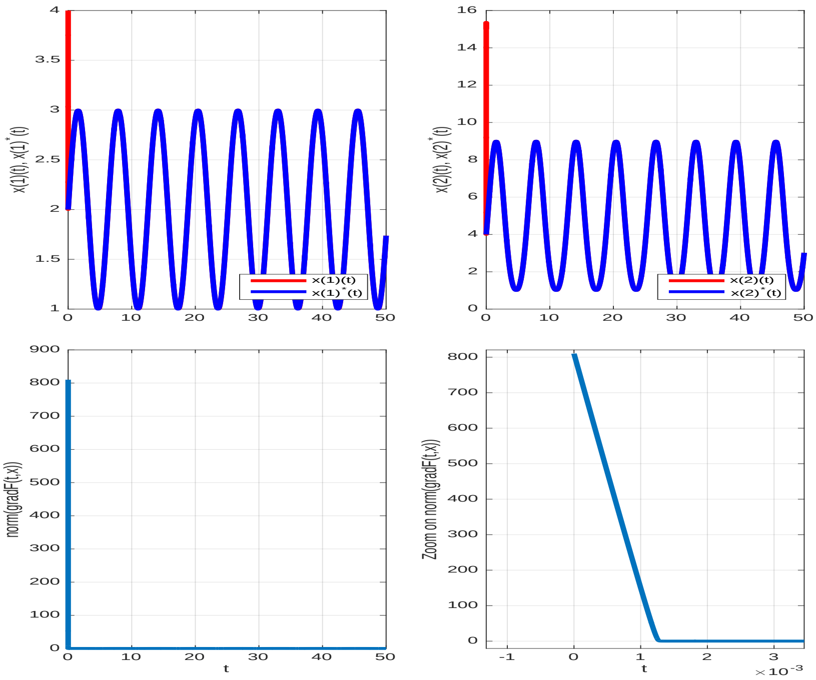


Figure 13.: Time-varying cost function: Robust flow (36),(37) with $k = 10^2$, under additive uncertainty $\epsilon_1(t) = -10^{-6}(1 + sin(t))$.

MA, from May to August of 2019.

Appendix

Proof of Lemma 1: By definition of Jacobian matrices as total derivatives, we have

$$\|\varphi(\tilde{x}') - \varphi(\tilde{x}) - \mathbf{J}_\varphi(\tilde{x})(\tilde{x}' - \tilde{x})\| = o(\|\tilde{x}' - \tilde{x}\|), \tag{52}$$

for $\tilde{x}, \tilde{x}' \in \mathbb{R}^m$. Furthermore, the $n \times m$ Jacobian matrix $\mathbf{J}_\varphi(\tilde{x})$ is injective, *i.e.* $\text{rank}(\mathbf{J}_\varphi(\tilde{x})) = m$. Therefore, we have $\text{rank}(\mathbf{J}_\varphi(\tilde{x})^\top \mathbf{J}_\varphi(\tilde{x})) = \text{rank}(\mathbf{J}_\varphi(\tilde{x})) = m$, and thus $\mathbf{J}_\varphi(\tilde{x})^\top \mathbf{J}_\varphi(\tilde{x})$ is invertible. Therefore, $\mathbf{J}_\varphi(\tilde{x})$ has a left inverse, given by the Moore-Penrose pseudoinverse $\mathbf{J}_\varphi(\tilde{x})^\dagger = (\mathbf{J}_\varphi(\tilde{x})^\top \mathbf{J}_\varphi(\tilde{x}))^{-1} \mathbf{J}_\varphi(\tilde{x})^\top$. Therefore, its induced Euclidean norm satisfies

$$\|\mathbf{J}_\varphi(\tilde{x})^\dagger\| \geq \frac{\|\tilde{x}' - \tilde{x}\|}{\|\mathbf{J}_\varphi(\tilde{x})(\tilde{x}' - \tilde{x})\|} \tag{53}$$

for every $\tilde{x}' \neq \tilde{x}$, since $\mathbf{J}_\varphi(\tilde{x})^\dagger \mathbf{J}_\varphi(\tilde{x})(\tilde{x}' - \tilde{x}) = \tilde{x}' - \tilde{x}$, and $\mathbf{J}_\varphi(\tilde{x})(\tilde{x}' - \tilde{x}) \neq 0$ due to the injectivity of $\mathbf{J}_\varphi(\tilde{x})$. Putting everything together, we have

$$\|\varphi(\tilde{x}') - \varphi(\tilde{x})\| \geq \|\mathbf{J}_\varphi(\tilde{x})(\tilde{x}' - \tilde{x})\| - \|\varphi(\tilde{x}') - \varphi(\tilde{x}) - \mathbf{J}_\varphi(\tilde{x})(\tilde{x}' - \tilde{x})\| \tag{54a}$$

$$\overset{(52)}{=} \|\mathbf{J}_\varphi(\tilde{x})(\tilde{x}' - \tilde{x})\| + o(\|\tilde{x}' - \tilde{x}\|) \tag{54b}$$

$$\overset{(53)}{\geq} \frac{\|\tilde{x}' - \tilde{x}\|}{\|\mathbf{J}_\varphi(\tilde{x})^\dagger\|} + o(\|\tilde{x}' - \tilde{x}\|) \tag{54c}$$

$$= \underbrace{\left(\frac{1}{\|\mathbf{J}_\varphi(\tilde{x})^\dagger\|} + \frac{o(\|\tilde{x}' - \tilde{x}\|)}{\|\tilde{x}' - \tilde{x}\|}\right)}_{(\star)} \|\tilde{x}' - \tilde{x}\|. \tag{54d}$$

Finally, for each $\tilde{x} \in \mathbb{R}^m$, if $\|\tilde{x}' - \tilde{x}\| > 0$ is sufficiently small, then $(\star) > 0$. Therefore, $\|\varphi(\tilde{x}') - \varphi(\tilde{x})\| > 0$. ■

Statements and proofs of Lemma 2 and 3:

Recall that a function $x : I \to \mathbb{R}^n$ defined over an interval $I \subset \mathbb{R}$ is *absolutely continuous* if, for every $\varepsilon > 0$, there exists some $\delta > 0$ such that

$$\sum_{j=1}^{k} (t'_j - t_j) < \delta \implies \sum_{j=1}^{k} \|x(t'_j) - x(t_j)\| < \varepsilon \tag{55}$$

for any disjoint subintervals $[t_1, t'_1], \ldots, [t_k, t'_k] \subseteq I$.

Lemma 2. *If $x : I \to \mathbb{R}^n$ is absolutely continuous, then so is $t \mapsto (t, x(t))$.*

Proof 10. *We start by fixing an arbitrarily small $\varepsilon > 0$. Since $x(\cdot)$ is absolutely continuous, we can choose some $\delta > 0$ such that (55) holds. Furthermore, we can clearly always make δ smaller, and thus assume $0 < \delta \leq \varepsilon$. Let $\varepsilon' = \varepsilon - \delta$. Once again invoking the absolute continuity of $x(\cdot)$, we can choose some $\delta' > 0$ such that (55) holds for δ' and ε' instead of δ and ε. Furthermore, we can choose δ' in the interval $(0, \delta]$. Therefore, we have, for any disjoint subintervals $[t_1, t'_1], \ldots, [t_k, t'_k] \subset I$ such that $\sum_{j=1}^{k} (t'_j - t_j) < \delta$,*

$$\sum_{j=1}^{k} \|(t'_j, x(t'_j)) - (t_j, x(t_j))\| \tag{56a}$$

$$= \sum_{j=1}^{k} \|(t'_j - t_j, x(t'_j) - x(t_j))\| \tag{56b}$$

$$\leq \sum_{j=1}^{k} [(t'_j - t_j) + \|x(t'_j) - x(t_j)\|] \tag{56c}$$

$$< \delta' + \varepsilon' \tag{56d}$$

$$\leq \varepsilon. \tag{56e}$$

Therefore, $t \mapsto (t, x(t))$ is absolutely continuous in I.

As a direct corollary, we have the following result.

Lemma 3. *If $x : [0, \tau] \to \mathbb{R}^n$ is absolutely continuous and $V : [0, \tau] \times \mathcal{D} \to \mathbb{R}^n$ is Lipschitz continuous (in both variables), where $\mathcal{D} \subset \mathbb{R}^n$ is an open set that contains the trajectory $x(\cdot)$, then $t \mapsto V(t, x(t))$ is absolutely continuous.*

Proof 11. *By Lemma 2, we know that $t \mapsto (t, x(t))$ is absolutely continuous in $[0, \tau]$. Therefore, given that $V(\cdot)$ is Lipschitz continuous, it follows that its composition with $t \mapsto (t, x(t))$ is absolutely continuous in $[0, \tau]$.*

References

K. B. Ariyur and M. Krstić, *Real-Time Optimization by Extremum-Seeking Control.* Wiley, 2003.

H. Attouch and J. Peypouquet, "The rate of convergence of nesterov's accelerated forward-backward method is actually $o(k^{-2})$," *SIAM Journal on Optimization*, no. 3, pp. 1–7, 2015.

H. Attouch, J. Peypouquet, and P. Redont, "Fast convex optimization via inertial dynamics with hessian driven damping," *Journal of Differential Equations*, no. 10, pp. 5734–5783, 2016.

——, "Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity," *Mathematical Programming*, no. 168, pp. 123–175, 2018.

R. Brockett, "Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems," in *IEEE Conference on Decision and Control*, 1988, pp. 799–803.

A. Bacciotti and F. Ceragioli, "Stability and stabilization of discontinuous systems and nonsmooth lyapunov functions," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 4, pp. 361–376, 1999.

M. Benosman and K.-Y. Lum, "Nonlinear control allocation for non-minimum phase systems," *IEEE Transactions on Control Systems Technology*, vol. 17, no. 2, pp. 394–404, 2009.

M. Benosman and K.-Y. Lum, "Passive actuators' fault tolerant control for affine nonlinear systems," *IEEE, Transactions on Control Systems Technology*, vol. 18, pp. 152–163, 2010.

F. Clarke, "Nonsmooth analysis in control theory: A survey," *European Journal of Control*, vol. 7, no. 2, pp. 145 – 159, 2001.

J. Cortés, "Finite-time convergent gradient flows with applications to network consensus," *Automatica*, vol. 42, pp. 1993–2000, 03 2006.

J. Cortes, "Discontinuous dynamical systems," *IEEE Control Systems Magazine*, vol. 28, no. 3, pp. 36–73, June 2008.

L. Faybusovich, "Dynamical systems which solve optimization problems with linear constraints," *IMA Journal of Mathematical Control & Information*, pp. 135–149, 1991.

A. F. Filippov and F. M. Arscott, "Differential equations with discontinuous righthand sides," in *Mathematics and its Applications*, Springer, 1988.

G. Franca, D. Robinson, and R. Vidal, "A dynamical systems perspective on nonsmooth constrained optimization," *arXiv:1808.04048v2 [math.OC]*, 2019(a).

G. Franca, J. Sulam, D. Robinson, and R. Vidal, "Conformal symplectic and relativistic optimization," *arXiv:1903.04100v3 [math.OC]*, 2019(b).

——, "ADMM and accelerated ADMM as continuous dynamical systems," *Proceedings of the 35th International Conference on Machine Learning, Stockholm*, 2018.

M. Guay and T. Zhang, "Adaptive extremum seeking control of nonlinear dynamic systems with parametric uncertainties," *Automatica*, vol. 39, pp. 1283–1293, 2003.

V. Grushkovskaya, A. Zuyev, and C. Ebenbauer, "On a class of generating vector fields for the extremum seeking problem: Lie bracket approximation and stability properties," *Automatica*, no. 94, pp. 151–160, 2018.

L. Grüne, and I. Karafyllis, "Lyapunov function based step size control for numerical ODE solvers with application to optimization algorithms ," *Mathematical System Theory, K. Huper and J. Trumpf, eds., CreateSpace*, pp. 183–210, 2013.

U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, 1996.

I. Karafyllis, and M. krstic, "Global dynamical solvers for nonlinear programming problems," *SIAM J. Control Optim.*, Vol. 55, No. 2, pp. 1302–1331, 2017.

I. Karafyllis, "Feedback stabilization methods for the solution of nonlinear programming problems ," *J. Optim. Theory Appl.*, No. 161, pp. 783–806, 2014.

M. Krstić, "Stability of extremum seeking feedback for general nonlinear dynamic agents," *Automatica*, vol. 36, pp. 595–601, 2000.

Y.-J. Liu and S. Tong, "Barrier Lyapunov functions-based adaptive control for a class of nonlinear pure-feedback systems with full state constraints," *Automatica*, no. 64, pp. 70–75, 2016.

J. Nocedal and S. J. Wright, *Numerical Optimization.* Springer, 1999.

B. Paden and S. Sastry, "A calculus for computing Filippov as differential inclusion with application to the variable structure control of robot manipulators," *Circuits and Systems, IEEE Transactions on*, vol. 34, pp. 73 – 82, 02 1987.

E. Polak, *Optimization Algorithms and Consistent Approximations.* Springer, 1997.

J. I. Poveda and N. Li, "Inducing Uniform Asymptotic Stability in Non-Autonomous Accelerated Optimization Dynamics via Hybrid Regularization', *Proceedings of the 58th IEEE Conference on Decision and Control* (to appear), 2019.

J. Poveda and A. Teel, "A framework for a class of hybrid extremum seeking controllers with dynamic inclusions," *Automatica*, no. 76, pp. 113–126, 2017.

D. Scieur, V. Roulet, F. Bach, and A. d'Aspremont, "Integration methods and accelerated optimization algorithms," *31st Conference on Neural Information Processing Systems*, 2017.

A. Scheinker and M. Krstić, *Model-free stabilization by extremum seeking.* Springer, 2016.

W. Su, S. Boyd, and E. J. Candes, "A differential equation for modeling Nesterovs accelerated gradient method: Theory and insights," *Journal of Machine Learning Research*, no. 17, pp. 1–43, 2016.

J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *IEEE Conference on Decision and Control and European Control Conference*, 2011, pp. 3800–3805.

C. Wang and Y. M. Lu, "The scaling limit of high-dimensional online independent component analysis," in *31st Conference on Neural Information Processing Systems*, 2017.

A. C. Wilson, B. Recht, and M. I. Jordan, "A lyapunov analysis of momentum methods in optimization," *arXiv:1611.02635*, 2016.

J. Wang and N. Elia, "Control approach to distributed optimization," in *48th Annual Allerton Conference*, 2010, pp. 557–561.

Y. Yang, K. G. Vamvoudakis, H. Modares, W. He, Y. Yin, and D. C. Wunsch, "Safety-aware reinforcement learning framework with an actor-critic-barrier structure," in *IEEE American Control Conference*, 2019.

C. Zhang and R. Ordónez, *Extremum-Seeking Control and Applications.* Springer, 2012.

J. Zhang, A. Mokhtari, S. Sra, and A. Jadbabaie, "Direct runge-kutta discretization achieves acceleration," in *32st Conference on Neural Information Processing Systems*, 2018.