Implementing adjoint subgradient evaluation for use in global dynamic optimization

Yulan Zhang [a], Kamil A. Khan [a,1]

Department of Chemical Engineering, McMaster University, Hamiton, ON, Canada

Abstract

Global dynamic optimization problems emerge in several engineering applications, such as global optimal control, parameter estimation, and safety verification. Such problems are often represented as nonlinear optimization problems with embedded parametric ordinary differential equations (ODEs). Typically, deterministic methods for global optimization employ subgradients of convex relaxations, to construct lower bounds that provide crucial global intuition. New results show that convex relaxation subgradients for dynamic optimization problems may be obtained by adapting standard forward or adjoint sensitivity approaches for smooth problems, and combining these with generalized McCormick relaxations. This adjoint approach aims to compute subgradients of an optimization problem's objective function directly, and ought to be computationally favorable except for small problems. However, established adjoint implementations are incompatible with established software libraries for evaluating subgradient information of an ODE's right-hand side function, such as EAGO.jl and MC++. Thus, this article describes a new fully-automated proof-of-concept implementation of our adjoint subgradient evaluation approach in C++, by combining the convex relaxation package MC++, the ODE solver CVODES, and our own differentiation and code generation tools. In this implementation, the adjoint sensitivity system can be constructed with either the forward mode or the reverse mode of automatic differentiation, adapting recent subgradient propagation approaches. Numerical examples are presented for illustration.

Keywords

Global dynamic optimization, Convex relaxations, Ordinary differential equations, Adjoint sensitivity analysis.

I. Introduction

This article considers sensitivity analysis methods for a nonlinear dynamic optimization problem with an embedded system of parametric ordinary differential equations (ODEs), expressed as:

$$\begin{aligned} \min_{\mathbf{p}^{\mathrm{L}}\leq\mathbf{p}\leq\mathbf{p}^{\mathrm{U}}} \quad & J(\mathbf{p}) := g(t_f,\mathbf{p},\mathbf{x}(t_f,\mathbf{p})) \\ \text{s.t.} \quad & \dot{\mathbf{x}}(t,\mathbf{p}) = \mathbf{f}(t,\mathbf{p},\mathbf{x}), \quad \forall t\in(t_0,t_f], \\ & \mathbf{x}(t_0,\mathbf{p}) = \mathbf{x}_0(\mathbf{p}). \end{aligned} \tag{1}$$

In this formulation, $\mathbf{p}\in\mathbb{R}^{n_p}$ denotes system parameters and $\mathbf{x}(t,\mathbf{p})\in\mathbb{R}^{n_x}$ denotes state variables, solving an ODE over the time horizon $I := [t_0,t_f]$. $J:\mathbb{R}^{n_p}\to\mathbb{R}$ is an objective function based on a cost function $g: I\times P\times\mathbb{R}^{n_x}\to\mathbb{R}$. We consider applications in which the optimization problem (1) must be solved to global optimality. Such global dynamic optimization problems arise in a wide variety of engineering applications, including global optimal control (Houska and Chachuat, 2014), parameter estimation (Esposito and Floudas, 2000), and safety verification (Huang et al., 2002).

For global optimization problems, deterministic methods require lower bounds of the globally optimal objective function value. To construct a lower bound of J in (1) this way, we seek to minimize a convex relaxation of J, which is typically built from convex relaxations of the ODE solution $\mathbf{x}$ with respect to $\mathbf{p}$, known as *state relaxations*. Scott and Barton (2013) developed a general ODE relaxation framework that describes nonsmooth state relaxations for ODE solutions as the solutions of auxiliary ODE systems. There are currently two established state relaxation methods (Scott and Barton, 2013; Song and Khan, 2021) in this framework, which construct different right-hand side (RHS) functions in the auxiliary ODE system. The RHS functions constructed by Scott and Barton (2013) employ generalized McCormick relaxations (Scott et al., 2011), and we call the resulting state relaxations the *Scott-Barton state relaxations*. As an alternative, our previous work (Song and Khan, 2021) constructs auxiliary RHS functions as optimal-value functions with embedded convex optimization problems whose objective functions employ convex and concave relaxations of $\mathbf{f}$.

Subgradients provide global sensitivity information for convex relaxations, and are used in typical convex minimization methods. However, established methods for subgradient propagation are somewhat limited for ODE state relaxations.

[1] Corresponding author: Kamil A. Khan
(E-mail: `kamilkhan@mcmaster.ca`).

Some established propagation methods (Khan and Barton, 2014, 2017; Hannemann-Tamas et al., 2015) require solving auxiliary discontinuous ODE systems, and thus require either tailored solvers that are not yet implemented, or assurance that nonsmoothness in RHS functions will generally be avoided. Another approach (Khan, 2018) requires either differentiability or convexity of the RHS itself, and applies to earlier state relaxation formulations, but not the later superior formulations of Scott and Barton (2013); Song and Khan (2021).

Our new subgradient evaluation framework (Song, 2021; Song and Khan, 2022) describes valid subgradients for the two established state relaxations (Scott and Barton, 2013; Song and Khan, 2021). These subgradients are described as solving a forward auxiliary parametric ODE system, analogously to classical forward sensitivity analysis (summarized by Hartman, 2002, Theorem 3.1) for smooth dynamic systems. For Scott-Barton relaxations, Song and Khan (2022) also developed an adjoint ODE subgradient evaluation system, which essentially extends the adjoint sensitivity analysis methods (Cao et al., 2003) for smooth dynamic systems to nonsmooth convex relaxations, without the differentiability or transversality requirements of Hannemann-Tamas et al. (2015). In principle, such an adjoint ODE system permits evaluation of subgradients of the objective function J in (1) without constructing full generalized derivatives of state relaxations.

Thus, this article is concerned with the implementation of our new adjoint ODE subgradient evaluation system, to obtain subgradients of convex relaxations of the objective function J of (1). In this implementation, the adjoint ODE system's RHS can be constructed by either the forward-mode automatic differentiation (AD) procedure for subgradient computation described by (Mitsos et al., 2009), or the corresponding reverse AD mode described by (Beckers et al., 2012). We expect that this implementation would reduce computational effort for exploring sensitivity information that is required by nonsmooth optimizers, and ultimately speed up an overarching global dynamic optimization method.

This article is structured as follows. Section II establishes the notational conventions we employ. Section III formalizes the goal of this article, and Section IV summarizes relevant mathematical preliminaries. Section V describes our new proof-of-concept adjoint sensitivity implementation in C++, and Section VI presents two numerical examples for illustration.

II. Notation

Throughout this article, scalars will be denoted as lowercase letters (e.g. $\xi \in \mathbb{R}$), vectors will be denoted as boldface lowercase letters (e.g. $\boldsymbol{\xi} \in \mathbb{R}^n$), and the i^{th} row of any vector is denoted as ξ_i. Matrices are denoted as boldface uppercase letters (e.g. $\mathbf{M} \in \mathbb{R}^{m\times n}$), and the i^{th} row of a matrix is denoted as $m_{(i)} \in \mathbb{R}^m$. Given any quantity, an over-dot (e.g. $\dot{\mathbf{x}}$) indicates the partial derivative with respect to t (e.g. $\frac{\partial x}{\partial t}$). Sets are denoted as uppercase letters (e.g. $X \subset \mathbb{R}^n$). Intervals in $\mathbb{R}^n$ are denoted either as uppercase letters or explicitly in terms of their bounds (e.g. $X := [\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}]$). The standard Euclidean norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$ are employed.

The following definitions of *subgradients* and *subdifferentials* are standard in convex analysis.

Definition 2.1. Consider a convex set $X \in \mathbb{R}^n$ and a convex function $h^{\mathrm{cv}} : X \to \mathbb{R}$. A vector $\mathbf{s}^{\mathrm{cv}} \in \mathbb{R}^n$ is a *subgradient* of h^{cv} at $\mathbf{y} \in X$ if

$$h^{\mathrm{cv}}(\mathbf{x}) \geq h^{\mathrm{cv}}(\mathbf{y}) + \langle s^{\mathrm{cv}}, \mathbf{x}-\mathbf{y}\rangle, \quad \forall \mathbf{x} \in X.$$

The affine mapping $\mathbf{x} \mapsto h^{\mathrm{cv}}(\mathbf{y}) + \langle s^{\mathrm{cv}}, \mathbf{x}-\mathbf{y}\rangle$ is called a *subtangent* of h^{cv} at $\mathbf{y}$. The collection of all subgradients of h^{cv} at $\mathbf{y}$ is the *subdifferential* $\partial h^{\mathrm{cv}}(\mathbf{y}) \subset \mathbb{R}^n$.

Analogously, given a concave function $h^{\mathrm{cc}} : X \to \mathbb{R}$, a vector $\mathbf{s}^{\mathrm{cc}} \in \mathbb{R}^n$ is a subgradient of h^{cc} at $\mathbf{y} \in X$ if

$$h^{\mathrm{cc}}(\mathbf{x}) \leq h^{\mathrm{cc}}(\mathbf{y}) + \langle s^{\mathrm{cc}}, \mathbf{x}-\mathbf{y}\rangle, \quad \forall \mathbf{x} \in X,$$

the affine mapping $\mathbf{x} \mapsto h^{\mathrm{cc}}(\mathbf{y}) + \langle s^{\mathrm{cc}}, \mathbf{x}-\mathbf{y}\rangle$ is called a subtangent of h^{cc} at $\mathbf{y}$, and the subdifferential $\partial h^{\mathrm{cc}}(\mathbf{y})$ is the collection of all subgradients of h^{cc} at $\mathbf{y}$.

III. Problem formulation

To formalize the considered global dynamic optimization problem (1), we first suppose that the following assumption holds.

Assumption 3.1. Given an open convex set $D \subset \mathbb{R}^{n_x}$, and an interval $P := [\mathbf{p}^{\mathrm{L}}, \mathbf{p}^{\mathrm{U}}] \subset \mathbb{R}^{n_p}$, suppose that a function $\mathbf{x}_0 : P \to D$ is Lipschitz continuous. Given an interval $I := [t_0, t_f] \subset \mathbb{R}$, suppose that a continuous function $\mathbf{f} : I \times P \times D \to \mathbb{R}^{n_x}$ satisfies the following conditions:

- the mapping $\mathbf{f}(\cdot, \mathbf{p}, \boldsymbol{\eta})$ is (Lebesgue) measurable on I, uniformly over $(\mathbf{p}, \boldsymbol{\eta}) \in P \times D$,
- for each $\mathbf{p} \in P$, there exists a solution $\mathbf{x}(\cdot, \mathbf{p})$ in D on I of the following ODE system:

$$\begin{aligned} &\dot{\mathbf{x}}(t,\mathbf{p}) = \mathbf{f}(t,\mathbf{p},\mathbf{x}), \quad \forall t \in (t_0, t_f], \\ &\mathbf{x}(t_0,\mathbf{p}) = \mathbf{x}_0(\mathbf{p}), \end{aligned} \tag{2}$$

- there exists $l \geq 0$, so that the following Lipschitz condition holds for each $t \in I$, $\mathbf{p} \in P$, and $\mathbf{x}^{\mathrm{A}}, \mathbf{x}^{\mathrm{B}} \in D$:

$$\|\mathbf{f}(t,\mathbf{p},\mathbf{x}^{\mathrm{A}}) - \mathbf{f}(t,\mathbf{p},\mathbf{x}^{\mathrm{B}})\| \leq l\|\mathbf{x}^{\mathrm{A}} - \mathbf{x}^{\mathrm{B}}\|.$$

Under this assumption, the ODE solution $\mathbf{x}$ is unique. Furthermore, if we replace $\mathbf{f}$ with an appropriate Lipschitz extension on the domain $I \times P \times \mathbb{R}^{n_x}$, then the global existence of $\mathbf{x}$ is guaranteed.

Definition 3.2. Continuous functions $\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}} : I \to \mathbb{R}^{n_x}$ are called *state bounds* for (2) on $I \times P$ if, $\mathbf{x}^{\mathrm{L}}(t) \leq \mathbf{x}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{U}}(t)$ for all $(t,\mathbf{p}) \in I \times P$.

Definition 3.3. Given a convex set $Y \subset \mathbb{R}^n$ and a function $\mathbf{h} : Y \to \mathbb{R}^m$, a function $\mathbf{h}^{\mathrm{cv}} : Y \to \mathbb{R}^m$ is a *convex relaxation* of $\mathbf{h}$ on Y if, for all $i \in \{1, \cdots, m\}$, ${h_i}^{\mathrm{cv}}(\mathbf{y})$ is convex on Y and ${h_i}^{\mathrm{cv}}(\mathbf{y}) \leq h_i(\mathbf{y})$, for each $\mathbf{y} \in Y$.

Similarly, a function $\mathbf{h}^{\mathrm{cc}} : Y \to \mathbb{R}^m$ is a *concave relaxation* of $\mathbf{h}$ on Y if, for all $i \in \{1, \cdots, m\}$, ${h_i}^{\mathrm{cc}}(\mathbf{y})$ is concave on Y and ${h_i}^{\mathrm{cc}}(\mathbf{y}) \geq h_i(\mathbf{y})$, $\forall \mathbf{y} \in Y$.

Definition 3.4 (Scott and Barton (2013))**.** Continuous functions $\mathbf{x}^{\text{cv}}, \mathbf{x}^{\text{cc}} : I \times P \to \mathbb{R}^{n_x}$ are called *state relaxations* for (2) on $I \times P$ if, for each $t \in I$, $\mathbf{x}^{\text{cv}}(t,\cdot)$ is a convex relaxation of $\mathbf{x}(t,\cdot)$ on P and $\mathbf{x}^{\text{cc}}(t,\cdot)$ is a concave relaxation of $\mathbf{x}(t,\cdot)$ on P.

Definition 3.5 (Song (2021); Song and Khan (2022))**.** Denote the interior of P as $\tilde{P}$. For any state relaxations $(\mathbf{x}^{\text{cv}}, \mathbf{x}^{\text{cc}})$ for (2), functions $\mathbf{S}^{\text{cv}}, \mathbf{S}^{\text{cc}} : I \times \tilde{P} \to \mathbb{R}^{n_x \times n_p}$ are *state relaxation subgradients* if, for each $i \in \{1, \ldots, n_x\}$, $\mathbf{p} \in \tilde{P}$ and $t \in I$, $s^{\text{cv}}_{(i)}(t,\mathbf{p})$ (resp. $s^{\text{cc}}_{(i)}(t,\mathbf{p})$) is the transpose of a subgradient of $x^{\text{cv}}_i(t,\cdot)$ (resp. $x^{\text{cc}}_i(t,\cdot)$) at $\mathbf{p}$.

Given any state relaxations $(\mathbf{x}^{\text{cv}}, \mathbf{x}^{\text{cc}})$ for (2), we consider the following convex relaxation of (1):

$$\begin{aligned} \min_{\mathbf{p} \in \mathbb{R}^{n_p}} \quad & J^{\text{cv}}(\mathbf{p}) := g^{\text{cv}}(t_f, \mathbf{p}, \mathbf{x}^{\text{cv}}(t_f,\mathbf{p}), \mathbf{x}^{\text{cc}}(t_f,\mathbf{p})) \\ \text{s.t.} \quad & \mathbf{p}^{\text{L}} \leq \mathbf{p} \leq \mathbf{p}^{\text{U}}, \end{aligned} \tag{3}$$

where $J^{\text{cv}} : \mathbb{R}^{n_p} \to \mathbb{R}$ is a convex relaxation of J, constructed by composing a generalized McCormick relaxation (Scott et al., 2011) $g^{\text{cv}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}$ of g with state relaxations for (2).

With state relaxation subgradients $(\mathbf{S}^{\text{cv}}, \mathbf{S}^{\text{cc}})$, the subgradients of J^{cv} may be computed using established subgradient propagation approaches, which we summarize in the next section. However, we expect that computing $(\mathbf{S}^{\text{cv}}, \mathbf{S}^{\text{cc}})$ via a classical forward sensitivity approach (summarized in (Song, 2021; Song and Khan, 2022)) would be computationally expensive for problem when n_x and n_p are large. With this in mind, this article focuses on implementing a newer adjoint subgradient evaluation system (Song, 2021; Song and Khan, 2022), to evaluate subgradients of J^{cv} without actually constructing $(\mathbf{S}^{\text{cv}}, \mathbf{S}^{\text{cc}})$.

IV. Preliminaries

This section summarizes relevant prior theoretical results concerning subgradient evaluation for the ODE state relaxations of Scott and Barton (2013), when applied to (3). These summaries are necessarily brief; for rigorous statements and discussion, the reader is directed to the cited references.

A. Scott-Barton state relaxations

The ODE state relaxations of Scott and Barton (2013) require access to *state bounds* $\mathbf{x}^{\text{L}}, \mathbf{x}^{\text{U}} : I \to \mathbb{R}^{n_x}$ for (2); these are functions for which

$$\mathbf{x}^{\text{L}}(t) \leq \mathbf{x}(t,\mathbf{p}) \leq \mathbf{x}^{\text{U}}(t), \qquad \forall t \in I, \quad \mathbf{p} \in P,$$

and can be furnished via Harrison's bounding method (Harrison et al., 1977).

In the Scott-Barton relaxation framework, the strict inequality $\mathbf{x}^{\text{cv}}(t,\mathbf{p}) < \mathbf{x}^{\text{cc}}(t,\mathbf{p})$ can be guaranteed by adding an arbitrarily small perturbation $\varepsilon > 0$ to the defining ODE system. Moreover, $\mathbf{x}^{\text{L}}(t) < \mathbf{x}^{\text{cv}}(t,\mathbf{p})$ and $\mathbf{x}^{\text{cc}}(t,\mathbf{p}) < \mathbf{x}^{\text{U}}(t)$ are guaranteed to be satisfied for sufficiently small P, and very often in practice. Hence we assume that (Scott and Barton, 2013) satisfies $\mathbf{x}^{\text{L}}(t) < \mathbf{x}^{\text{cv}}(t,\mathbf{p}) < \mathbf{x}^{\text{cc}}(t,\mathbf{p}) < \mathbf{x}^{\text{U}}(t)$. Under this assumption, the Scott-Barton-McCormick relaxations for (2) are as follows:

$$\begin{aligned} \dot{\mathbf{x}}^{\text{cv}}(t,\mathbf{p}) &= \mathbf{f}^{\text{cv}}(t,\mathbf{p},\mathbf{x}^{\text{cv}}(t,\mathbf{p}),\mathbf{x}^{\text{cc}}(t,\mathbf{p})),\ \mathbf{x}^{\text{cv}}(t_0,\mathbf{p}) = \mathbf{x}^{\text{cv}}_0(\mathbf{p}), \\ \dot{\mathbf{x}}^{\text{cc}}(t,\mathbf{p}) &= \mathbf{f}^{\text{cc}}(t,\mathbf{p},\mathbf{x}^{\text{cv}}(t,\mathbf{p}),\mathbf{x}^{\text{cc}}(t,\mathbf{p})),\ \mathbf{x}^{\text{cc}}(t_0,\mathbf{p}) = \mathbf{x}^{\text{cc}}_0(\mathbf{p}), \end{aligned} \tag{4}$$

where continuous functions $\mathbf{x}^{\text{cv}}_0, \mathbf{x}^{\text{cc}}_0 : P \to \mathbb{R}^{n_x}$ are respectively convex and concave relaxations for $\mathbf{x}_0$ on P, and continuous functions $\mathbf{f}^{\text{cv}}, \mathbf{f}^{\text{cc}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ are constructed by combining generalized McCormick relaxations (Scott et al., 2011) of $\mathbf{f}$ in (2) with interval flattening operations, within the interval bounds $[\mathbf{x}^{\text{L}}(t), \mathbf{x}^{\text{U}}(t)]$. This auxiliary ODE (4) is guaranteed to have valid state relaxations $\mathbf{x}^{\text{cv}}, \mathbf{x}^{\text{cc}}$ as its unique solution.

B. Established subgradient evaluation framework

As shown in (Song, 2021; Song and Khan, 2022), state relaxation subgradients $(\mathbf{S}^{\text{cv}}, \mathbf{S}^{\text{cc}})$ uniquely solve an auxiliary parametric ODE system that may be integrated simultaneously with (4), constructed as follows.

Considering the interior $\tilde{P}$ of P, functions $\mathbf{V}, \mathbf{W} : I \times \tilde{P} \times \mathbb{R}^{n_x \times n_p} \times \mathbb{R}^{n_x \times n_p} \to \mathbb{R}^{n_x \times n_p}$ are called *subgradient propagation functions* for $(\mathbf{f}^{\text{cv}}, \mathbf{f}^{\text{cc}})$ if, regarding $\mathbf{f}^{\text{cv}}, \mathbf{f}^{\text{cc}}$ as flattened generalized McCormick relaxations of $\mathbf{f}$, $\mathbf{V}$ and $\mathbf{W}$ are the corresponding operations that would be evaluated as `cvsub` and `ccsub` in MC++ (Chachuat, 2001) or as `cv_grad` and `cc_grad` in McCormick.jl (Wilhelm and Stuber, 2020). These are computed in MC++ and McCormick.jl according to a tailored forward mode of automatic differentiation (AD) (Mitsos et al., 2009), although in principle they could be computed by a corresponding reverse AD mode (Beckers et al., 2012).

Suppose we have subgradient propagation functions $\mathbf{V}, \mathbf{W}$ available. Consider functions $\mathbf{S}^{\text{cv}}_0, \mathbf{S}^{\text{cc}}_0 : \tilde{P} \to \mathbb{R}^{n_x \times n_p}$, so that for each $i \in \{1, \ldots, n_x\}$ and $\mathbf{p} \in \tilde{P}$, $s^{\text{cv}}_{(i),0}(\mathbf{p})$ is the transpose of a subgradient of $x^{\text{cv}}_{i,0}$ at $\mathbf{p}$, and $s^{\text{cc}}_{(i),0}(\mathbf{p})$ is the transpose of a subgradient of $x^{\text{cc}}_{i,0}$ at $\mathbf{p}$. Consider the following ODE system:

$$\begin{aligned} \dot{\mathbf{S}}^{\text{cv}}(t,\mathbf{p}) &= \mathbf{V}(t,\mathbf{p},\mathbf{S}^{\text{cv}}(t,\mathbf{p}),\mathbf{S}^{\text{cc}}(t,\mathbf{p})),\ \ \mathbf{S}^{\text{cv}}(t_0,\mathbf{p}) = \mathbf{S}^{\text{cv}}_0(\mathbf{p}), \\ \dot{\mathbf{S}}^{\text{cc}}(t,\mathbf{p}) &= \mathbf{W}(t,\mathbf{p},\mathbf{S}^{\text{cv}}(t,\mathbf{p}),\mathbf{S}^{\text{cc}}(t,\mathbf{p})),\ \mathbf{S}^{\text{cc}}(t_0,\mathbf{p}) = \mathbf{S}^{\text{cc}}_0(\mathbf{p}). \end{aligned} \tag{5}$$

For each $\mathbf{p} \in \tilde{P}$, the uniqueness and local existence of $\mathbf{S}^{\text{cv}}(\cdot,\mathbf{p})$ and $\mathbf{S}^{\text{cc}}(\cdot,\mathbf{p})$ on I are guaranteed (Song and Khan, 2022), and $\mathbf{S}^{\text{cv}}, \mathbf{S}^{\text{cc}}$ are state relaxation subgradients for (2).

C. Adjoint subgradient evaluation system

State relaxation subgradients $(\mathbf{S}^{\text{cv}}, \mathbf{S}^{\text{cc}})$ can be computed numerically by solving (5) in the "forward mode" of increasing t, analogously to the classical forward sensitivity analysis approach (Hartman, 2002, Theorem 3.1) for smooth systems. However, $(\mathbf{S}^{\text{cv}}, \mathbf{S}^{\text{cc}})$ may be computationally expensive to evaluate when n_x and n_p are large. Here we note that, in smooth dynamic optimization, adjoint sensitivity analysis (Cao et al. (2003)) provides an alternative way to evaluate derivatives of an objective function without computing the partial derivatives of state variables with respect to system parameters.

Given an ODE system as described in (2), let $\boldsymbol{\lambda} \in \mathbb{R}^{n_x}$ be an adjoint variable, and denote the partial derivatives of a function by its subscript. The sensitivity of a function $g(t_f, \mathbf{p}, \mathbf{x})$ with respect to $\mathbf{p}$ at t_f is computed through the following adjoint ODE system directly:

$$\begin{cases} \dot{\boldsymbol{\lambda}}(t) = -\mathbf{f}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\lambda}(t), \\ \boldsymbol{\lambda}(t_f) = g_{\mathbf{x}}^{\mathrm{T}}, \end{cases} \tag{6}$$

and thus $\frac{dg}{d\mathbf{p}}(t_f) = \boldsymbol{\lambda}^{\mathrm{T}}(t_0)\mathbf{x}_{\mathbf{p}}(t_0) + g_{\mathbf{p}}(t_f) + \int_{t_0}^{t_f} \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{f}_{\mathbf{p}}\,dt$. The term $\int_{t_0}^{t_f} \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{f}_{\mathbf{p}}\,dt$ can be computed simultaneously during numerical integration of (6).

Song (2021); Song and Khan (2022) extended this adjoint sensitivity analysis approach to nonsmooth subgradient evaluation for the problem (3) with (2) embedded, as summarized in the following proposition. Through this approach, reformulating the ODE system (5) into an adjoint ODE formulation allows evaluating subgradients of J^{cv} with respect to $\mathbf{p}$ without computing $(\mathbf{S}^{\mathrm{cv}}, \mathbf{S}^{\mathrm{cc}})$.

Here we suppose that subgradient propagation functions $(\mathbf{V}, \mathbf{W})$ are available, and that these are affine with the form:

$$\begin{bmatrix} \mathbf{V}(t,\mathbf{p},\mathbf{M},\mathbf{N}) \\ \mathbf{W}(t,\mathbf{p},\mathbf{M},\mathbf{N}) \end{bmatrix} \equiv \Theta_{\mathrm{A}}(t,\mathbf{p}) \begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix} + \Theta_{\mathrm{B}}(t,\mathbf{p}).$$

As shown by Song and Khan (2022), this is actually the case for the generalized McCormick relaxations.

Proposition 4.1 (Song (2021); Song and Khan (2022))**.** Consider a lower-bounding problem (3), where state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ are constructed by solving (4). Given a fixed point $\tilde{\mathbf{p}} \in \tilde{P}$, and a transposed subgradient $\boldsymbol{\rho}$ of $g^{\mathrm{cv}}(t_f, \cdot, \cdot, \cdot)$ at $(\tilde{\mathbf{p}}, \mathbf{x}^{\mathrm{cv}}(t_f, \tilde{\mathbf{p}}), \mathbf{x}^{\mathrm{cc}}(t_f, \tilde{\mathbf{p}}))$, the transpose $\tilde{\mathbf{s}}$ of some subgradient of J^{cv} at $\tilde{\mathbf{p}}$ is described as follows. Let $\boldsymbol{\lambda} : I \to \mathbb{R}^{2n_x}$ be a Carathéodory solution (Filippov, 1988) on I of the following adjoint ODE system:

$$\dot{\boldsymbol{\lambda}}(t) = -\Theta_{\mathrm{A}}^{\mathrm{T}}(t, \tilde{\mathbf{p}})\boldsymbol{\lambda}(t), \quad \boldsymbol{\lambda}(t_f) = \boldsymbol{\rho}^{\mathrm{T}}. \tag{7}$$

Then, $\tilde{\mathbf{s}}$ is represented as:

$$\tilde{\mathbf{s}} \equiv (\boldsymbol{\lambda}(t_0))^{\mathrm{T}}\mathbf{S}_0^{\mathrm{C}} + \boldsymbol{\rho} + \int_{t_0}^{t_f} (\boldsymbol{\lambda}(t))^{\mathrm{T}}\Theta_{\mathrm{B}}(t, \tilde{\mathbf{p}})\,dt. \tag{8}$$

The term $\int_{t_0}^{t_f} (\boldsymbol{\lambda}(t))^{\mathrm{T}}\Theta_{\mathrm{B}}(t, \tilde{\mathbf{p}})\,dt$ can be integrated simultaneously with (7).

V. Implementation

A proof-of-concept C++17 implementation was developed to evaluate subgradients of the objective function relaxation J^{cv} in (3) with respect to $\mathbf{p}$, with the ultimate goal of aiding deterministic algorithms for global dynamic optimization. In this implementation, Harrison's bounding method (Harrison et al., 1977) is applied with MC++ v1.0 (Chachuat, 2001) to compute state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ automatically, and the right-hand sides $\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}}$ in the Scott-Barton relaxations framework (4) are also constructed automatically via MC++. The corresponding adjoint ODE subgradient evaluation system is established according to (7), by using Julia's operator overloading capabilities to generate the corresponding adjoint right-hand sides as C++ code. The Scott-Barton relaxations and this adjoint system are integrated together by CVODES v5.8.0 (Hindmarsh et al., 2005), which is an ODE solver capable of performing adjoint sensitivity analysis. Our overall approach is depicted in Figure 1.

In our Julia code generation step, we can employ either the forward-mode AD of (Mitsos et al., 2009) or the reverse-mode AD of (Beckers et al., 2012) methods to construct the required right-hand sides. The forward mode subgradient AD for computing $(\mathbf{V}, \mathbf{W})$ was implemented via MC++, and we have developed a ReverseADforVW module in Julia v1.0.5 (Bezanson et al., 2017) to aid the implementation of reverse mode subgradient AD (Beckers et al., 2012) in C++. For any user-defined factorable function, ReverseADforVW automatically constructs its computational graph and generates C++ code for computing subgradient propagation functions $(\mathbf{V}, \mathbf{W})$ in reverse mode. Furthermore, a C++ class called RevMcCormick was developed to store subgradient values for each intermediate quantity of a factorable function during the reverse AD sweep.

For comparison, we also evaluate the subgradients of J^{cv} in (3) with respect to $\mathbf{p}$ using finite difference approximation in MATLAB. However, we note that it is possible for a finite difference approximation of subgradients of J^{cv} to be poor even in the absence of numerical error (Song et al., 2021), and the numerical ODE solution contributes additional error, so these are unreliable in global dynamic optimization.

All computation in this section was conducted in a Lenovo Laptop with 1.8 GHz AMD Ryzen 7 4800U CPUs and 16.0GB of RAM.

VI. Examples

This section presents two numerical examples using our new implementation. First, the following example shows that the new adjoint ODE subgradient evaluation system (7) appears to yield valid subgradients of an optimization problem's objective function, when the adjoint system is constructed using forward-mode AD.

Example 6.1. Consider the following nonlinear ODE system from (Song and Khan, 2021, Example 2) with one state variable x on a time interval $I := [0.0, 0.15]$ and one parameter $p \in P := [-2, 2]$:

$$\begin{aligned} &\dot{x}(t) = p(x^2 - 1), \quad t \in I \\ &x(0) = -2, \end{aligned} \tag{9}$$

Suppose we seek to minimize $x(0.15, p)$ in (1). A corresponding lower-bounding problem with (9) embedded is then:

$$\begin{aligned} \min_{p \in P} \quad & J^{\mathrm{cv}}(p) := x^{\mathrm{cv}}(t_f, p) \\ \text{s.t.} \quad & -2 \leq p \leq 2, \end{aligned} \tag{10}$$

where the function $J^{\mathrm{cv}} : P \to \mathbb{R}$ is a convex relaxation of $J \equiv x(t_f, \cdot)$ on P based on a convex relaxation $x^{\mathrm{cv}} : I \times P \to \mathbb{R}$ of x in (9) on P.

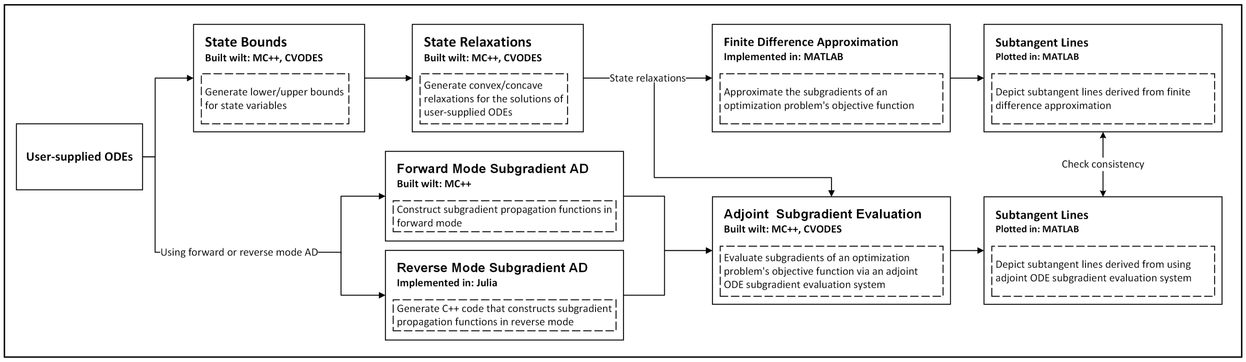

Figure 1: An overview of the new implementation

Scott-Barton relaxations $(x^{\mathrm{cv}}, x^{\mathrm{cc}})$ for (9) on $I \times P$ were generated numerically by applying our C++ implementation, with a subgradient $\tilde{s}$ of J^{cv} at $\tilde{p} := -1$ generated numerically by applying our C++ implementation of the new adjoint ODE system (7). This adjoint ODE system's RHS was automatically constructed with forward-mode subgradient AD (Mitsos et al., 2009) for constructing $(\mathbf{V}, \mathbf{W})$.

In this example, $\tilde{s}$ is numerically equal to the subgradient of $x^{\mathrm{cv}}(t_f, \tilde{p})$ with respect to $\tilde{p}$. Similarly, the subgradient of $x^{\mathrm{cc}}(t_f, \tilde{p})$ with respect to $\tilde{p}$ can be computed by constructing a concave relaxed problem of (1). For comparison, we also applied finite difference approximation in MATlAB for which the established state relaxation is differentiable.

Figure 2 illustrates the objective function J, along with convex/concave relaxations and subtangent lines derived from $\tilde{s}$ and finite difference approximations. Observe that the generated subtangents do indeed appear to be correct.

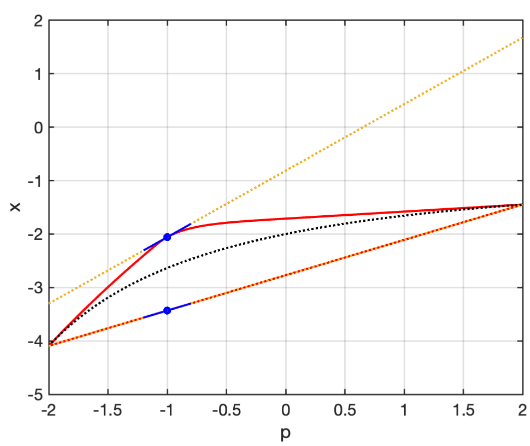

Figure 2: The objective function J in Example 1 (dotted black), along with corresponding Scott-Barton relaxations (solid red), subtangent lines obtained by a new adjoint ODE system (solid blue) and finite difference approximations (dashed yellow), plotted against p and $t_f = 0.15$ fixed.

In the following example, we show that the adjoint ODE subgradient evaluation system (7) can employ the reverse-mode subgradient AD (Beckers et al., 2012) for computing $(\mathbf{V}, \mathbf{W})$ to generate valid subgradients for an objective function in lower-bounding problems.

Example 6.2 (adapted from Scott and Barton (2013))**.** Consider parameters $\mathbf{p} \in P := [-6.5, 6.5] \times [0.01, 0.5]$, and a time interval $I := [0.0, 2.0]$. Consider the following instance of (2):

$$\begin{aligned} \dot{x}_1 &= -(2 + \sin(p_1/3))x_1^2 + p_2 x_1 x_2, \quad & x_1(0, \mathbf{p}) = 1, \\ \dot{x}_2 &= \sin(p_1/3)x_1^2 - p_2 x_1 x_2, & x_2(0, \mathbf{p}) = 0.5. \end{aligned} \tag{11}$$

Suppose we seek to minimize $x(2.0, p)$ in (1). A corresponding lower-bounding problem with (11) embedded is then:

$$\begin{aligned} \min_{\mathbf{p} \in P} \quad & J^{\mathrm{cv}}(\mathbf{p}) := x_2^{\mathrm{cv}}(t_f, \mathbf{p}) \\ \text{s.t.} \quad & \mathbf{p}^{\mathrm{L}} \leq \mathbf{p} \leq \mathbf{p}^{\mathrm{U}}, \end{aligned} \tag{12}$$

where the function $J^{\mathrm{cv}} : P \to \mathbb{R}$ is a convex relaxation of $J \equiv x(t_f, \cdot)$ on P based on a convex relaxation $x^{\mathrm{cv}} : I \times P \to \mathbb{R}$ of x in (11) on P.

Scott-Barton relaxations $(x^{\mathrm{cv}}, x^{\mathrm{cc}})$ for (9) on $I \times P$ were generated numerically by applying our C++ implementation. Then, a subgradient $\tilde{s}$ of J^{cv} at $\tilde{p} := (1, 0.5)$ was generated numerically by applying our C++ implementation of the new adjoint ODE system (7), this time using reverse-mode AD to construct this system automatically.

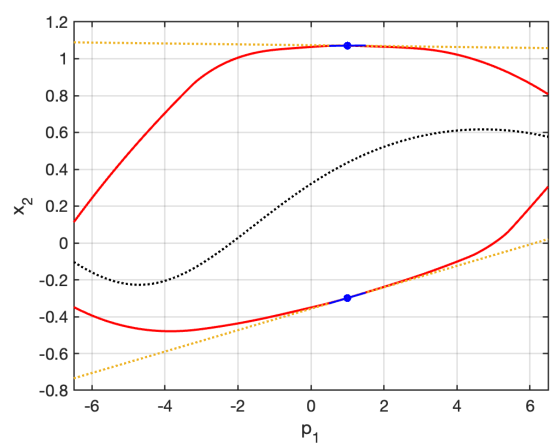

Figure 3: The objective function J in Example 2 (dotted black), along with corresponding Scott-Barton relaxations (solid red), subtangent lines obtained by a new adjoint ODE system (solid blue) and finite difference approximations (dashed yellow), plotted against p_1 with $p_2 = 0.5$ and $t_f = 2$ fixed.

Figure 3 depicts the solution of (11), state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$, subtangents derived from $\tilde{\mathbf{s}}$ and corresponding finite difference approximations. Noting the visual agreement between these subtangents and their approximations, our implementation of the new adjoint ODE subgradient propagation system (7) constructed using reverse-mode AD appears to generate valid subgradients of J^{cv} in this case.

VII. Conclusion

For parametric ODE relaxations established by Scott and Barton (2013), we have developed a proof-of-concept implementation of our new adjoint ODE subgradient evaluation approach (Song, 2021; Song and Khan, 2022) in C++17. This implementation constructs the adjoint ODE system (7) using either forward-mode (Mitsos et al., 2009) or reverse-mode (Beckers et al., 2012) subgradient AD for computing the crucial subgradient propagation functions $(\mathbf{V}, \mathbf{W})$. Numerical examples based on our implementation show that the new adjoint ODE system (7) do indeed appear to compute subgradients for a convex relaxation of the objective function of a dynamic optimization.

Future work will involve embedding this adjoint ODE subgradient evaluation implementation within methods for deterministic global dynamic optimization, and to assess their ability to enhance computational performance.

References

Beckers, M., V. Mosenkis, and U. Naumann (2012). Adjoint mode computation of subgradients for McCormick relaxations. In *Recent Advances in Algorithmic Differentiation*, pp. 103–113. Springer.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah (2017). Julia: A fresh approach to numerical computing. *SIAM Review 59*(1), 65–98.

Cao, Y., S. Li, L. Petzold, and R. Serban (2003). Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution. *SIAM Journal on Scientific Computing 24*(3), 1076–1089.

Chachuat, B. (2001). MC++: a toolkit for bounding factorable functions.
`https://github.com/coin-or/MCpp`.

Esposito, W. R. and C. A. Floudas (2000). Global optimization for the parameter estimation of differential-algebraic systems. *Industrial & Engineering Chemistry Research 39*(5), 1291–1310.

Filippov, A. F. (1988). *Differential Equations with Discontinuous Righthand Sides.* Dordrecht: Kluwer Academic Publishers.

Hannemann-Tamas, R., D. A. Munoz, and W. Marquardt (2015). Adjoint sensitivity analysis for nonsmooth differential-algebraic equation systems. *SIAM Journal on Scientific Computing 37*(5), A2380–A2402.

Harrison, G. W. et al. (1977). Dynamic models with uncertain parameters. In *Proceedings of the First International Conference on Mathematical Modeling*, Volume 1, pp. 295–304. University of Missouri Rolla.

Hartman, P. (2002). *Ordinary Differential Equations*. Philadelphia: SIAM.

Hindmarsh, A. C., P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS) 31*(3), 363–396.

Houska, B. and B. Chachuat (2014). Branch-and-lift algorithm for deterministic global optimization in nonlinear optimal control. *Journal of Optimization Theory and Applications 162*(1), 208–248.

Huang, H., C. S. Adjiman, and N. Shah (2002). Quantitative framework for reliable safety analysis. *AIChE Journal 48*(1), 78–96.

Khan, K. A. (2018). Subtangent-based approaches for dynamic set propagation. In *2018 IEEE Conference on Decision and Control (CDC)*, pp. 3050–3055. IEEE.

Khan, K. A. and P. I. Barton (2014). Generalized derivatives for solutions of parametric ordinary differential equations with non-differentiable right-hand sides. *Journal of Optimization Theory and Applications 163*(2), 355–386.

Khan, K. A. and P. I. Barton (2017). Generalized derivatives for hybrid systems. *IEEE Transactions on Automatic Control 62*(7), 3193–3208.

Mitsos, A., B. Chachuat, and P. I. Barton (2009). McCormick-based relaxations of algorithms. *SIAM Journal on Optimization 20*(2), 573–601.

Scott, J. K. and P. I. Barton (2013). Improved relaxations for the parametric solutions of ODEs using differential inequalities. *Journal of Global Optimization 57*(1), 143–176.

Scott, J. K., M. D. Stuber, and P. I. Barton (2011). Generalized McCormick relaxations. *Journal of Global Optimization 51*(4), 569–606.

Song, Y. (2021). *New Bounding Methods for Global Dynamic Optimization*. Ph. D. thesis, Mcmaster University.

Song, Y., H. Cao, C. Mehta, and K. A. Khan (2021). Bounding convex relaxations of process models from below by tractable black-box sampling. *Computers and Chemical Engineering 153*, 107413.

Song, Y. and K. A. Khan (2021). Optimization-based convex relaxations for nonconvex parametric systems of ordinary differential equations. *Mathematical Programming*, in press.

Song, Y. and K. A. Khan (2022). Computing subgradients of convex relaxations for the solutions of parametric systems of ordinary differential equations, in preparation.

Wilhelm, M. E. and M. D. Stuber (2020). EAGO.jl: easy advanced global optimization in Julia. *Optimization Methods and Software, in press*.