A discretization algorithm for time-varying composite gradient flow dynamics

Kanat Camlibel, Aneel Tanwani

A discretization algorithm for time-varying composite gradient flow dynamics

Kanat Camlibel * **Aneel Tanwani** **

* *Bernoulli Institute of Mathematics, Computer Science, and Artificial Intelligence, University of Groningen, The Netherlands.*
Email: `m.k.camlibel@rug.nl`
** *LAAS – CNRS, University of Toulouse,*
7 Avenue Colonel du Roche, Toulouse 31400 France.
Email: `aneel.tanwani@laas.fr`

Abstract: The problem of minimizing the sum, or composition, of two objective functions is a frequent sight in the field of optimization. In this article, we are interested in studying relations between the discrete-time gradient descent algorithms used for optimization of such functions and their corresponding gradient flow dynamics, when one of the functions is in particular time-dependent. It is seen that the subgradient of the underlying convex function results in differential inclusions with time-varying maximal monotone operator. We describe an algorithm for discretization of such systems which is suitable for numerical implementation. Using appropriate tools from convex and functional analysis, we study the convergence with respect to the size of the sampling interval. As an application, we study how the discretization algorithm relates to gradient descent algorithms used for constrained optimization.

Keywords: Convex functions; Gradient flow dynamics; Maximal monotone operators; Time-stepping algorithm

1. INTRODUCTION

One of the recent trends in research at the intersection of optimization and dynamical systems has been to develop connections between optimization algorithms and dynamical systems to provide qualitative or quantitative assessment of the performance of certain optimization algorithms. Conventionally, unconstrained optimization problems over a continuous domain have been solved by gradient descent algorithms, and one can directly draw connections with Euler discretization of the corresponding gradient flow dynamics (Su et al., 2014). When studying constrained optimization problems, we can apply the gradient descent algorithm to the composite function described by the sum of the objective function and the indicator function associated to the constraint set. The performance of the resulting algorithm, more commonly known as the proximal algorithm, has also been reported in several studies (Nesterov, 2013; Parikh and Boyd, 2014; Attouch and Peypouquet, 2019).

We are interested in studying connections between proximal algorithms for sum-type composite functions and the corresponding differential inclusions obtained by taking the gradient (or subdifferential) of the sum of two convex functions. In particular, we consider the case when one of the functions could be time-dependent and extended real-valued (to describe constraints in an optimization problem). Our goal is to study a discretization algorithm for such continuous-time dynamical systems and compare the continuous-time solution with the discrete approximations. Our analysis makes use of the fact that the right-hand side of such differential inclusions involves maximal monotone operators, so that the central object of our study is abstractly written as

$$\dot{x} \in -F(t,x), \quad x(0) \in \operatorname{dom} F(0,\cdot), \tag{1}$$

where $F : [0,\infty) \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ has the property that, for each $t \geqslant 0$, $F(t,\cdot)$ is a maximal monotone operator.

A comprehensive reference for time-invariant differential inclusions with maximal monotone operators can be traced back to (Brézis, 1973), or see a recent reference (Camlibel and Schumacher, 2016) for how such inclusions are obtained by interconnections. Generalizing such methods to time-varying maximal monotone operators has been rather challenging. When $F(t,x)$ is the subdifferential of a time-dependent, proper, lower semicontinuous, and convex function $g_t(\cdot)$, that is, $F(t,x) = \partial g_t(x)$, then $F(t,\cdot)$ is a maximal monotone operator. Such systems, involving time-dependent subdifferentials, have been particularly studied in (Arseni-Benou et al., 1999; Kandilakis, 1996; Kartsatos and Parrott, 1984; Kuttler, 2000; Otani, 1994; Yamazaki, 2005) under varying degrees of regularity on the system data. Imposing further structure on the operator $F(t,\cdot)$, if we take $F(t,x) = \partial\psi_{S(t)}(x)$, where $S : [0,\infty) \rightrightarrows \mathbb{R}^n$ is closed and convex-valued mapping and $\psi_{S(t)}$ is the indicator function associated with $S(t)$, then the resulting dynamics have been more commonly studied under the topic of *sweeping processes.* Starting from the seminal work of (Moreau, 1977), the research in this area has grown to study several generalizations of the fundamental model, see for example, the monographs

⋆ Corresponding author: Aneel Tanwani. This work is partially supported by ANR project ConVan, grant number ANR-17-CE40-0019-01. The manuscript was first submitted on February 3rd, 2020.

(Adly, 2018; Monteiro Marques, 1993) for an overview. Besides the cases where F is expressed as a subdifferential of a convex function, certain classes of evolution variational inequalities (Brogliato and Tanwani, 2020; Pang and Stewart, 2008) can also be embedded in the framework of (1). Moving away from the problems related to existence and uniqueness of solutions, we see that the researchers have started addressing analysis and design related questions for system class (1); for example, use of optimization algorithms for computing Lyapunov functions for stability analysis (Camlibel et al., 2006; Souaiby et al., 2021), and some control design problems (Tanwani et al., 2018; Cao et al., 2021).

In this paper, we consider the problem of studying solutions of (1) using a time-discretization algorithm. Our approach builds on using the time-stepping algorithm pioneered in (Moreau, 1977), which was also used for studying existence of solutions for system (1) in (Kunze and Monteiro Marques, 1997). This algorithm constructs a sequence of solutions, where each element of the sequence is an interpolation of points obtained by applying the proximal operator. When the differential inclusion (1) is obtained from a constrained optimization problem, we can naturally make connections with the solution of the differential inclusion and a generalized proximal algorithm used for solving the optimization problem. The basic toolset used in this paper comes from the field of convex analysis, and we refer the interested to (Bauschke and Combettes, 2017), and (Hiriart-Urruty and Lemaréchal, 2001) for the basic definitions and results.

2. BACKGROUND AND MOTIVATION

Consider a continuously differentiable convex function $f : \mathbb{R}^n \to \mathbb{R}$, and the following unconstrained optimization problem:

$$\min_{z\in\mathbb{R}^n} f(z). \tag{2}$$

Commonly employed gradient descent algorithm for solving this optimization problem is

$$z_{k+1} = z_k - h\nabla f(z_k) \tag{3}$$

where h is the step size, and z_0 is an initial condition, that need to be chosen for implementing this algorithm. One can easily draw comparisons between the gradient descent algorithm (3), and the gradient flow dynamics, where the later is described by

$$\dot{x} = -\nabla f(x). \tag{4}$$

Indeed, if we partition an interval $[0,T]$ as $\{0 = t_0, t_1, \dots, t_N = T\}$, so that $t_{i+1} - t_i = h$, for each $i \geqslant 0$, the corresponding Euler discretization of (4) matches (3), that is, $x(t_{i+1}) = x(t_i) - h\nabla f(x(t_i))$. By choosing $x(0) = z_0$, it can be established (with ∇f Lipschitz) that

$$|x(t_i; z_0) - z_i| = O(h), \quad \text{as} \quad h \to 0, \tag{5}$$

where $x(t_i; z_0)$ denotes the solution of (4) at time $t_i \in [0,T]$ with initial condition $x(t_0) = z_0$.

In case of constrained optimization over a closed convex set $\mathcal{C}$,

$$\min_{z\in\mathcal{C}} f(z), \tag{6}$$

we reformulate the problem as,

$$\min_{z\in\mathbb{R}^n} f(z) + \psi_{\mathcal{C}}(z) \tag{7}$$

where $\psi_{\mathcal{C}}$ denotes the indicator function of the set $\mathcal{C}$ defined as

$$\psi_{\mathcal{C}}(z) = \begin{cases} 0, & \text{if } z \in \mathcal{C}, \\ +\infty & \text{if } z \notin \mathcal{C}. \end{cases}$$

It is a common practice to consider the proximal gradient descent algorithm, which is described as,

$$z_{k+1} = \operatorname{prox}_{\psi_{\mathcal{C}}}(z_k - h\nabla f(z_k)), \tag{8}$$

where $\operatorname{prox}_{\psi_{\mathcal{C}}}$ is the *proximal operator* of $\psi_{\mathcal{C}}$, and is defined as,

$$\operatorname{prox}_{\psi_{\mathcal{C}}}(v) = \operatorname{argmin}_{x\in\mathcal{C}} \|x - v\|^2.$$

The gradient flow dynamics associated with (7) are described as,

$$\dot{x} \in -\nabla f(x) - \partial\psi_{\mathcal{C}}(x) \tag{9}$$

where $\partial\psi_{\mathcal{C}} : \mathbb{R}^n \to \mathbb{R}^n$ denotes the subdifferential [1] of $\psi_{\mathcal{C}}$, and is defined as,

$$v \in \partial\psi_{\mathcal{C}}(x) \quad \Leftrightarrow \quad \langle v, z - x\rangle \leqslant 0, \quad \forall\, z \in \mathcal{C}.$$

It is a natural question to ask how the iterates in (8) relate to the solution of the differential inclusion (9). Here again, one can draw similar connections between the two as in the case of unconstrained optimization using the results on maximal monotone differential inclusions, that is, the estimate (5) holds. We will provide the details of this observation in the next section as a particular case of our main result.

In this paper, we are interested in optimization problems with time-varying constraint sets. That is, given a set-valued mapping $\mathcal{C} : [0,T] \rightrightarrows \mathbb{R}^n$, such that for each $t \in [0,T]$, the set $\mathcal{C}(t)$ is closed and convex, we are interested in solving

$$\min_{z\in\mathcal{C}(t)} f(z) \tag{10}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable convex function such that $\nabla f(\cdot)$ is Lipschitz continuous. The minimizer of the optimization problem (10) is in general time-dependent, that is, $z_t^\star \in \mathcal{C}(t)$. There is no proximal gradient descent algorithm for this type of problems, but we are interested in connections between the algorithm solving (10) and the corresponding gradient flow dynamics:

$$\dot{x} \in -\nabla f(x) - \partial\psi_{\mathcal{C}(t)}(x). \tag{11}$$

The right-hand side of (11), for each t, is a maximal monotone operator. We are interested in knowing whether, under certain cases,

$$\lim_{t\to\infty} |x(t; x_0) - z_t^\star| = 0.$$

Thus, $x(t; x_0)$, obtained as a solution of the differential inclusion (11), approximates the minimizer of (10) for sufficiently large values of $t \geqslant 0$. In what follows, we study discretization of (11) which indeed allows us to compute an approximation of z_t^* in some appropriate sense.

3. MAIN RESULTS

Based on the discussions in the previous section, our starting point is the following data:

(H1) A convex, continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ such that ∇f is Lipschitz continuous.

[1] We recall that $v \in \mathbb{R}^n$ is a subgradient of a convex function $g : \mathbb{R}^n \to \mathbb{R}$ at $x \in \operatorname{dom} g$, if for all $z \in \operatorname{dom} g$, we have $\langle v, z - x\rangle \leqslant g(z) - g(x)$. The subdifferential of g at $x \in \operatorname{dom} g$, denoted by $\partial g(x)$, is the closed convex set of all subgradients of g at x.

(H2) A function $g : [0,T] \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ such that $g(t,\cdot)$ is proper, convex and lower semicontinuous for each $t \in [0,T]$.

For simplicity, we denote the function $g(t,\cdot)$ by $g_t(\cdot)$. Associated with functions $f(\cdot)$ and $g_t(\cdot)$, we consider the following differential inclusion which is obtained by taking the generalized gradient of their sum:

$$\dot{x} \in -\nabla f(x) - \partial g_t(x), \qquad x(0) \in \operatorname{dom} \partial g_0(\cdot), \qquad (12)$$

where $\nabla f(\cdot)$ denotes the conventional gradient of the function $f(\cdot)$, and $\partial g_t(\cdot)$ denotes the subdifferential of the convex function $g_t(\cdot)$. It readily follows that the right-hand side of (12) is a maximal monotone operator, for each $t \geqslant 0$, whose domain is $\operatorname{dom} \partial g_t(\cdot)$. We first provide a time-discretization algorithm for (12).

3.1 Discretization of (12)

To describe the discretization of (12), let us take $\Delta = \{t_0, t_1, \dots, t_{K_\Delta} : 0 = t_0 < t_1 < \cdots < t_k < t_{k+1} < \cdots < t_{K_\Delta} = T\}$ to be a partition of the interval $[0,T]$. We call K_Δ the *size* of the partition of Δ, and we denote the *granularity* of Δ by h_Δ, that is,

$$0 < h_\Delta := \max_{k \in \{1,2,\dots,K_\Delta\}} t_k - t_{k-1}.$$

Assuming for the sake of simplicity that the nodes $\{t_i\}_{i=0}^{K_\Delta}$ are equidistant, we therefore have $h_\Delta = t_k - t_{k-1}$ for each $k \in \{1,2,\dots,K_\Delta\}$. For simplicity, we write $K = K_\Delta$, and $h = h_\Delta$, when Δ is clear from the context.

Next, for a fixed partition of the interval $[0,T]$ denoted by Δ, consider the discretization of (12), that is,

$$\frac{x_{k+1} - x_k}{h} \in -\nabla f(x_k) - \partial g_{t_{k+1}}(x_{k+1}) \qquad (13)$$

for $k \in \{0,1,\dots,K-1\}$. Alternatively, we have

$$\boxed{x_{k+1} = \big(I + h\,\partial g_{t_{k+1}}(\cdot)\big)^{-1}(x_k - h\nabla f(x_k))} \qquad (14)$$

where $x_0 \in \operatorname{dom} \partial g_0(\cdot)$ is the initial condition of (12).

3.2 Approximation Result

We now formally state the result which shows that the discrete-time sequence in (14) indeed approximates the solution of (12) under following assumptions:

(H3) The set-valued mapping $\operatorname{dom} g : [0,T] \rightrightarrows \mathbb{R}^n$ is closed and convex-valued, and satisfies

$$\sup_{z \in \operatorname{dom} g(s,\cdot)} \operatorname{dist}\big(z, \operatorname{dom} g(t,\cdot)\big) \leqslant \varphi(t) - \varphi(s),$$

for all s,t with $0 \leqslant s \leqslant t \leqslant T$, and some nondecreasing absolutely continuous function $\varphi : [0,T] \to \mathbb{R}^n$.

(H4) There exists a continuous $\sigma : [0,T] \to \mathbb{R}_+$ such that

$$|\operatorname{proj}(0, \partial g(t,x))| \leqslant \sigma(t)(1 + |x|)$$

for all $t \in [0,T]$ and $x \in \operatorname{dom} \partial g_t(\cdot)$.

(H5) The mapping $t \mapsto \operatorname{graph} \partial g(t,\cdot)$ is outer semicontinuous.

Within this setup, we can now state the following result:

Theorem 1. Consider the gradient flow dynamics (12) such that the mappings f, g satisfy hypotheses **(H1)**, **(H2)**, **(H3)**, **(H4)** and **(H5)**. For each initial condition $x_0 \in \operatorname{dom} \partial g_0(\cdot)$, system (12) admits a unique absolutely continuous solution $x : [0,T] \to \mathbb{R}^n$, and

$$\lim_{h_\Delta \to 0} |x(t_k; x_0) - x_k| = 0, \quad \forall\, k = 0, 1, \cdots, K_\Delta,$$

where Δ is a partition of $[0,T]$ with granularity h_Δ and nodes $t_0, t_1, \dots, t_{K_\Delta}$, and $\{x_k\}_{k=1}^{K_\Delta}$ is obtained from (14).

3.3 Application to Constrained Optimization

Coming back to the application discussed in Section 2, we can now talk about the performance of algorithms that can be used for solving (10) using the result of Theorem 1. To do so, we let

$$g_t = \psi_{\mathcal{C}(t)}$$

so that the differential inclusion (12) takes the form:

$$\dot{x} \in -\nabla f(x) - \mathcal{N}_{\mathcal{C}(t)}(x), \qquad x(0) \in \mathcal{C}(0), \qquad (15)$$

where $\mathcal{N}_{\mathcal{C}(t)}(x)$ denotes the outward normal cone to $\mathcal{C}(t)$ at x. To define the discretization of (15) over the interval $[0,\infty)$, we introduce the discretization nodes $t_0, t_1, \cdots$, so that $h := t_{k+1} - t_k$, for each $k \in \mathbb{N}$. Hence, the iterates in (14) take the following form:

$$\boxed{x_{k+1} = \operatorname{prox}_{\mathcal{C}(t_{k+1})}(x_k - h\nabla f(x_k))} \qquad (16)$$

where we used the fact that $\big(I + h\,\partial\psi_{\mathcal{C}(t_{k+1})}\big)^{-1} = \operatorname{prox}_{\mathcal{C}(t_{k+1})}$. To relate the solution of (10) with (15) and (16), we introduce the following assumptions:

(A1) The optimal solution of (10), denoted by z_t^*, is a solution of (15).

(A2) For each $x_0 \in \mathcal{C}(0)$, the corresponding solution $x(t;x_0)$ has the property that

$$\lim_{t \to \infty} |x(t;x_0) - z_t^*| = 0.$$

If $\mathcal{C}(\cdot)$ is a constant set-valued map, then **(A1)** and **(A2)** hold if f satisfies **(H1)**. Let us now look at an example with $\mathcal{C}(\cdot)$ time-varying where **(A1)** and **(A2)** hold as well.

Example 1. Let $f(x) := x^2$ and $\mathcal{C}(t) := [t, t+1]$, $t \geqslant 0$. The optimal solution to (10) is

$$z_t^* = t.$$

It is readily seen that

$$\mathcal{N}_{\mathcal{C}(t)}(x) = \begin{cases} 0, & \text{if } x \in]t, t+1[, \\ \mathbb{R}_-, & \text{if } x = t, \\ \mathbb{R}_+, & \text{if } x = t+1, \end{cases}$$

and the solution of the differential inclusion

$$\dot{x} \in -2x - \mathcal{N}_{\mathcal{C}(t)}(x),$$

with $x(0) = x_0 \in [0,1]$, is exactly given by

$$x(t;x_0) = \begin{cases} e^{-2t}x_0, & \text{if } e^{-2t}x_0 \geqslant t \\ t, & \text{otherwise.} \end{cases}$$

As a result, **(A1)** and **(A2)** are satisfied.

In general, it remains to be seen under what additional assumptions on the dynamics, it can be guaranteed that **(A1)** and **(A2)** hold. Here, we observe that these assumptions combined with Theorem 1 yield the following result:

Corollary 2. Consider the optimization problem (10) under assumptions **(A1)** and **(A2)**. Suppose that f satisfies **(H1)**, and the set-valued mapping $\mathcal{C}(\cdot)$ is closed and convex-valued such that

$$d_{\text{Haus}}(\mathcal{C}(s), \mathcal{C}(t)) \leqslant \varphi(t) - \varphi(s), \tag{17}$$

for a nondecreasing locally absolutely continuous function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$, and all $t \geqslant s \geqslant 0$. For each $h > 0$ sufficiently small, consider the unbounded sequence $t_{k+1}^h := t_k^h + h$, $k \in \mathbb{N}$, $t_0^h = 0$. It then holds that

$$\lim_{h \to 0, k \to \infty} \left| z_{t_k^h}^* - x_k \right| = 0,$$

where $z_{t_k^h}^* := \operatorname{argmin}_{z \in \mathcal{C}(t_k^h)} f(z)$, and x_k is obtained by recursion from (16).

The statement can be proved by checking that the conditions stated on the function f and the set-valued mapping $\mathcal{C}(\cdot)$ lead to hypotheses **(H1)**–**(H5)** in the statement of Theorem 1. Moreover, under **(A1)** and **(A2)**, any solution of (15) converges to the optimal solution of (10), denoted by z_t^*. Thus, the discrete approximants (16) indeed approach the minimizer as time gets large and sampling time gets small.

3.4 Discussions around Corollary 2

Let us provide some comments on the additional assumptions **(A1)** and **(A2)** used in Corollary 2.

(1) The assumption **(A1)** always holds if $\mathcal{C}$ in (10) is closed, convex, and time-invariant. This is because the minimizer z^* of (10) satisfies

$$0 \in \nabla f(z^*) + \partial \psi_{\mathcal{C}}(z^*).$$

Such a point z^* is exactly an equilibrium point of (15). From Theorem 1, the solutions of (15) are unique and hence **(A1)** holds.

(2) To see that **(A1)** may not hold in general with $\mathcal{C}$ time-varying, we present an example where the optimal solution of (10) is *not* a solution of the inclusion (15).

Example 2. Let $f(x) := x^2$ and let $\mathcal{C}(t)$ be defined as

$$\mathcal{C}(t) := \begin{cases} [t, t+1], & \text{if } 0 \leqslant t \leqslant 1, \\ [e^{-4(t-1)}, 1 + e^{-4(t-1)}], & \text{if } t \geqslant 1. \end{cases} \tag{18}$$

In this case, the optimal solution of (10) is given by

$$z_t^* = \begin{cases} t, & \text{if } 0 \leqslant t \leqslant 1, \\ e^{-4(t-1)}, & \text{if } t \geqslant 1. \end{cases} \tag{19}$$

It is easily checked that z_t^* is not a solution of (15), as the unique solution of (15) with $x(0) = 0$ is

$$x(t; 0) = \begin{cases} t, & \text{if } 0 \leqslant t \leqslant 1, \\ e^{-2(t-1)}, & \text{if } t \geqslant 1. \end{cases} \tag{20}$$

(3) To provide an interpretation of **(A2)**, note that if z_t^* is a solution (15), then **(A2)** requires all solutions of (15), starting with different initial conditions, to converge to z_t^*. Such a property of the dynamical systems is studied under the notion of incremental stability. Design of feedback controls, which guarantee convergence with respect to a desired trajectory in the presence of constraints, have been studied in (Tanwani et al., 2018). Once again, we observe that, if $\mathcal{C}$ is time-invariant, incremental stability boils down to the stability of an equilibrium point z^*, and such standard stability notions for constrained dynamical systems have been studied in (Goeleven and Brogliato, 2004; Camlibel et al., 2006; Souaiby et al., 2021).

3.5 Quadratic program and complementarity systems

Let us use the result of Corollary 2 for the case when the optimization problem in (10) is described by a quadratic f, and the set $\mathcal{C}$ is time-invariant and described by linear inequalities. More precisely, we consider

$$\mathbb{R}^n \ni x \mapsto f(x) := x^\top A x + b^\top x \tag{21}$$

where $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix, and $b \in \mathbb{R}^n$ is a known vector, and the constraint set

$$\mathcal{C} := \{x \in \mathbb{R}^n \,|\, Cx + d \geqslant 0\} \tag{22}$$

for some matrix $C \in \mathbb{R}^{p \times n}$ and a vector $d \in \mathbb{R}^p$. Using some basic relations from convex analysis, it can be shown that, system (15) is described by,

$$\dot{x} = -Ax - b + C^\top \eta \tag{23a}$$

$$0 \leqslant \eta \perp Cx + d \geqslant 0. \tag{23b}$$

In other words, (15) takes the form of a linear complementarity system, where the expression in (23b) means that, $\eta \geqslant 0$, $Cx + d \geqslant 0$, and $\eta^\top (Cx + d) = 0$. One can easily verify that the minimizer of the function $f(\cdot)$ in (21) over the set $\mathcal{C}$ given in (22), as described by the Karush-Kuhn-Tucker conditions, is exactly the equilibrium point of the system (23). Since the system (23) has a unique solution (Camlibel and Schumacher, 2016), **(A1)** holds in this case. Conditions for stability of system (23) can be derived using the results given in (Goeleven and Brogliato, 2004; Souaiby et al., 2021), which basically follows due to positive definiteness of the matrix A. It can also be verified that the conditions **(H1)**–**(H5)** hold in this case. One can therefore apply the result of Corollary 2, and compute the minimizer of $f(\cdot)$ subject to the constraint $\mathcal{C}$ by discretizing (23), which is described by the following recursion:

$$x_{k+1} = -Ax_k - b + C^\top \eta_{k+1}$$

$$0 \leqslant \eta_{k+1} \perp Cx_{k+1} + d \geqslant 0.$$

We refer the reader to (Acary et al., 2010) for more discussions on numerical aspects concerning simulation of complementarity systems.

4. PROOF SKETCH OF THEOREM 1

In this section, we provide a sketch of the proof of Theorem 1. Here, we only provide the main steps involved in the proof and refer the reader to (Camlibel et al., 2021) for all the details.

4.1 Step 1: Getting the bounds

Let φ satisfy **(H3)** and let α be such that

$$\alpha = |x_0| + \varphi(T) - \varphi(0). \tag{24}$$

Take σ as in **(H4)** and let β, and γ be such that

$$\beta = \alpha + \varphi(T) - \varphi(0) + (1 + \alpha) \int_0^T \sigma(s)\, ds \tag{25}$$

$$\gamma = \beta + \varphi(T) - \varphi(0) \tag{26}$$

Define $\phi : [0, T] \to \mathbb{R}_+$ by

$$\phi(t) := t + 2\varphi(t) + (1 + \gamma) \int_0^t \sigma(s)\, ds \quad \forall t \in [0, T]. \tag{27}$$

With the help of these definitions, we provide uniform bounds on x_k values in the following lemma. These bounds are required for invoking the convergence theorems.

Lemma 3. For any partition Δ, we have

$$|x_k| \leqslant \beta \tag{28}$$

$$|x_k - x_{k-1}| \leqslant \phi(t_k) - \phi(t_{k-1}) \tag{29}$$

for each $k \in \{1, 2, \ldots, K_\Delta\}$.

4.2 Construction of a sequence of approximate solutions

Based on the x_k values, we construct a sequence of absolutely continuous (in time) functions which approximate the actual solution of the system. To this end, note that the function ϕ defined above is strictly increasing and absolutely continuous. Now, define the piecewise continuous function x_Δ as

$$x_\Delta(t) := \frac{\phi(t_{k+1}) - \phi(t)}{\phi(t_{k+1}) - \phi(t_k)} x_k + \frac{\phi(t) - \phi(t_k)}{\phi(t_{k+1}) - \phi(t_k)} x_{k+1} \tag{30}$$

where $t \in [t_k, t_{k+1}]$ and $k \in \{0, 1, \ldots, K-1\}$. By definition, x_Δ is a continuous function and

$$x_\Delta(t_k) = x_k \tag{31}$$

for all $k \in \{0, 1, \ldots, K\}$. We will show that

$$x(t) := \lim_{|\Delta| \to 0} x_\Delta(t)$$

is the unique solution to the inclusion (12). An important intermediate step in studying the convergence of the sequence x_Δ is to obtain the following uniform bound.

Lemma 4. Let $\underline{\tau}$ and $\overline{\tau}$ be such that $0 \leqslant \underline{\tau} < \overline{\tau} \leqslant T$. For any partition Δ, it holds that

$$|x_\Delta(\overline{\tau}) - x_\Delta(\underline{\tau})| \leqslant \phi(\overline{\tau}) - \phi(\underline{\tau}). \tag{32}$$

4.3 Limit of the sequence

The bounds established in Lemma 3 and Lemma 4 allow us to study the limiting behaviour of the sequence $(x_{\Delta_\ell})_{\ell \in \mathbb{N}}$.

Lemma 5. Consider a sequence of partitions $(\Delta_\ell)_{\ell \in \mathbb{N}}$ with $|\Delta_\ell| \to 0$ as ℓ tends to infinity. The sequence $(x_{\Delta_\ell})_{\ell \in \mathbb{N}}$ is equicontinuous.

Let $(\Delta_\ell)_{\ell \in \mathbb{N}}$ be a sequence of partitions with $|\Delta_\ell| \to 0$ as ℓ tends to infinity. Since the sequence $(x_{\Delta_\ell})_{\ell \in \mathbb{N}}$ is also uniformly bounded in view of Lemma 3, Arzelá-Ascoli theorem implies that it converges uniformly to a continuous function x on a subsequence. We claim that x is absolutely continuous. To see this, let $\underline{\tau}, \overline{\tau} \in [0, T]$ with $\underline{\tau} \leqslant \overline{\tau}$ and note that

$$\begin{aligned} |x(\overline{\tau}) - x(\underline{\tau})| &\leqslant |x(\overline{\tau}) - x_{\Delta_\ell}(\overline{\tau})| + |x_{\Delta_\ell}(\overline{\tau}) - x_{\Delta_\ell}(\underline{\tau})| \\ &\quad + |x_{\Delta_\ell}(\underline{\tau}) - x(\underline{\tau})| \\ &\leqslant |x(\overline{\tau}) - x_{\Delta_\ell}(\overline{\tau})| + \phi(\overline{\tau}) - \phi(\underline{\tau}) \\ &\quad + |x_{\Delta_\ell}(\underline{\tau}) - x(\underline{\tau})| \\ &\leqslant \phi(\overline{\tau}) - \phi(\underline{\tau}) \end{aligned} \tag{33}$$

where the first inequality follows from the triangle inequality, the second from (32), and the third by taking the limit on the convergent subsequence N. Thus, absolute continuity of x follows from absolute continuity of the function ϕ.

Now, we want to show that x is a solution of (12), that is

$$x(t) \in \operatorname{dom} \partial g_t(\cdot) \quad \text{and} \quad \dot{x}(t) \in -\nabla f(x(t)) - \partial g_t(x(t)) \tag{34}$$

for almost all $t \in [0, T]$.

Let the set $\Gamma \subseteq [0, T]$ be defined by $\Gamma = \{t \in (0, T) :$ ϕ and x are differentiable at t and $t \notin \cup_{\ell \in N} \Delta_\ell\}$. Since ϕ and x are both absolutely continuous and $\cup_{\ell \in N} \Delta_\ell$ is countable, it is enough to show (34) for almost all $t \in \Gamma$.

For a partition Δ, define

$$y_\Delta(t) = \frac{x_{k+1} - x_k}{\phi(t_{k+1}) - \phi(t_k)} \tag{35}$$

for $t \in (t_k, t_{k+1})$ and $y_\Delta(t_k) = 0$ for $t_k \in \Delta$.

From (30), we see that

$$\dot{x}_{\Delta_\ell}(t) = \dot{\phi}(t) \frac{x_{k+1} - x_k}{\phi(t_{k+1}) - \phi(t_k)} = \dot{\phi}(t) y_{\Delta_\ell}(t) \tag{36}$$

for all $t \in \Gamma$.

In view of (30) and Lemma 4, we see that $|y_{\Delta_\ell}|_{L_\infty} \leqslant 1$ for all ℓ. Therefore, the sequence $(y_{\Delta_\ell})_{\ell \in N}$ is contained in the closed ball with radius $\sqrt{\phi(T) - \phi(0)}$ of the Hilbert space $L_2(d\phi, [0, T], \mathbb{R}^n)$. As such, there exists a subsequence N' of N such that $(y_{\Delta_\ell})_{\ell \in N'}$ converges to y weakly in $L_2(d\phi, [0, T], \mathbb{R}^n)$. It then follows that

$$\dot{x}(t) = \dot{\phi}(t) y(t) \tag{37}$$

for almost all $t \in \Gamma$.

Now, let $t^* \in \Gamma$. Then, for every $\ell \in N$, there must exist $k_\ell \in \{1, 2, \ldots, K(\Delta_\ell)\}$ with the property that $t_{k_\ell} < t^* < t_{k_\ell + 1}$. Note that $\lim_{\ell \uparrow \infty} t_{k_\ell} = \lim_{\ell \uparrow \infty} t_{k_\ell + 1} = t^*$ since $|\Delta_\ell|$ converges to zero as ℓ tends to infinity. By construction, we have

$$\left(x_{t_{k_\ell + 1}}, -\frac{x_{t_{k_\ell + 1}} - x_{t_{k_\ell}}}{t_{k_\ell + 1} - t_{k_\ell}} - \nabla f(x_{t_{k_\ell}}) \right) \in \operatorname{graph} \partial g(t_{k_\ell + 1}, \cdot).$$

Equivalently, we have

$$\left(x_{\Delta_\ell}(t_{k_\ell + 1}), -\frac{\phi(t_{k_\ell + 1}) - \phi(t_{k_\ell})}{t_{k_\ell + 1} - t_{k_\ell}} y_{\Delta_\ell}(t) - x_{\Delta_\ell}(t_{k_\ell}) \right) \in \operatorname{graph} \partial g(t_{k_\ell + 1}, \cdot). \tag{38}$$

Let $S_\ell(t^*) := -\frac{t_{k_\ell + 1} - t_{k_\ell}}{\phi(t_{k_\ell + 1}) - \phi(t_{k_\ell})} \Big(\partial g(t_{k_\ell + 1}, x_{\Delta_\ell}(t_{k_\ell + 1})) + \nabla f(x_{\Delta_\ell}(t_{k_\ell})) \Big)$. From (38), we have that $y_{\Delta_\ell}(t^*) \in S_\ell(t^*)$. It can be shown that

$$y(t^*) \in \operatorname{cl} \left(\operatorname{conv} \left(\limsup_{\ell \to \infty} S_\ell(t^*) \right) \right).$$

Due to the outer-semicontinuity assumption, we have $\limsup_{\ell \to \infty} \partial g(t_{k_\ell + 1}, x_{\Delta_\ell}(t_{k_\ell + 1})) \subseteq \partial g(t^*, x(t^*))$. The set $\partial g(t^*, x(t^*))$ is closed and convex because of the maximal monotonicity property, and hence

$$y(t^*) \in \frac{-1}{\dot{\phi}(t^*)} \Big(\partial g(t^*, x(t^*)) + \nabla f(x(t^*)) \Big).$$

Since $\dot{\phi}(t^*) \geqslant 1$, we get

$$\dot{x}(t^*) \overset{(37)}{=} \dot{\phi}(t^*) y(t^*) \in -\nabla f(x(t^*)) - \partial g(t^*, x(t^*))$$

for each $t^* \in \Gamma$.

5. CONCLUSIONS

We considered the problem of relating discretization algorithms for differential inclusions with maximal monotone operators and first order gradient descent algorithms. The emphasis is on the case where the objective function is described as a sum of two functions, one of which is possibly time-varying and extended real-valued. We show that, under certain assumptions, one can approximate the value of the minimizer by the solution of the differential inclusion.

Several questions raised in this manuscript need further investigation. While most of the hypotheses required for Theorem 1 seem rather natural, there is a possibility to work out the proof without the need of **(H4)**. Our paper (Camlibel et al., 2021) provides an alternate condition based on the Yosida approximation, and it needs to be seen if this alternate route allows us to study a broader class of optimization problems.

More investigation is required for better understanding **(A1)** and **(A2)**, the two assumptions that were required for applying Theorem 1 in the context of constrained optimization. As stated earlier, **(A1)** does not necessarily hold for differential inclusions when $\mathcal{C}(\cdot)$ is time-varying, and it is of interest to identify the class of set-valued mappings where the stationary solution of the gradient flow dynamics is the minimizer of the corresponding optimization problem. Similarly, **(A2)** refers to incremental stability for the system class (15) with respect to a nominal trajectory. Stability of (15) with respect to a stationary equilibrium has received considerable attention in the literature, but studying convergence of trajectories to another system trajectory needs more attention.

REFERENCES

Acary, V., Bonnefon, O., and Brogliato, B. (2010). Time-stepping numerical simulation of switched circuits within the nonsmooth dynamical systems approach. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 29(7), 1042–1055.

Adly, S. (2018). *A Variational Approach to Nonsmooth Dynamics: Applications in Unilateral Mechanics and Electronics*. SpringerBriefs in Mathematics. Springer International Publishing.

Arseni-Benou, K., Halidias, N., and Papageorgiou, N. (1999). Nonconvex evolution inclusions generated by time-dependent subdifferential operators. *J. Applied Mathematics and Stochastic Analysis*, 12(3), 233–252.

Attouch, H. and Peypouquet, J. (2019). Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. *Mathematical Programming, Ser. B*, 174(1-2), 391–432.

Bauschke, H. and Combettes, P. (2017). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, NewYork, 2nd edition.

Brézis, H. (1973). *Operateurs Maximaux Monotones*. North-Holland.

Brogliato, B. and Tanwani, A. (2020). Dynamical systems coupled with monotone set-valued operators: Formalisms, applications, well-posedness, and stability. *SIAM Review*, 62, 3–129.

Camlibel, M.K., Iannelli, L., and Tanwani, A. (2021). Convergence of proximal solutions for evolution inclusions with time-dependent maximal monotone operators. *Mathematical Programming, Ser. A*. DOI: 10.1007/s10107-021-01666-7.

Camlibel, M.K., Pang, J.S., and Shen, J. (2006). Lyapunov stability of complementarity and extended systems. *SIAM Journal on Optimization*, 17(4), 1056–1101.

Camlibel, M.K. and Schumacher, J.M. (2016). Linear passive systems and maximal monotone mappings. *Mathematical Programming, Ser. B*, 157(2), 397–420.

Cao, T., Colombo, G., Mordukhovich, B., and Nguyen, D. (2021). Optimization and discrete approximation of sweeping processes with controlled moving sets and perturbations. *J. Differential Equations*, 274, 461–509.

Goeleven, D. and Brogliato, B. (2004). Stability and instability matrices for linear evolution variational inequalities. *IEEE Transactions on Automatic Control*, 49(4), 521–534.

Hiriart-Urruty, J.B. and Lemaréchal, C. (2001). *Fundamentals of Convex Analysis*. Grundlehren Text Editions. Springer Verlag.

Kandilakis, D. (1996). Nonlinear evolution equations involving time-dependent subdifferentials of opposite sign. *Nonlinear Analysis: Theory, Methods & Applications*, 27(3), 321–326.

Kartsatos, A. and Parrott, M. (1984). Functional evolution equations involving time dependent maximal monotone operators in banach spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 8(7), 817–833.

Kunze, M. and Monteiro Marques, M.D.P. (1997). BV solutions to evolution problems with time-dependent domains. *Set-Valued Analysis*, 5, 57–72.

Kuttler, K. (2000). Evolution inclusions for time dependent families of subgradients. *Applicable Analysis*, 76(3-4), 185–201.

Monteiro Marques, M.D.P. (1993). *Differential Inclusions in Nonsmooth Mechanical Problems: Shocks and Dry Friction*, volume 9 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser.

Moreau, J.J. (1977). Evolution problem associated with a moving convex set in a Hilbert space. *J. Differential Equations*, 26, 347–374.

Nesterov, Y. (2013). Gradient methods for minimizing composite objective function. *Mathematical Programming*, 140(1), 125–161.

Otani, M. (1994). Nonlinear evolution equations with time-dependent constraints. *Advances in Mathematical Sciences and Applications*, 3.

Pang, J.S. and Stewart, D. (2008). Differential variational inequalities. *Mathematical Programming, Ser. A*, 113, 345–424.

Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Foundations & Trends in Optimization*, 1(3), 127–239.

Souaiby, M., Tanwani, A., and Henrion, D. (2021). Cone-copositive Lyapunov functions for complementarity systems: Converse result and polynomial approximation. *IEEE Transactions on Automatic Control*. DOI: 10.1109/TAC.2021.3061557.

Su, W., Boyd, S., and Candes, E. (2014). A differential equation for modeling nesterov's accelerated gradient method: Theory and insights. In *Advances in Neural Information Processing Systems 27*, 2510–2518. Curran Associates, Inc.

Tanwani, A., Brogliato, B., and Prieur, C. (2018). Well-posedness and output regulation for implicit time-varying evolution variational inequalities. *SIAM J. Control and Optimization*, 56(2), 751–781.

Yamazaki, N. (2005). Nonlinear evolution equations with time-dependent constraints. *Hokkaido University Preprint Series in Mathematics*, 696, 1–16.