NEW BOUNDING METHODS FOR GLOBAL DYNAMIC OPTIMIZATION

NEW BOUNDING METHODS FOR GLOBAL DYNAMIC OPTIMIZATION

BY

YINGKAI SONG, MS, BE

A THESIS

SUBMITTED TO THE DEPARTMENT OF CHEMICAL ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING

Doctor of Philosophy (2021) McMaster University
(chemical engineering) Hamilton, Ontario, Canada

TITLE: New Bounding Methods for Global Dynamic Optimization

AUTHOR: Yingkai Song
MS (Chemical Engineering),
Carnegie Mellon University, Pittsburgh, PA, USA,
BE (Chemical Engineering),
China University of Petroleum-Beijing, Beijing, China

SUPERVISOR: Dr. Kamil A. Khan

NUMBER OF PAGES: xix, 268

Abstract

Global dynamic optimization arises in many engineering applications such as parameter estimation, global optimal control, and optimization-based worst-case uncertainty analysis. In branch-and-bound deterministic global optimization algorithms, a major computational bottleneck is generating appropriate lower bounds for the globally optimal objective value. These bounds are typically constructed using convex relaxations for the solutions of dynamic systems with respect to decision variables. Tighter convex relaxations thus translate into tighter lower bounds, which will typically reduce the number of iterations required by branch-and-bound. Subgradients, as useful local sensitivities of convex relaxations, are typically required by nonsmooth optimization solvers to effectively minimize these relaxations. This thesis develops novel techniques for efficiently computing tight convex relaxations with the corresponding subgradients for the solutions of ordinary differential equations (ODEs), to ultimately improve efficiency of deterministic global dynamic optimization.

Firstly, new bounding and comparison results for dynamic process models are developed, which are more broadly applicable to engineering models than previous results. These new results show for the first time that in a state-of-the-art ODE relaxation framework, tighter enclosures of the original ODE system's right-hand side will necessarily translate into enclosures for the state variables that are at least as tight, which paves the

way towards new advances for bounding in global dynamic optimization.

Secondly, new convex relaxations are proposed for the solutions of ODE systems. These new relaxations are guaranteed to be at least as tight as state-of-the-art ODE relaxations. Unlike established ODE relaxation approaches, the new ODE relaxation approach can employ any valid convex and concave relaxations for the original right-hand side, and tighter such relaxations will necessarily yield ODE relaxations that are at least as tight. In a numerical case study, such tightness does indeed improve computational efficiency in deterministic global dynamic optimization. This new ODE relaxation approach is then extended in various ways to further tighten ODE relaxations.

Thirdly, new subgradient evaluation approaches are proposed for ODE relaxations. Unlike established approaches that compute valid subgradients for nonsmooth dynamic systems, the new approaches are compatible with reverse automatic differentiation (AD). It is shown for the first time that subgradients of dynamic convex relaxations can be computed via a modified adjoint ODE sensitivity system, which could speed up lower bounding in global dynamic optimization.

Lastly, in the situation where convex relaxations are known to be correct but subgradients are unavailable (such as for certain ODE relaxations), a new approach is proposed for tractably constructing useful correct affine underestimators and lower bounds of the convex relaxations just by black-box sampling. No additional assumptions are required, and no subgradients must be computed at any point. Under mild conditions, these new bounds are shown to converge rapidly to an original nonconvex function as the domain of interest shrinks. Variants of the new approach are presented to account for numerical error or noise in the sampling procedure.

Acknowledgements

Looking back on my journey in the past four years, I would first like to thank to my supervisor Professor Kamil Khan for his endless patience, guidance, and support. Prof. Khan provided meticulous feedback on my manuscripts and was always reachable whenever I had difficulties. His insightful and inspiring comments helped me to form my own opinions and become an independent researcher. His simple but precise writing style is what I hope to learn and carry forward throughout my career. Thank you, Prof. Khan! I am very lucky to be one of your first PhD students. I would express my gratitude to my thesis committee, Professor Christopher Swartz and Professor Antoine Deza, for their helpful remarks about improving my research work. I am also grateful to my previous supervisors, Professor Ignacio Grossmann and Professor Xingliang Huang, for their support even after my graduation.

I am grateful to Matthew Wilhelm and Professor Matthew Stuber for helpful discussions on their Julia global optimization package EAGO, which is the fundamental platform for almost all implementations in my thesis work.

My life in MACC has been awesome. I am grateful to my MACC colleagues who made the office JHE-370 positive, supportive, and full of joyness. This pandemic is not easy for anybody, and I truly hope we meet again in the future. Special thanks to my comrade-in-arms Huiyi for enjoyable communication about our research and life over years. Hope we both have a bright future.

I would like to thank my for-whole-life friends Zuwei and Yuzhu for their everlasting support out of my professional career. It is good to know that I will always have your backup. I also appreciate all friends who have walked into my life; all of you made me a better Yingkai.

Last but not the least, it comes to the most important persons in my life, my fiancée Xiaoying and my parents Shuxia and Yongjun. I am extremely grateful to my parents for their love and dedication to my education. They have always been pushing me out of my comfort zone, convincing me that I deserve something bigger, and encouraging me to explore the fascinating world. Xiaoying, I feel blessed to meet with you on the land of maple leaf. I am willing to experience all unknowns in the future holding your hands.

The work in this thesis was supported by McMaster University's Department of Chemical Engineering and the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN-2017-05944.

Contents

Abstract **iii**

Acknowledgements **v**

Abbreviations **xviii**

1 Introduction **1**

1.1 Motivation 1

1.2 Established relaxation methods for composite functions 6

1.3 Established ODE relaxation approaches 7

1.4 Scott–Barton ODE relaxation framework 9

1.5 Goals 12

1.6 Contributions and thesis structure 13

2 Mathematical Preliminaries **16**

2.1 Notation 16

2.2 Established relaxations for composite functions 17

2.2.1 Natural interval extensions and McCormick relaxations 18

2.2.2 αBB relaxations 20

2.2.3 Auxiliary Variable Method 21

2.3 Underlying ODE system 23

2.4 Established ODE relaxation formulations 24

3 Comparing Solutions of Related ODEs Using New Differential Inequalities 31

3.1 Introduction 32

3.2 Background: Carathéodory solutions of ODE systems 39

3.3 One-sided bounding 39

3.4 Bound comparison 44

3.5 Tightness results for ODE relaxations 49

3.6 Conclusions and future work 61

4 Optimization-Based Convex Relaxations for Nonconvex Parametric Systems of ODEs 63

4.1 Introduction 64

4.2 New state relaxation formulation 67

4.3 Properties of new state relaxations 72

4.3.1 Continuity 73

4.3.2 Existence and right-uniqueness 75

4.3.3 Bounding of original ODE solution 77

4.3.4 Convexity 78

4.3.5 Tightness 80

4.3.6 Convergence as domain shrinks 89

4.3.7 Comparison to established relaxation methods 95

4.4 Implementation and examples 101

4.4.1 Implementation 101
4.4.2 Numerical examples 101
4.5 Conclusions and future work 119

5 Extending Optimization-Based Convex Relaxations for Dynamic Process Models 121

5.1 Introduction 122
5.2 Background: optimization-based relaxation formulation 126
5.3 Using closed-form minima of convex envelopes 128
5.4 New state relaxation approach 133
5.4.1 Background: convex relaxations of factorable functions 134
5.4.2 New state relaxation formulation 141
5.4.3 Comparison to optimization-based relaxations 145
5.4.4 Properties of new state relaxations 149
5.4.5 Implementation 165
5.4.6 Numerical examples 166
5.5 Conclusions and future work 181

6 Constructing Subgradients for Convex Relaxations of Nonconvex Parametric Systems of ODEs 183

6.1 Introduction 184
6.2 Mathematical preliminaries 190
6.2.1 Directional derivatives and subgradients 190
6.2.2 Directional derivatives of parametric ODE solutions 192
6.3 Problem statement 193

6.4 Subgradient evaluation framework . . . 196
6.4.1 Subgradient propagation functions . . . 196
6.4.2 Subgradient evaluation ODE system . . . 197
6.4.3 Constructing subgradient propagation functions . . . 203
6.4.4 Accommodating coinciding state relaxations . . . 210
6.4.5 Adjoint subgradient evaluation . . . 214
6.5 New subgradients of established state relaxations . . . 220
6.5.1 Subgradients of Scott–Barton–McCormick relaxations . . . 220
6.5.2 Subgradients of optimization-based relaxations . . . 227
6.6 Implementation and examples . . . 234
6.6.1 Implementation . . . 234
6.6.2 Numerical examples . . . 235
6.7 Conclusions and future work . . . 238

7 Bounding Convex Relaxations of Process Models from Below by Tractable Black-Box Sampling 240

7.1 Introduction . . . 241
7.2 New affine relaxation formulation . . . 244
7.3 Tightness and convergence properties . . . 247
7.4 Accounting for noise of sampling . . . 255
7.5 Moving the sampled set . . . 258
7.6 Conclusions and future work . . . 262

8 Conclusions and Future Work 264

8.1 Conclusions . . . 264

8.2 Future work 266

List of Figures

1.1 The original objective function c (*solid*) in (1.1.1), along with a convex relaxation (*dot-dashed*) and a tighter convex relaxation (*dashed*) with corresponding lower bounds $c^{\mathrm{L,B}}$ and $c^{\mathrm{L,A}}$ of c on $[p^{\mathrm{L}}, p^{\mathrm{U}}]$, respectively. 4

1.2 An illustration of how tighter state relaxations for x typically lead to tighter convex relaxations for c in (1.1.1). On the left: a solution $x(t_f, \cdot)$ (*solid*) of the parametric ODE system (1.1.2) at $t := t_f$, along with state relaxations (*dot-dashed*) and tighter state relaxations (*dashed*). 5

4.1 The solution $x(0.15, p)$ (*solid black*) of the parametric ODE (4.4.1) from Example 4.2, plotted against p along with corresponding state relaxations obtained by the OB-αTI method (*dotted blue*), the OB-αTV method (*dashed red*), and the PA method [1] (*squared green*). 105

4.2 A plot on logarithmic axes of $\mathrm{wid}(X_k^{\mathrm{C}}) := \sup_{\mathbf{p}\in P(k)} \left(x_{P(k)}^{\mathrm{cc}}(0.15, p) - x_{P(k)}^{\mathrm{cv}}(0.15, p)\right)$ vs. $\mathrm{wid}(P(k)) := 2^{k+1}$ (*blue circles*) for Example 4.2, for $k := -10, -9, \ldots, 0$, with $(\mathbf{x}_{P(k)}^{\mathrm{cv}}, \mathbf{x}_{P(k)}^{\mathrm{cc}})$ generated by the OB-αTV method, along with a reference line (*dotted red*) corresponding to second-order pointwise convergence. 105

4.3 A plot of the convex state relaxation discrepancy $\varepsilon^{\mathrm{cv}}(t) := \max_{p\in P}\big(x(t,p) - x^{\mathrm{cv}}(t,p)\big)$ (*dotted blue*) and the concave state relaxation discrepancy $\varepsilon^{\mathrm{cc}}(t) := \max_{p\in P}\big(x^{\mathrm{cc}}(t,p) - x(t,p)\big)$ (*dashed red*) vs. time $t \in [0.1, 0.7]$, with $(x^{\mathrm{cv}}, x^{\mathrm{cc}})$ generated by the OB-αTV method, for Example 4.3. 107

4.4 A cross-section at $(p_1, p_2, p_3, p_4, p_5, p_6, p_8) := (42.14, 116.5, 269, 50.6, 343.6, 450, 1)$ of the solution $x_5(2,\cdot)$ (*solid black*) of the ODE system (4.4.3) from Example 4.4, along with corresponding state relaxations obtained by the new OBM method (*dashed red*) and by the SBM method [2] (*dotted blue*). . . . 109

4.5 A cross-section at $(p_1, p_2, p_3, p_4, p_5, p_6, p_7) := (42.14, 116.5, 269, 50.6, 343.6, 450, 9.28)$ of the solution $x_5(2,\cdot)$ (*solid black*) of the ODE system (4.4.3) from Example 4.4, along with corresponding state relaxations obtained by the new OBM method (*dashed red*) and by the SBM method [2] (*dotted blue*). . . . 110

4.6 The solution $x(0.9,\cdot)$ (*solid black*) of the ODE (4.4.4) on P from Example 4.5, along with corresponding state relaxations obtained by the OB-ENV method (*squared green*), the SBM method [2] (*dashed red*), and the OB-αTV method (*dotted blue*). (In this example, the OB-αTI relaxations overlap with the OB-αTV relaxations, and the OB-αTV convex relaxations overlap with the SBM convex relaxations). 114

5.1 The solution $x_1(2.5,p)$ (*left, solid black*) and $x_2(2.5,p)$ (*right, solid black*) of the parametric ODEs (5.3.1) from Example 5.1, plotted against p along with corresponding SBM relaxations [2] (*dotted blue*) and the OB relaxations (*dashed red*) obtained by solving (2.4.1) with closed-form (5.3.4) and (5.3.6) embedded. (In the left subfigure, the concave SBM relaxation overlaps with the new concave relaxation for x_1.) 133

5.2 The solution $x(0.06, p)$ (*solid black*) of the parametric ODE (5.4.24) from Example 5.3, plotted against p along with corresponding SBM relaxations [2] (*dotted blue*) and the AVM-based relaxations (*dashed red*). 168

5.3 The solution $x(0.1, p)$ (*solid black*) of the parametric ODE (5.4.25) from Example 5.4, plotted against p along with corresponding SBM relaxations [2] (*dotted blue*) and the AVM-based state relaxations (*dashed red*). 172

5.4 The solution $x_2(0.02, T)$ (*solid black*) of the parametric ODE (5.4.27) from Example 5.5, plotted against T along with corresponding SBM relaxations [2] (*dotted blue*), CF-AVM-based relaxations (*dashed red*), NLP-AVM-based relaxations (*dashed green*), and LP-AVM-based relaxations (*dashed orange*). Note that the OBM relaxations [3] are identical to the SBM relaxations in this case, and all state convex relaxations overlap. 180

5.5 The solution $x_3(0.02, T)$ (*solid black*) of the parametric ODE (5.4.27) from Example 5.5, plotted against T along with corresponding SBM relaxations [2] (*dotted blue*), CF-AVM-based relaxations (*dashed red*), NLP-AVM-based relaxations (*dashed green*), and LP-AVM-based relaxations (*dashed orange*). Note that the OBM relaxations [3] are identical to the SBM relaxations in this case. 180

6.1 A cross-section at $p_2 := 0.4$ of the solution $x_1(4.0, \cdot)$ (*left, dashed black*) and $x_2(4.0, \cdot)$ (*right, dashed black*) of the ODE system (6.6.1) from Example 6.1, along with corresponding SBM relaxations (*solid red*) and subtangent cross-sections (*dotted blue*) derived from the subgradients at two reference points (*blue circle*) 236

6.2 A cross-section at $(p_2, p_3) := (8.5, 2.0)$ of the solution $x_1(0.95, \cdot)$ (*left, dashed black*) and $x_2(0.95, \cdot)$ (*right, dashed black*) of the ODE system (6.6.2) from Example 6.2, along with corresponding OB relaxations (*solid red*) and subtangent cross-sections (*dotted blue*) derived from the subgradients at two reference points (*blue circle*) 238

7.1 Illustration of the domain of the function f described in Definition 7.2.2. . . 246

7.2 Illustration of an original function ϕ (*dotted blue*), along with convex relaxations (*dash-dotted*) on various subdomains, and corresponding new sampling-based affine relaxations (*solid*), from [4, Fig. 6]. One sample data point $(\mathbf{w}^{(0)}, y_0)$ (*circle*) is depicted in each case. 247

List of Tables

3.1 Summary of established comparison results for bounding ODEs with C-solutions. 35

4.1 A list of considered state relaxation methods and their abbreviations. 102

4.2 Average computational times for evaluating state relaxations $(x^{\mathrm{cv}}(0.15,p),x^{\mathrm{cc}}(0.15,p))$ for (4.4.1) in Example 1. 104

4.3 The interval bounds of uncertain parameters **p** in Example 4.4. 108

4.4 Average computational times for evaluating state relaxations $(\mathbf{x}^{\mathrm{cv}}(2,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(2,\mathbf{p}))$ for (4.4.3) in Example 4.4. 109

4.5 Average computational times for evaluating state relaxations $(x^{\mathrm{cv}}(0.9,p),x^{\mathrm{cc}}(0.9,p))$ for (4.4.4) in Example 4.5. 113

4.6 Non-default IPOPT options specified for nonsmooth convex minimization, for Example 4.6. 116

4.7 Global optimization results for the problem (4.4.5) in Example 4.6. 116

4.8 Computational results for solving (4.4.6) with BARON in Example 4.6. . . 118

5.1 Average computational times for evaluating state relaxations $(\mathbf{x}^{\mathrm{cv}}(2.5,p),\mathbf{x}^{\mathrm{cc}}(2.5,p))$ for (5.3.1) in Example 5.1. 132

5.2 Average computational times for evaluating state relaxations $(x^{\mathrm{cv}}(0.06,p),x^{\mathrm{cc}}(0.06,p))$ for (5.4.24) in Example 5.3. 169

5.3 Average computational times for evaluating state relaxations $(x^{\mathrm{cv}}(0.1,p),x^{\mathrm{cc}}(0.1,p))$ for (5.4.25) in Example 5.4. 172

5.4 The values of the constant physical quantities in (5.4.27), for Example 5.5. . 174

5.5 Various state relaxation methods that were applied to (5.4.27), for Example 5.5. 179

5.6 Average computational times for evaluating state relaxations $(\mathbf{x}^{\mathrm{cv}}(0.02,T),\mathbf{x}^{\mathrm{cc}}(0.02,T))$ for (5.4.25) in Example 5.5. 181

Abbreviations

a.e.	"almost every" in the sense of Lebesgue measure
NLP	Nonlinear programming
LP	Linear programming
ODE	Ordinary differential equation
DAE	Differential algebraic equation
KKT	Karush–Kuhn–Tucker optimality conditions
C-solutions	Carathéodory solutions of ODE systems
RHS	Right-hand-side of ODE systems
AD	Automatic differentiation
MC	McCormick relaxations [5]
gMC	Generalized McCormick relaxations [6]
TMC	Tsoukalas–Mitsos–McCormick relaxations [7]
AVM	Auxiliary Variable Method [8]

αBB	Name of the convex relaxation scheme proposed by Adjiman et al. [9]
SBM	Scott–Barton–McCormick ODE relaxations [2]
OB	Optimization-based ODE relaxations [3]
OB-ENV	OB relaxations derived from convex envelopes
OB-αTV	OB relaxations derived from αBB relaxations with time-varying α
OB-αTI	OB relaxations derived from αBB relaxations with time-invariant α
OBM	OB relaxations derived from McCormick relaxations
PA	Dynamic αBB relaxations by Papamichail and Adjiman [1]

Chapter 1

Introduction

1.1 Motivation

This thesis considers dynamic optimization problems that are formulated as optimization problems with embedded parametric systems of ordinary differential equations (ODEs). These ODE systems may exhibit significant nonconvexity, and thus a local nonlinear programming (NLP) solver may converge to suboptimal local optima when applied to such dynamic optimization problems. Compared with suboptimal local optima, a global optimum represents the most desirable outcome subject to the predefined constraints, such as the lowest operating cost or the highest production rate. Globally optimal solutions for dynamic optimization problems are sought in engineering applications such as optimal control of batch processes [10–13], optimal catalyst blending [14], and optimal drug scheduling [15, 16]. Moreover, in several other applications, suboptimal local optima are inappropriate and a global optimum is essential. For example, a typical dynamic parameter estimation and model identification problem [17–19] aims to choose parameters to minimize the discrepancy between the prediction of a dynamic model and experiment data, and global optimization can confirm that a particular model is inappropriate regardless of parameter choice. A worst-case uncertainty analysis determines the maximal possible cost or potential safety hazard [20] based on an assessment of the probability distribution of parameter values.

For simplicity of analysis, this thesis considers a generic nonconvex dynamic optimization problem:

$$\begin{aligned} \min_{\mathbf{p}} \quad & c(\mathbf{p}) := g(t_f, \mathbf{p}, \mathbf{x}(t_f, \mathbf{p})) \\ \text{s.t.} \quad & \mathbf{p}^{\mathrm{L}} \leq \mathbf{p} \leq \mathbf{p}^{\mathrm{U}}, \end{aligned} \tag{1.1.1}$$

where $\mathbf{p} \in \mathbb{R}^{n_p}$ denotes decision variables with known bounds, $c : \mathbb{R}^{n_p} \to \mathbb{R}$ is an objective function based on a cost function g of appropriate domain and range dimension, and $\mathbf{x}$ denotes the solution of the parametric ODE system:

$$\begin{aligned} \dot{\mathbf{x}}(t, \mathbf{p}) &= \mathbf{f}(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p})), \quad \forall t \in (t_0, t_f], \\ \mathbf{x}(t_0, \mathbf{p}) &= \mathbf{x}_0(\mathbf{p}), \end{aligned} \tag{1.1.2}$$

This ODE system will be formalized in Section 2.3. Established global optimization algorithms may be classified into deterministic algorithms and stochastic algorithms. The main types of stochastic algorithms are genetic algorithms [21], differential evolution algorithms [22], and particle swarm algorithms [23], whose applications on dynamic optimization problems are discussed in [24–26]. While these stochastic global search algorithms are effective in certain situations, deterministic global optimization methods are guaranteed to locate a global optimum to within a predefined tolerance in finite computational time. Deterministic global optimization algorithms [27–31] are typically based on branch-and-bound frameworks, wherein upper and lower bounds of the globally optimal objective value are evaluated. These are then refined progressively as the considered decision space is subdivided. Upper bounds are typically computed from local minima obtained by local NLP solvers, and lower bounds are typically computed by minimizing *convex relaxations* c^{cv} of c in (1.1.1). A convex relaxation is an auxiliary function that underestimates the original function and is convex on the considered domain. Thus, any local minimum of c^{cv} obtained using local NLP solvers would also be a global minimum, which is guaranteed to be a valid lower bound of c. *Tighter* convex relaxations — namely relaxations that are pointwise closer to the original function — thus translate into tighter bounds supplied to branch-and-bound, and could in turn reduce the number of iterations required by an overarching optimization method. Figure 1.1 depicts a convex relaxation and a tighter convex relaxation of c for illustration. While state-of-the-art deterministic global optimization solvers including BARON [28] and ANTIGONE [32] are effective at solving problems that are not dynamic, current deterministic algorithms for global dynamic optimization can only currently solve problems with no more than around five state variables and five decision variables. Hence, there is a need in global dynamic optimization to develop efficient and accurate computational tools for automatically generating tight convex relaxations for dynamic systems, to ultimately extend the scope of these deterministic algorithms to problems of practical interest.

State relaxations are respective underestimators and overestimators of $\mathbf{x}$ in (1.1.2) whose components are respectively convex and concave with respect to $\mathbf{p}$ for each fixed t. State relaxations may be used in deterministic global dynamic optimization to construct convex relaxations of the objective function c of (1.1.1) by various adaptations [6, 7, 33–35] of McCormick's relaxation method [5]. Except for bounding and convexity, state relaxations should have desirable tightness and convergence properties. Firstly, state relaxations

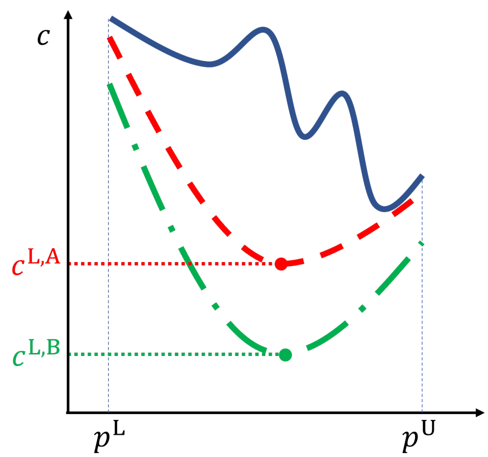

Figure 1.1: The original objective function c (*solid*) in (1.1.1), along with a convex relaxation (*dot-dashed*) and a tighter convex relaxation (*dashed*) with corresponding lower bounds $c^{\mathrm{L,B}}$ and $c^{\mathrm{L,A}}$ of c on $[p^{\mathrm{L}}, p^{\mathrm{U}}]$, respectively.

should be tight. As illustrated in Figure 1.2, tighter state relaxations for (1.1.2) will typically translate into tighter convex relaxations for c in (1.1.1) [6], which will in turn translate into tighter lower bounds in global optimization, and thus reduce the number of iterations required in a branch-and-bound algorithm. Secondly, state relaxations should converge rapidly to $\mathbf{x}$ as the domain $P := [\mathbf{p}^{\mathrm{L}}, \mathbf{p}^{\mathrm{U}}]$ is subdivided. The convex relaxations of c will necessarily inherit this rapid convergence [36]. Such rapid convergence can mitigate the *cluster effect* [37, 38], wherein a branch-and-bound method must branch many times near a global minimum even in the best case. This notion of rapid convergence has been formalized as *second-order pointwise convergence* [36]. State relaxations may also be used for similarly constructing lower bounding problems in global optimization for an original problem that is more complicated than (1.1.1), such as a dynamic optimization problem with typical path or endpoint constraints and an objective function with integral. The extension of state relaxations to parametric systems of differential-algebraic equations (DAEs) is discussed in [39].

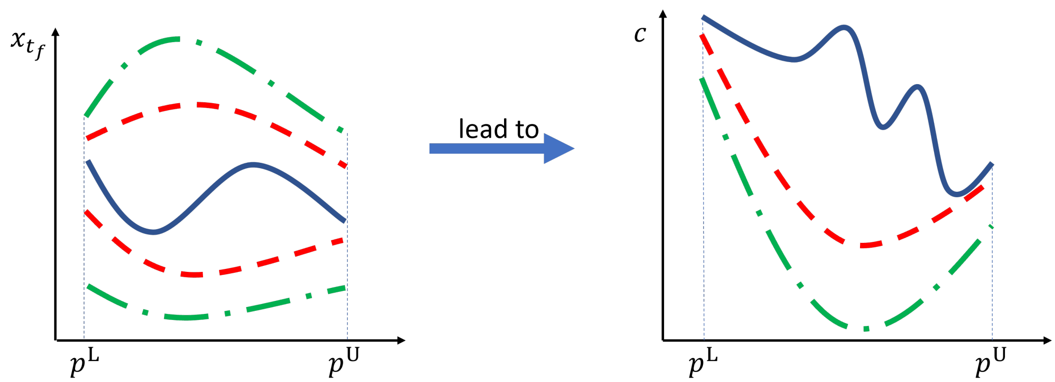

Figure 1.2: An illustration of how tighter state relaxations for x typically lead to tighter convex relaxations for c in (1.1.1). On the left: a solution $x(t_f,\cdot)$ (*solid*) of the parametric ODE system (1.1.2) at $t := t_f$, along with state relaxations (*dot-dashed*) and tighter state relaxations (*dashed*).

Subgradients are useful local sensitivities for nonsmooth convex functions (analogously for concave functions), which reduce to the usual gradients for smooth convex functions. Subgradients of state relaxations may be used to compute subgradients of convex relaxations of c [35]. When computing lower bounds by minimizing convex relaxations in global optimization, subgradients are typically required by nonsmooth convex optimization methods such as *Nesterov's Level Method* and *Subgradient Method* [40] and general nonsmooth local optimization methods such as *bundle methods* [41–43] to proceed effectively. Without subgradients, an overarching global optimization method may fail to compute the required lower bounds by minimizing convex relaxations. Moreover, subgradients are useful for constructing piecewise affine relaxations by a finite combination of the corresponding subtangents [34, 44, 45], and each subtangent can be efficiently constructed by a single evaluation of the original convex relaxation and an associated subgradient [46]. Subgradients of state relaxations are also useful in dynamic reachable set generation, for constructing a convex polyhedral enclosure of the reachable set for $\mathbf{x}$ in (1.1.2) [47].

1.2 Established relaxation methods for composite functions

This section introduces several established convex relaxation generation methods for composite functions, which may be used for relaxing ODE systems' right-hand side functions. Such relaxations may then be embedded in approaches for computing ODE relaxations.

Established convex relaxation generation methods for composite functions mainly include natural interval extensions [48], McCormick relaxations [5], generalized McCormick relaxations [6], αBB relaxations [9,49], and relaxations obtained using the Auxiliary Variable Method (AVM) [8,50]. These relaxation methods will be further illustrated via examples in Section 2.2. The natural interval extension [48] employs predefined upper and lower bounds for simple arithmetic operations that are used to define the overall composite function, and constructs bounds for the composite function by propagating the bounds for these operations. These bounds are constant on the domain of interest, and thus are trivially convex and concave. Based on the underlying bounds, the McCormick relaxation method [5] employs predefined convex and concave relaxations of intrinsic functions, and constructs closed-form nonsmooth convex and concave relaxations for a composite function in a similar fashion. The generalized McCormick relaxation method [6] adapts the McCormick relaxations to better handle each composed function in a composition. The αBB relaxation method [9,49] applies to twice-differentiable functions, and constructs convex relaxations by adding a sufficiently large convex quadratic term to the original nonconvex function. The AVM [8,50] typically applies to nonconvex optimization problems. The AVM constructs an auxiliary convex optimization problem, by first introducing auxiliary variables to capture any nonlinearities in the original nonconvex optimization problem, and then bounding these variables by appropriate convex and concave relaxations. The optimal objective values of this auxiliary optimization problem are guaranteed to be valid lower bounds of the optimal objective values of the original nonconvex problem. Moreover, Chapter 5 of this thesis will show that the AVM may be adapted to construct convex relaxations for composite functions as well. Other convex relaxations include piecewise-affine relaxations constructed from subtangents [35, 44] and convex envelope which is the tightest possible convex relaxation for a nonconvex function. Convex envelopes of certain functions [51–56] have been proposed, but there is no established method for constructing convex envelopes for general composite functions.

1.3 Established ODE relaxation approaches

This section summarizes established approaches for constructing either state relaxations or *state bounds* that are $\mathbf{p}$-invariant bounds for $\mathbf{x}$ in (1.1.2). These approaches may be classified into two broad categories: *discretize-then-relax* approaches and *relax-then-discretize* approaches. Discretize-then-relax approaches may be further classified into two subclasses based on how the ODE system (1.1.2) is handled. The first subclass [35, 57–59] discretizes the ODE system (1.1.2) into approximating equations, by approaches including the explicit Euler method or orthogonal collocation. These equations are then relaxed using algebraic relaxation methods such as αBB [57], the McCormick relaxation method [5], or the Auxiliary Variable Method [8, 60]. Most recently, Yang and Scott [59] extended the continuous-time theory of differential inequalities to bound discrete-time nonlinear dynamic systems. Wilhelm et al. [58] proposed a relaxation method efficient for stiff dynamic systems utilizing implicit functions [61].

The second subclass of discretize-then-relax approaches includes the methods in [62–66], which discretize the dynamic systems using Taylor expansions which are then bounded by various approaches. These methods originate from the high-order-enclosure (HOE) method [63], where bounds for Taylor expansions are propagated over t. The HOE method was extended to propagate McCormick relaxations [64], Taylor models [62], and McCormick-Taylor models [66]. Recently, Pérez-Galván and Bogle [67] tightened bounds for Taylor models using interval contractors. The McCormick-Taylor method [66] has been shown to yield state relaxations with second-order pointwise convergence [68], and this method was applied to solve a global optimal control problem with the branch-and-lift algorithm [12].

Relax-then-discretize approaches [1, 2, 69–76], on the other hand, are based directly on differential inequalities [77], since these preserve the differential equation nature of (1.1.2). These approaches have the advantage of being able to exploit the adaptive time-stepping and error control of numerical ODE solvers. Papamichail and Adjiman [1] extended αBB relaxations from closed-form functions to parametric ODEs (1.1.2), with the required Hessian bounds computed via numerical integration. This method was recently applied to solve global optimal control problems with the direct multiple shooting method [13]. Baja and Hasan [78] proposed dynamic edge-concave relaxations, which are also based on Hessian bounding of the ODE solutions. Other methods in this class compute state relaxations directly by solving auxiliary ODE systems, whose right-hand sides are related to various relaxations of the right-hand side $\mathbf{f}$ of (1.1.2). The function $\mathbf{f}$ has been relaxed in these methods using interval extension [69, 73, 74], affine relaxation [70–72, 76], and generalized McCormick relaxation [2, 6, 75]. As shown in [79], the relaxations proposed in [2] are at least as tight as those in [75], and have second-order pointwise convergence. Other methods [71–74, 76] in this class can incorporate additional bounding information from physical or mathematical arguments, to yield significantly tighter state bounds or relaxations when such bounding information is available. Recently, methods that combine differential inequalities and Taylor models have been also proposed [80, 81]. Other established methods for generating convex enclosures of reachable sets include bounding the reachable sets by ellipsoids [82], zonotopes [83], or polytopes [84]. The ellipsoidal calculus method was applied to account for the parameterization errors of control inputs for affine control systems in global optimal control [12].

1.4 Scott–Barton ODE relaxation framework

This section introduces a state-of-the-art relax-then-discretize framework by Scott and Barton [2], for relaxing the underlying ODE system (1.1.2). This thesis aims to improve this framework, to aid in lower bounding for global dynamic optimization. The Scott–Barton framework requires furnishing the following crucial functions:

- convex and concave relaxations $(\mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}})$ of the initial-value function $\mathbf{x}_0$,
- predefined $\mathbf{p}$-invariant bounds $(\mathbf{x}^{\mathrm{L}}\ \mathbf{x}^{\mathrm{U}})$ for $\mathbf{x}$,
- functions $(\mathbf{u}, \mathbf{o})$ (c.f. [2, Definitions 6 and 7]) that are modified relaxations of $\mathbf{f}$.

Then, valid state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ for (1.1.2) are computed by solving the following auxiliary ODE system: for each component i,

$$\begin{aligned} \dot{x}_i^{\mathrm{cv}}(t,\mathbf{p}) &= \begin{cases} u_i(t,\mathbf{p},\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}}), & \text{if } x_i^{\mathrm{cv}} > x_i^{\mathrm{L}}(t), \\ \max\left(\dot{x}_i^{\mathrm{L}}(t), u_i(t,\mathbf{p},\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})\right), & \text{if } x_i^{\mathrm{cv}} = x_i^{\mathrm{L}}(t), \end{cases} \\ x_i^{\mathrm{cv}}(0,\mathbf{p}) &= x_{i,0}^{\mathrm{cv}}(\mathbf{p}). \end{aligned} \tag{1.4.1}$$

Dynamics of $\mathbf{x}^{\mathrm{cc}}$ are described similarly. This formulation will be formalized in Section 2.4. Scott and Barton [2] proceeded to construct appropriate $(\mathbf{u}, \mathbf{o})$ based on the generalized McCormick relaxations [6] of $\mathbf{f}$. The generalized McCormick relaxation method [6] has been shown to be efficient for constructing useful convex and concave relaxations for composite functions [6, 35]. The resulting state relaxations of this approach will be denoted as the *Scott–Barton-McCormick (SBM) relaxations*. The state-of-the-art SBM relaxation approach has the following advantages over other established ODE relaxation approaches. Unlike general discretize-then-relax approaches, the SBM relaxation approach is able to exploit the adaptive time-stepping and error control of numerical ODE solvers, since the SBM relaxations are computed by solving an auxiliary ODE system. Compared with the discretize-then-relax approaches based on Taylor models [62, 63, 65, 66], the SBM relaxations are more efficient and relatively simple to implement as discussed in [2]. The SBM relaxations may also be more efficient than the αBB ODE relaxations [1], since the SBM relaxations do not require expensive second-order sensitivity information of $\mathbf{x}$. For solving a global dynamic optimization problem, the embedded ODE system may be discretized first, and then the resulting non-dynamic optimization problem is supplied to the state-of-the-art deterministic global optimization solver BARON [28, 60]. However, as will be seen in a global dynamic optimization case study in Chapter 4, this approach failed to converge to a global optimum, while a branch-and-bound algorithm with the SBM relaxations embedded successfully converged. Lastly, Schaber et al. [79] recently showed that the SBM relaxations are guaranteed to be at least as tight as the earlier relax-then-discretize ODE relaxations in [75]. This tightness result benefits from a general relaxation preserving dynamics nature that is exclusive for the Scott–Barton framework. However, while the SBM relaxations exhibit these advantages, the Scott–Barton framework has the following limitations, which impede developing new ODE relaxations that outperform the SBM relaxations:

1. In the Scott–Barton framework, it was unknown prior to this thesis whether the tightness of the original right-hand side relaxations would translate into tightness of state relaxations. Since several recent advances [7, 34, 45, 85] have been proposed for constructing tight convex relaxations for closed-form functions, new methods for constructing tighter state relaxations could significantly benefit from a greater understanding of the tightness properties of the Scott–Barton framework. Establishing the tightness results requires fundamental results for comparing ODE solutions based on differential inequalities [77, 86]. However, existing ODE comparison results [39, 77, 79, 87–89] (detailed in Chapter 3 below) are insufficient to address this problem; new ODE comparison results must be developed.

2. Prior to this thesis, only generalized McCormick relaxations [6] were allowed to be embedded into the Scott–Barton framework. If tighter non-McCormick relaxations of $\mathbf{f}$ are available, they cannot be used in this framework to yield potentially tighter state relaxations. It is not obvious how to embed non-McCormick relaxations into Scott–Barton framework in its current setting, and new versatile state relaxation formulation must be developed for this task.

3. Due to current limitations in convex analysis theory and nonsmooth dynamic sensitivity analysis, there was previously no dynamic subgradient evaluation methods for state relaxations in the Scott–Barton framework. As mentioned in Section 1.1, subgradients of convex relaxations can help nonsmooth local optimizer proceed effectively. Without subgradients, an overarching global optimization method may fail to compute the required lower bounds by minimizing convex relaxations.

1.5 Goals

This thesis aims to resolve the limitations outlined in the previous section of the Scott–Barton ODE relaxation framework [2], to ultimately improve computational efficiency of deterministic algorithms of global dynamic optimization. Specifically, the goals of this thesis are to:

1. Develop new results for comparing solutions of related ODE systems. Based on these new results, develop new tightness results for state relaxations obtained using Scott–Barton framework, to pave the way towards computing tighter state relaxations from tighter relaxations for the original right-hand side $\mathbf{f}$.

2. Develop new state relaxation approaches by using the Scott–Barton relaxation theory [2] in a new way, to yield tighter state relaxations than the SBM state relaxations in [2] and permit tighter non-McCormick relaxations for $\mathbf{f}$ to be used when these are available.

3. Develop new approaches for efficiently computing subgradients of state relaxations obtained using the Scott–Barton framework.

4. Develop new approaches for using convex relaxations with unknown subgradients in global optimization, by exploring derivative-free optimization technique.

5. Embed the new techniques for lower bounding into a branch-and-bound-based deterministic algorithms, to obtain a new efficient implementation for deterministic global dynamic optimization.

The long-term goal for this line of research is to develop deterministic global dynamic optimization algorithms that are efficient enough for engineering applications. Once the goals of this thesis are achieved, one could employ any tight convex relaxations for right-hand sides of the underlying ODE system, to construct tight ODE relaxations using the new versatile ODE relaxation formulation proposed in this thesis. Since the first approaches for computing subgradients of state-of-the-art ODE relaxations are proposed in this thesis, one could for the first time use these ODE relaxations in deterministic global dynamic optimization to compute the required lower bounds for the globally optimal objective values. All these efforts aim to improve computational efficiency for the lower-bounding procedure in an overarching global optimization method, which would ultimately facilitate the long-term goal.

1.6 Contributions and thesis structure

This thesis proposes novel state relaxation approaches for the original parametric system (1.1.2), which allow using tighter non-McCormick relaxations of $\mathbf{f}$ for constructing state relaxations, and guarantee to yield state relaxations that are at least as tight as the state-of-the-art SBM state relaxations [2]. This thesis also proposes new dynamic subgradient evaluation approaches for state relaxations obtained using the Scott–Barton framework, which show for the first time that dynamic subgradients may be computed using adjoint sensitivity approaches [90]. These approaches may improve computational efficiency for lower bounding in deterministic global dynamic optimization. A new Julia implementation of deterministic global dynamic optimization is being developed in collaboration with colleagues. The work in this thesis has appeared in the published articles [3, 4], articles currently in review [91, 92], and a manuscript in preparation [93]. The contents and contributions of each chapter of this thesis are summarized below.

Chapter 2 presents the notational conventions throughout this thesis, formalizes the original parametric ODE system (1.1.2), and summarizes the Scott–Barton ODE relaxation framework [2] and the SBM relaxations, which are the actual state relaxations proposed in [2]. Recall that the main goal of this thesis is to resolve the limitations of the Scott–Barton framework described in Section 1.3, to improve efficiency of deterministic global dynamic optimization.

Chapter 3, reproduced from the submitted journal article [91], presents new results for comparing solutions of related ODE systems, which are more broadly applicable to engineering models than previous results. By applying these results, it is shown for the first time that in the Scott–Barton ODE relaxation framework, tighter enclosures of ODE right-hand side functions will necessarily translate into enclosures of the ODE solutions that are at least as tight.

Chapter 4, reproduced from the published journal article [3], proposes a new approach for constructing useful convex and concave relaxations for the solutions of parametric ODE systems. This new approach allows using any convex and concave relaxations for the underlying ODE system's right-hand sides, and is guaranteed to yield ODE relaxations that are at least as tight as the state-of-the-art SBM relaxations [2]. In a global optimization case study, the new ODE relaxations indeed lead to fewer branch-and-bound global optimization iterations than the SBM relaxations.

Chapter 5, reproduced from the manuscript in preparation [93], describes two extensions of the new ODE relaxations proposed in Chapter 4. Firstly, a new implementation method of the new ODE relaxations is presented, which is more efficient than the implementation used in Chapter 4. By employing this new implementation, the new ODE relaxations may be significantly tighter, and at the same time as efficient as the SBM relaxations. Secondly, another new ODE relaxation method is proposed based on Chapter 4, which can effectively exploit the structure of the underlying ODE system's right-hand sides. It will be shown that the ODE relaxations in Chapter 5 are guaranteed to be at least as tight as both the ODE relaxations in Chapter 4 and the SBM relaxations.

Chapter 6, reproduced from the submitted journal article [92], proposes new methods for evaluating subgradients of ODE relaxations obtained using the Scott–Barton framework. These methods for the first time enable using these state-of-the-art ODE relaxations in deterministic global dynamic optimization, to compute the required lower bounds of the globally optimal objective values. Moreover, this work extends the efficient classical dynamic adjoint gradient evaluation methods to nonsmooth dynamic subgradient evaluation, which may speed up the computation of lower bounds by minimizing convex relaxations.

Chapter 7, reproduced from the published journal article [4], proposes a new approach for tractably constructing useful, correct affine underestimators and lower bounds of convex relaxations via black-box sampling. This approach enables computing lower bounds in global optimization using convex relaxations whose subgradients are unavailable such as certain ODE relaxations. The resulting affine underestimators are shown to converge rapidly to an original nonconvex function as the domain of interest shrinks, and variants are proposed to account for numerical error or noise in the sampling procedure. Thus, this approach essentially extends derivative-free techniques to the computation of lower bounds in global optimization. The associated article [4] was written in collaboration with colleagues, but Chapter 7 only presents the contributions of the author of this thesis.

Chapter 2

Mathematical Preliminaries

This chapter summarizes mathematical background information that is used throughout this thesis, including notational conventions, established convex relaxation methods for composite functions, a formalized underlying ODE system (1.1.2), and the established ODE relaxations by Scott and Barton [2]. Mathematical background that is specific to one chapter will be introduced later in that chapter.

2.1 Notation

Throughout this thesis, scalars are denoted as lowercase letters (e.g. $\xi \in \mathbb{R}$), vectors are denoted as boldface lowercase letters (e.g. $\boldsymbol{\xi} \in \mathbb{R}^n$), and the i^{th} component of a vector $\boldsymbol{\xi}$ is denoted as ξ_i. The symbol $\mathbf{e}^{(i)} \in \mathbb{R}^n$ denotes the i^{th} unit coordinate vector in $\mathbb{R}^n$. Inequalities involving vectors are to be interpreted componentwise. Sets are denoted as uppercase letters (e.g. $\Xi \subseteq \mathbb{R}^n$). 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 $\mathbf{M}$ is denoted as $\mathbf{m}_{(i)}$. Let $\mathbb{R}^0$ denote a set of null vector, and for any $\mathbf{v} \in \mathbb{R}^n$, let $\mathbf{v}_{r:s}$ denote the vector $(v_r, v_{r+1}, \ldots, v_s)$ for $s > r$, denote the scalar v_r for $s = r$, and denote a null vector for $s < r$. Lower case Greek letters (e.g. $\boldsymbol{\xi}$) typically denote dummy variables standing in for analogous English-lettered quantities (e.g. $\mathbf{x}$). A dot above a quantity (e.g. $\dot{\mathbf{x}}$) indicates a partial derivative with respect to t (e.g. $\frac{\partial \mathbf{x}}{\partial t}$). For any $\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}} \in \mathbb{R}^n$ with $\mathbf{x}^{\mathrm{L}} \leq \mathbf{x}^{\mathrm{U}}$, an *interval* $X := [\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}]$ is defined as the compact set $\{\boldsymbol{\xi} \in \mathbb{R}^n : \mathbf{x}^{\mathrm{L}} \leq \boldsymbol{\xi} \leq \mathbf{x}^{\mathrm{U}}\}$. For any $Q \subseteq \mathbb{R}^n$, let $\mathbb{I}Q$ denote the set of all nonempty interval subsets of Q. Superscripts "L" and "U" will be used to denote lower and upper bounds of intervals, and superscripts "cv" and "cc" will be used to denote convex and concave relaxations. The Euclidean norm $\|\cdot\|$ or the l-infinity norm $\|\cdot\|_\infty$ and inner product $\langle \cdot, \cdot \rangle$ are considered on $\mathbb{R}^n$. The matrix norm induced by the Euclidean norm is employed. A sum such as $\sum_{j \in J} g(j)$ is understood to be 0 if the index set J is empty. The abbreviation "a.e." stands for "almost every" in the sense of Lebesgue measure.

2.2 Established relaxations for composite functions

This section illustrates several established *convex relaxation* methods for composite functions via examples, including natural interval extension [48], McCormick relaxation [5], generalized McCormick relaxation [2], αBB relaxation [9, 49], and the Auxiliary Variable Method [8, 50]. Full mathematical descriptions of these methods are somewhat cumbersome, and are available at the cited references. These methods may be used for constructing relaxations for the right-hand side function $\mathbf{f}$ in (1.1.2). Prior to this thesis, the Scott–Barton ODE relaxation framework only allowed using generalized McCormick relaxations of $\mathbf{f}$ for constructing ODE relaxations. On the other hand, the new ODE relaxation formulations that will be presented in Chapters 4 and 5 permit all these different relaxations of $\mathbf{f}$ to be embedded.

Definition 2.2.1 (adapted from [35])**.** Let $Y \subseteq \mathbb{R}^n$ be convex, and consider a function $\mathbf{h} : Y \to \mathbb{R}^m$. Vector functions $\mathbf{h}^{\mathrm{cv}}, \mathbf{h}^{\mathrm{cc}} : Y \to \mathbb{R}^m$ are respectively called to be *convex* and *concave*, if for each $i \in \{1, ..., m\}$, h_i^{cv} and h_i^{cc} are respectively convex and concave. Moreover, if for each $\mathbf{y} \in Y$, $\mathbf{h}^{\mathrm{cv}}(\mathbf{y}) \leq \mathbf{h}(\mathbf{y})$ and $\mathbf{h}^{\mathrm{cc}}(\mathbf{y}) \geq \mathbf{h}(\mathbf{y})$, then $\mathbf{h}^{\mathrm{cv}}$ and $\mathbf{h}^{\mathrm{cc}}$ are respectively called a *convex relaxation* and a *concave relaxation* of $\mathbf{h}$ on Y.

2.2.1 Natural interval extensions and McCormick relaxations

As introduced in Section 1.3, the natural interval extension [48] and the McCormick relaxation method [5] compute closed-form interval bounds and convex and concave relaxations for composite functions by propagating bounds and relaxations for simple intrinsic functions. This will be illustrated in the following example.

Example 2.1. Consider a function $h : (y,z) \mapsto e^z - (y+z)^2$ defined on the interval $[0,1]^2$. This example aims to construct lower and upper bounds $h^{\mathrm{L}}, h^{\mathrm{U}} \in \mathbb{R}$ for h for which $h^{\mathrm{L}} \leq h(y,z) \leq h^{\mathrm{U}}$, for all $(y,,z) \in [0,1]^2$ using natural interval extension, and construct convex and concave relaxations $h^{\mathrm{cv}}, h^{\mathrm{cc}} : [0,1]^2 \to \mathbb{R}$ for h using the McCormick relaxation method. These methods may employ the following evaluation procedure (also known as a *factored representation*) for h: for each $(y,z) \in [0,1]^2$,

$$\begin{aligned} v_1 &:= e^z, \\ v_2 &:= y+z, \\ v_3 &:= v_2^2, \\ v_4 &:= -v_3, \\ v_5 &:= v_1 + v_4, \\ \text{and} \quad h(y,z) &\equiv v_5. \end{aligned} \tag{2.2.1}$$

Observe that this evaluation procedure employs the following intrinsic functions:

$$w_1 : r \mapsto e^r, \quad w_2 : (r^{\mathrm{A}}, r^{\mathrm{B}}) \mapsto r^{\mathrm{A}} + r^{\mathrm{B}}, \quad w_3 : r \mapsto r^2, \quad \text{and} \quad w_4 : r \mapsto -r. \tag{2.2.2}$$

Based on the known behavior of these intrinsic functions, bounds and convex and concave relaxations for these functions on any interval $[r^{\mathrm{L}}, r^{\mathrm{U}}]$ can be easily constructed. For example, w_1 attains its maximum at r^{U} and attains its minimum at r^{L}. Since w_1 is convex, w_1 is a convex relaxation of itself, and a valid concave relaxation is the secant line connecting $(r^{\mathrm{L}}, w_1(r^{\mathrm{L}}))$ and $(r^{\mathrm{U}}, w_1(r^{\mathrm{U}}))$. Then, the bounds $(h^{\mathrm{L}}, h^{\mathrm{U}})$ for h may be computed based on the natural interval extension rules [48]. This computation requires computing intermediate quantities denoted as $(v_j^{\mathrm{L}}, v_j^{\mathrm{U}})$ for each v_j in (2.2.1), as follows:

$$\begin{aligned}
v_1^{\mathrm{L}} &:= 1, & v_1^{\mathrm{U}} &:= e, \\
v_2^{\mathrm{L}} &:= 0, & v_2^{\mathrm{U}} &:= 2, \\
v_3^{\mathrm{L}} &:= [v_2^{\mathrm{L}}]^2 = 0, & v_3^{\mathrm{U}} &:= [v_2^{\mathrm{U}}]^2 = 4, \\
v_4^{\mathrm{L}} &:= -v_3^{\mathrm{U}} = -4, & v_4^{\mathrm{U}} &:= -v_3^{\mathrm{L}} = 0, \\
v_5^{\mathrm{L}} &:= v_1^{\mathrm{L}} + v_4^{\mathrm{L}} = 1 + (-4) = -3, & v_5^{\mathrm{U}} &:= v_1^{\mathrm{U}} + v_4^{\mathrm{U}} = e + 0 = e, \\
h^{\mathrm{L}} &:= v_5^{\mathrm{L}} = -3, & h^{\mathrm{U}} &:= v_5^{\mathrm{U}} = e.
\end{aligned}$$

Similarly, for each $(y,z) \in [0,1]^2$, the *McCormick convex/concave relaxations* $h^{\mathrm{cv}}(y,z)$ and $h^{\mathrm{cc}}(y,z)$ may be computed based on the McCormick relaxation rules [5]. This computation requires the previously computed quantities $(v_j^{\mathrm{L}}, v_j^{\mathrm{U}})$ and also requires computing new quantities denoted as $(v_j^{\mathrm{cv}}, v_j^{\mathrm{cc}})$, as follows:

$$\begin{aligned}
v_1^{\mathrm{cv}} &:= e^z, & v_1^{\mathrm{cc}} &:= (e-1)z+1,\\
v_2^{\mathrm{cv}} &:= y+z, & v_2^{\mathrm{cc}} &:= y+z,\\
v_3^{\mathrm{cv}} &:= [v_2^{\mathrm{cv}}]^2, & v_3^{\mathrm{cc}} &:= (v_2^{\mathrm{U}}+v_2^{\mathrm{L}})(v_2^{\mathrm{cc}}-v_2^{\mathrm{L}})+[v_2^{\mathrm{L}}]^2,\\
v_4^{\mathrm{cv}} &:= -v_3^{\mathrm{cc}}, & v_4^{\mathrm{cc}} &:= -v_3^{\mathrm{cv}},\\
v_5^{\mathrm{cv}} &:= v_1^{\mathrm{cv}}+v_4^{\mathrm{cv}}, & v_5^{\mathrm{cc}} &:= v_1^{\mathrm{cc}}+v_4^{\mathrm{cc}},\\
h^{\mathrm{cv}}(y,z) &:= v_5^{\mathrm{cv}}, & h^{\mathrm{cc}}(y,z) &:= v_5^{\mathrm{cc}}.
\end{aligned} \tag{2.2.3}$$

The McCormick relaxations have been shown to be efficient in deterministic global optimization [35]. The generalized McCormick relaxation method [6] is a later variant of the McCormick relaxation method, which can better handle each composed function in a composition. The generalized McCormick relaxations have been implemented to automatically execute procedures similar to (2.2.3) for any composite functions, such as in the EAGO package [94] in Julia [95] and in the MC++ library [96]. Other variants of McCormick relaxations include a differentiable variant [97], a variant [7] to better handle multivariate intrinsic functions, and a tighter variant [34] via subgradient propagation.

2.2.2 αBB relaxations

The αBB relaxation method [9,49] applies to twice-differentiable functions, and constructs convex relaxations by adding a sufficiently large convex quadratic term to the original nonconvex function. This will be illustrated in the following example.

Example 2.2 (from [33])**.** Consider a function $h : (y,z) \mapsto z(y^2-1)$ defined on $[-4,4]^2$. The αBB relaxation method constructs a convex relaxation $h^{\mathrm{cv}} : [-4,4]^2 \to \mathbb{R}$ of h in the following form:

$$h^{\mathrm{cv}}(y,z) := h(y,z) + \alpha_1(y-4)(y+4) + \alpha_2(z-4)(z+4),$$

where $\alpha_1, \alpha_2 \geq 0$. Since $y, z \in [-4,4]$, the quadratic terms $\alpha_1(y-4)(y+4)$ and $\alpha_2(z-4)(z+4)$ above are always less or equal than zero. Thus, h^{cv} is guaranteed to underestimate h on $[-4,4]^2$. The key part for constructing h^{cv} is to find appropriate values of (α_1, α_2) so that the Hessian of h^{cv} is positive semidefinite at each $(y,z) \in [-4,4]^2$, which guarantees h^{cv} to be convex. There are various methods [9] for computing (α_1, α_2) typically based on estimating the Hessian of h. Using a nonuniform diagonal shift matrix method summarized in [9] for computing (α_1, α_2), a correct convex relaxation h^{cv} may be constructed as

$$h^{\mathrm{cv}}(y,z) := h(y,z) + 8(y^2 - 16) + 4(z^2 - 16).$$

A concave relaxation of h is constructed analogously. Note that the αBB relaxations only apply to twice-differentiable nonconvex functions, whereas the McCormick relaxations in the previous subsection apply to nonsmooth functions or even discontinuous functions [98].

2.2.3 Auxiliary Variable Method

The Auxiliary Variable Method (AVM) [8, 50] is used for constructing lower-bounding problems in global optimization, and is employed in the state-of-the-art deterministic global optimization solver BARON [28, 60]. The following example illustrates how the Auxiliary Variable Method (AVM) [8, 50] constructs an auxiliary convex optimization problem, whose optimal objective values are valid lower bounds for the optimal objective values of an original nonconvex optimization problem.

Example 2.3. Consider minimizing the function h defined in Example 2.1 in Section 2.2.1 on the box $[0,1]^2$:

$$\begin{aligned} \min_{y,z} \quad & e^z - (y+z)^2 \\ \text{s.t.} \quad & 0 \leq y \leq 1, \\ & 0 \leq z \leq 1. \end{aligned} \tag{2.2.4}$$

Observe that this problem is nonconvex since the objective function is nonconvex. The AVM considers the evaluation procedure (2.2.1) of h and sets up new decision variables $(v_1, v_2, v_3, v_4, v_5)$ in an auxiliary optimization problem. The decision variables v_1 and v_3 are bounded by convex and concave relaxations of the nonlinear intrinsic functions w_1 and w_3 in (2.2.2), respectively, and the linear expressions $v_2 = y+z$, $v_4 = -v_3$, and $v_5 = v_1 + v_4$ in (2.2.1) are employed directly as linear constraints. Thus, the auxiliary optimization problem constructed by AVM is as follows:

$$\begin{aligned} \min_{y,z,v_1,v_2,v_3,v_4,v_5} \quad & v_5 \\ \text{s.t.} \quad & v_5 = v_1 + v_4, \\ & v_4 = -v_3, \\ & v_2^2 \leq v_3 \leq (v_2^{\mathrm{U}} + v_2^{\mathrm{L}})(v_2 - v_2^{\mathrm{L}}) + [v_2^{\mathrm{L}}]^2, \\ & v_2 = y+z, \\ & e^z \leq v_1 \leq (e-1)z + 1, \\ & 0 \leq y \leq 1, \\ & 0 \leq z \leq 1. \end{aligned} \tag{2.2.5}$$

As established in [8, 50], the optimization problem above is convex, and its optimal objective values are valid lower bounds for the optimal objective values of (2.2.4). If we substitute the “min” with “max” in (2.2.5), then this optimization problem is now a concave maximization problem, and instead yields guaranteed upper bounds for (2.2.4).

Chapter 5 in this thesis will extend the AVM to construct convex and concave relaxations for composite functions, and will show that the AVM relaxations are guaranteed to be at least as tight as the McCormick relaxations.

2.3 Underlying ODE system

This section is adapted from [3, Section 2] and formalizes the ODE process model (1.1.2) for a generic dynamic process system considered throughout. This thesis considers deterministic global dynamic optimization problems with this system embedded, and proposes new approaches for computing convex relaxations with corresponding subgradients for the ODE solution, for efficiently computing lower-bounding information in an overarching global optimization method. The following definition of uniform Lipschitz continuity is adapted from [99].

Definition 2.3.1 (adapted from [99]). Consider a function $\mathbf{h} : Y \times Z \to \mathbb{R}^m$. The mapping $\mathbf{h}(\cdot, \mathbf{z})$ is said to be *Lipschitz continuous* on Y, *uniformly over* $\mathbf{z} \in Z$ if there exists a $l \geq 0$ so that for any $\mathbf{y}^{\mathrm{A}}, \mathbf{y}^{\mathrm{B}} \in Y$ and $\mathbf{z} \in Z$,

$$\|\mathbf{h}(\mathbf{y}^{\mathrm{A}}, \mathbf{z}) - \mathbf{h}(\mathbf{y}^{\mathrm{B}}, \mathbf{z})\|_\infty \leq l \|\mathbf{y}^{\mathrm{A}} - \mathbf{y}^{\mathrm{B}}\|_\infty.$$

The following assumption formalizes the parametric ODE system (1.1.2).

Assumption 2.3.2. Let $I := [t_0, t_f] \subsetneq \mathbb{R}$ and $P := [\mathbf{p}^{\mathrm{L}}, \mathbf{p}^{\mathrm{U}}] \subsetneq \mathbb{R}^{n_p}$ be nonempty intervals, and let $D \subseteq \mathbb{R}^{n_x}$ be open. Consider continuous functions $\mathbf{x}_0 : P \to D$ and $\mathbf{f} : I \times P \times D \to \mathbb{R}^{n_x}$. Suppose that $\mathbf{f}(t, \mathbf{p}, \cdot)$ is Lipschitz continuous on D, uniformly over $(t, \mathbf{p}) \in I \times P$. Consider the following parametric ODE system:

$$\begin{aligned} \dot{\mathbf{x}}(t,\mathbf{p}) &= \mathbf{f}(t,\mathbf{p},\mathbf{x}(t,\mathbf{p})), \quad \forall t \in (t_0, t_f], \\ \mathbf{x}(t_0,\mathbf{p}) &= \mathbf{x}_0(\mathbf{p}). \end{aligned} \tag{2.3.1}$$

Suppose that the ODE system (2.3.1) has exactly one solution in the classical sense on I.

Definition 2.3.3. Consider the parametric ODE system (2.3.1) formalized in Assumption 2.3.2. A function $\mathbf{x} : I \times P \to D$ is a solution *in the classical sense* of (2.3.1) on I if, for each $\mathbf{p} \in P$, $\mathbf{x}(\cdot, \mathbf{p})$ is continuously differentiable and satisfies (2.3.1) on I. A function $\mathbf{x} : I \times P \to D$ is a solution *in the Carathéodory sense* of (2.3.1) on I if, for each $\mathbf{p} \in P$, $\mathbf{x}(\cdot, \mathbf{p})$ is absolutely continuous on I, $\mathbf{x}(t_0, \mathbf{p}) = \mathbf{x}_0(\mathbf{p})$ is satisfied, and $\dot{\mathbf{x}}(t, \mathbf{p}) = \mathbf{f}(t, \mathbf{p}, \mathbf{x}(t, \mathbf{p}))$ is satisfied for a.e. $t \in I$. Solutions of other ODEs throughout this thesis are defined analogously.

Note that if an appropriate Lipschitz extension of $\mathbf{f}$ is applied from the domain $I \times P \times D$ to $I \times P \times \mathbb{R}^{n_x}$, then global existence and uniqueness of solutions of (2.3.1) on I are guaranteed by [100, Theorem 3.2].

2.4 Established ODE relaxation formulations

For reference, this section, adapted from [3, Section 3], summarizes established ODE relaxation formulations [2, 69] for the underlying ODE system (2.3.1). Firstly, Harrison's method [69] for constructing *state bounds* that are $\mathbf{p}$-invariant bounds for the solution trajectory $\mathbf{x}$ of (2.3.1) is summarized. This method is widely used in dynamic state bounding. Then, this section summarizes a general framework for constructing *state relaxations* for $\mathbf{x}$ in (2.3.1) by Scott and Barton [2], who then specialize this framework by applying Harrison's bounding method and the generalized McCormick relaxation method [6] to furnish crucial functions in the framework. Prior to this thesis, this was the only method for furnishing these functions, whereas this thesis proposes new ODE relaxation approaches in this framework, which can yield state relaxations that are at least as tight as the relaxations proposed by Scott and Barton [2].

Definition 2.4.1 (adapted from [2])**.** Functions $\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}} : I \times P \to \mathbb{R}^{n_x}$ are called *state relaxations* for (2.3.1) on $I \times P$ if, for every $t \in I$, $\mathbf{x}^{\mathrm{cv}}(t, \cdot)$ is a convex relaxation of $\mathbf{x}(t, \cdot)$ on P, and $\mathbf{x}^{\mathrm{cc}}(t, \cdot)$ is a concave relaxation of $\mathbf{x}(t, \cdot)$ on P.

Definition 2.4.2 (adapted from [2])**.** Functions $\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}} : I \to \mathbb{R}^{n_x}$ are called *state bounds* for (2.3.1) on $I \times P$ if $\mathbf{x}^{\mathrm{L}}(t) \leq \mathbf{x}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{U}}(t)$ for all $(t, \mathbf{p}) \in I \times P$. For each $t \in I$, denote the interval $[\mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{U}}(t)] \subsetneq \mathbb{R}^{n_x}$ as $X(t)$.

Definition 2.4.4 below formalizes a state bounding method by Harrison [69], which is widely used in dynamic state bounding [1, 2, 13, 74–76, 78]. This method describes state bounds as the unique solution of an auxiliary ODE system whose right-hand side involves the *natural interval extension* [48] of the original right-hand side $\mathbf{f}$ in (2.3.1).

Definition 2.4.3 (adapted from [2])**.** For each $i \in \{1, ..., n_x\}$, define *interval flattening functions* $\mathbf{r}^{i,\mathrm{L}}, \mathbf{r}^{i,\mathrm{U}} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ for which

1. $\mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\xi}^{\mathrm{L,A}}, \boldsymbol{\xi}^{\mathrm{U,A}}) := (\boldsymbol{\xi}^{\mathrm{L,A}}, \boldsymbol{\xi}^{\mathrm{U,B}})$, where $\xi_k^{\mathrm{U,B}} := \xi_k^{\mathrm{U,A}}$ for all $k \in \{1, ..., n_x\}$ and $k \neq i$, and $\xi_i^{\mathrm{U,B}} := \xi_i^{\mathrm{L,A}}$,

2. $\mathbf{r}^{i,\mathrm{U}}(\boldsymbol{\xi}^{\mathrm{L,A}}, \boldsymbol{\xi}^{\mathrm{U,A}}) := (\boldsymbol{\xi}^{\mathrm{L,B}}, \boldsymbol{\xi}^{\mathrm{U,A}})$, where $\xi_k^{\mathrm{L,B}} := \xi_k^{\mathrm{L,A}}$ for all $k \in \{1,...,n_x\}$ and $k \neq i$, and $\xi_i^{\mathrm{L,B}} := \xi_i^{\mathrm{U,A}}$.

Definition 2.4.4 (adapted from [69])**.** Consider the parametric ODE system (2.3.1) formalized in Assumption 2.3.2. Let $D \subseteq \bar{D} \subseteq \mathbb{R}^{n_x}$. Consider constant lower and upper bounds $\mathbf{x}_0^{\mathrm{L}}, \mathbf{x}_0^{\mathrm{U}} \in \bar{D}$ for the function $\mathbf{x}_0$ on P. Consider the natural interval extension [48] $\mathbf{f}^{\mathrm{L}}, \mathbf{f}^{\mathrm{U}} : I \times \bar{D} \times \bar{D} \to \mathbb{R}^{n_x}$ of $\mathbf{f}$ such that, for each $t \in I$ and $[\boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}}] \in \mathbb{I}\bar{D}$, $\mathbf{f}^{\mathrm{L}}(t, \boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}})$ and $\mathbf{f}^{\mathrm{U}}(t, \boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}})$ are respectively constant lower and upper bounds for $\mathbf{f}(t,\cdot,\cdot)$ on $P \times [\boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}}]$. Then, *Harrison's state bounding method* constructs the following auxiliary ODE system: for each $i \in \{1,...,n_x\}$,

$$\begin{aligned} \dot{x}_i^{\mathrm{L}}(t) &= f_i^{\mathrm{L}}\big(t, \mathbf{r}^{i,\mathrm{L}}(\mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{U}}(t))\big), \quad x_i^{\mathrm{L}}(t_0) = x_{0,i}^{\mathrm{L}}, \\ \dot{x}_i^{\mathrm{U}}(t) &= f_i^{\mathrm{U}}\big(t, \mathbf{r}^{i,\mathrm{U}}(\mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{U}}(t))\big), \quad x_i^{\mathrm{U}}(t_0) = x_{0,i}^{\mathrm{U}}. \end{aligned}$$

Let $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ be a solution of this ODE system in the classical sense on I. Then, $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ are valid state bounds for the underlying parametric ODE system (2.3.1).

Assumption 2.4.5. Throughout this thesis, suppose that the state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ are *LR-analytic* (see [101, Definition 2.3]) on I. Roughly, a LR-analytic function can be represented by the combination of finitely many pieces of sufficiently smooth functions.

Observe that Harrison's state bounds [69] formalized in Definition 2.4.4 are LR-analytic. Since the right-hand side of Harrison's auxiliary ODE system is *abs-factorable* as in [101, Definition 2.1], the resulting state bounds are LR-analytic according to [101, Theorem 3.6]. Note that the state bounds are assumed to be absolutely continuous in [2], so that the time derivatives $(\dot{\mathbf{x}}^{\mathrm{L}}, \dot{\mathbf{x}}^{\mathrm{U}})$ are well-defined for a.e. $t \in I$. The LR-analytic assumption here is a stronger, yet widely applicable assumption, which is essential for guaranteeing solutions' existence of the Scott–Barton ODE system (2.4.1) below by classical results [102, Chapter 2, §7].

Definition 2.4.6. Functions $\mathbf{x}_0^{\text{cv}}, \mathbf{x}_0^{\text{cc}} : P \to \mathbb{R}_0^{n_x}$ are called *initial relaxations* for (2.3.1) if $\mathbf{x}_0^{\text{cv}}$ and $\mathbf{x}_0^{\text{cc}}$ are respectively convex and concave relaxations for $\mathbf{x}_0$ in (2.3.1) on P.

Definition 2.4.7 (from [2])**.** Functions $\mathbf{u}, \mathbf{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ describe *bound preserving dynamics for (2.3.1) (based on bounds $X(t)$)* if, for any $i \in \{1, ..., n_x\}$, any $\mathbf{p} \in P$, a.e. $t \in I$, and any functions $\boldsymbol{\xi}, \boldsymbol{\xi}^{\text{cv}}, \boldsymbol{\xi}^{\text{cc}} : I \times P \to \mathbb{R}^{n_x}$ such that $\boldsymbol{\xi}(t,\mathbf{p}), \boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}), \boldsymbol{\xi}^{\text{cc}}(t,\mathbf{p}) \in X(t)$ and $\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}) \leq \boldsymbol{\xi}(t,\mathbf{p}) \leq \boldsymbol{\xi}^{\text{cc}}(t,\mathbf{p})$, the following holds:

1. if $\xi_i(t,\mathbf{p}) = \xi_i^{\text{cv}}(t,\mathbf{p})$, then $u_i(t,\mathbf{p},\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}),\boldsymbol{\xi}^{\text{cc}}(t,\mathbf{p})) \leq f_i(t,\mathbf{p},\boldsymbol{\xi}(t,\mathbf{p}))$, and
2. if $\xi_i(t,\mathbf{p}) = \xi_i^{\text{cc}}(t,\mathbf{p})$, then $o_i(t,\mathbf{p},\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}),\boldsymbol{\xi}^{\text{cc}}(t,\mathbf{p})) \geq f_i(t,\mathbf{p},\boldsymbol{\xi}(t,\mathbf{p}))$.

Definition 2.4.8 (from [2])**.** Functions $\mathbf{u}, \mathbf{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ describe *convexity preserving dynamics for (2.3.1) (based on bounds $X(t)$)* if, for any $i \in \{1, ..., n_x\}$, any $\lambda \in (0,1)$, any $\mathbf{p}^{\text{A}}, \mathbf{p}^{\text{B}} \in P$, $\bar{\mathbf{p}} := \lambda \mathbf{p}^{\text{A}} + (1-\lambda)\mathbf{p}^{\text{B}}$, a.e. $t \in I$, and any functions $\boldsymbol{\xi}^{\text{cv}}, \boldsymbol{\xi}^{\text{cc}} : I \times P \to \mathbb{R}^{n_x}$ such that

1. $\boldsymbol{\xi}^{\text{cv}}(t,\bar{\mathbf{p}}) \leq \lambda \boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}^{\text{A}}) + (1-\lambda)\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}^{\text{B}})$,
2. $\boldsymbol{\xi}^{\text{cc}}(t,\bar{\mathbf{p}}) \geq \lambda \boldsymbol{\xi}^{\text{cc}}(t,\mathbf{p}^{\text{A}}) + (1-\lambda)\boldsymbol{\xi}^{\text{cc}}(t,\mathbf{p}^{\text{B}})$, and
3. $\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{q}) \leq \boldsymbol{\xi}^{\text{cc}}(t,\mathbf{q})$ and $\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{q}), \boldsymbol{\xi}^{\text{cc}}(t,\mathbf{q}) \in X(t)$, for all $\mathbf{q} \in \{\mathbf{p}^{\text{A}}, \mathbf{p}^{\text{B}}, \bar{\mathbf{p}}\}$,

both of the following conditions hold:

1. if $\xi_i^{\text{cv}}(t,\bar{\mathbf{p}}) = \lambda \xi_i^{\text{cv}}(t,\mathbf{p}^{\text{A}}) + (1-\lambda)\xi_i^{\text{cv}}(t,\mathbf{p}^{\text{B}})$, then

$$\begin{aligned} u_i(t,\bar{\mathbf{p}},\boldsymbol{\xi}^{\text{cv}}(t,\bar{\mathbf{p}}),\boldsymbol{\xi}^{\text{cc}}(t,\bar{\mathbf{p}})) \leq{} & \lambda u_i(t,\mathbf{p}^{\text{A}},\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}^{\text{A}}),\boldsymbol{\xi}^{\text{cc}}(t,\mathbf{p}^{\text{A}})) \\ & + (1-\lambda)u_i(t,\mathbf{p}^{\text{B}},\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}^{\text{B}}),\boldsymbol{\xi}^{\text{cc}}(t,\mathbf{p}^{\text{B}})), \text{ and} \end{aligned}$$

2. if $\xi_i^{\mathrm{cc}}(t,\bar{\mathbf{p}}) = \lambda \xi_i^{\mathrm{cc}}(t,\mathbf{p}^{\mathrm{A}}) + (1-\lambda)\xi_i^{\mathrm{cc}}(t,\mathbf{p}^{\mathrm{B}})$, then

$$\begin{aligned} o_i(t,\bar{\mathbf{p}},\boldsymbol{\xi}^{\mathrm{cv}}(t,\bar{\mathbf{p}}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\bar{\mathbf{p}})) \geq{}& \lambda o_i(t,\mathbf{p}^{\mathrm{A}},\boldsymbol{\xi}^{\mathrm{cv}}(t,\mathbf{p}^{\mathrm{A}}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\mathbf{p}^{\mathrm{A}})) \\ &+ (1-\lambda) o_i(t,\mathbf{p}^{\mathrm{B}},\boldsymbol{\xi}^{\mathrm{cv}}(t,\mathbf{p}^{\mathrm{B}}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\mathbf{p}^{\mathrm{B}})). \end{aligned}$$

Definition 2.4.9 (adapted from [2])**.** Functions $\mathbf{u},\mathbf{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ are called *Scott–Barton right-hand side functions for (2.3.1) (based on bounds $X(t)$)* if,

1. $\mathbf{u}$ and $\mathbf{o}$ are continuous,

2. $\mathbf{u}(t,\mathbf{p},\cdot,\cdot)$ and $\mathbf{o}(t,\mathbf{p},\cdot,\cdot)$ are Lipschitz continuous on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly over $(t,\mathbf{p}) \in I \times P$, and

3. $\mathbf{u}$ and $\mathbf{o}$ describe both bound preserving dynamics and convexity preserving dynamics for (2.3.1) based on bounds $X(t)$.

Now, the following definition formalizes the Scott–Barton ODE relaxation framework [2].

Definition 2.4.10 (from [2])**.** Given state bounds $(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}})$ that satisfy Assumption 2.4.5, initial relaxations $(\mathbf{x}_0^{\mathrm{cv}},\mathbf{x}_0^{\mathrm{cc}})$, and valid Scott–Barton right-hand side functions $(\mathbf{u},\mathbf{o})$ in Definition 2.4.9, Scott and Barton provide valid state relaxations $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$ for (2.3.1) as the unique solution in the Carathéodory sense of an auxiliary ODE system constructed as follows: for each $i \in \{1,...,n_x\}$,

$$\dot{x}_i^{\mathrm{cv}}(t,\mathbf{p}) = \begin{cases} u_i(t,\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})), & \text{if } x_i^{\mathrm{cv}}(t,\mathbf{p}) > x_i^{\mathrm{L}}(t), \\ \max\left(\dot{x}_i^{\mathrm{L}}(t), u_i(t,\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}))\right), & \text{if } x_i^{\mathrm{cv}}(t,\mathbf{p}) \leq x_i^{\mathrm{L}}(t), \end{cases}$$
$$\dot{x}_i^{\mathrm{cc}}(t,\mathbf{p}) = \begin{cases} o_i(t,\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})), & \text{if } x_i^{\mathrm{cc}}(t,\mathbf{p}) < x_i^{\mathrm{U}}(t), \\ \min\left(\dot{x}_i^{\mathrm{U}}(t), o_i(t,\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}))\right), & \text{if } x_i^{\mathrm{cc}}(t,\mathbf{p}) \geq x_i^{\mathrm{U}}(t), \end{cases} \tag{2.4.1}$$
$$x_i^{\mathrm{cv}}(t_0,\mathbf{p}) = \max(x_i^{\mathrm{L}}(t_0), x_{0,i}^{\mathrm{cv}}(\mathbf{p})), \quad x_i^{\mathrm{cc}}(t_0,\mathbf{p}) = \min(x_i^{\mathrm{U}}(t_0), x_{0,i}^{\mathrm{cc}}(\mathbf{p})).$$

The auxiliary parametric ODE system (2.4.1) with any constructions of the functions $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}, \mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}}, \mathbf{u}, \mathbf{o})$ will be denoted as the *Scott–Barton ODE relaxation framework.*

Scott–Barton right-hand side functions $(\mathbf{u},\mathbf{o})$ may be constructed by first constructing functions $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ that satisfy the following assumption [2].

Assumption 2.4.11 (adapted from [2])**.** Suppose that functions $\tilde{\mathbf{u}},\tilde{\mathbf{o}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ satisfy all of the following conditions:

1. $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{o}}$ are continuous,

2. $\tilde{\mathbf{u}}(t,\mathbf{p},\cdot,\cdot)$ and $\tilde{\mathbf{o}}(t,\mathbf{p},\cdot,\cdot)$ are Lipschitz continuous on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly over $(t,\mathbf{p}) \in I \times P$, and

3. $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{o}}$ describe *bound amplifying dynamics* and *convexity amplifying dynamics* for (2.3.1) based on bounds $X(t)$, as defined in [2].

Roughly, the functions $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{o}}$ are respectively intended to describe convex and concave relaxations of the composition $\mathbf{f}(t,\cdot,\mathbf{x}(t,\cdot))$ on P, for each fixed $t \in I$.

Scott and Barton proceed by constructing functions $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ satisfying Assumption 2.4.11 using the generalized McCormick relaxation method [6]. Scott–Barton right-hand side functions $(\mathbf{u}, \mathbf{o})$ are then constructed by composing the functions $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ with the interval flattening operations in Definition 2.4.3 as follows: for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} u_i(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) &\equiv \tilde{u}_i(t, \mathbf{p}, \mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})) \\ \text{and} \quad o_i(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) &\equiv \tilde{o}_i(t, \mathbf{p}, \mathbf{r}^{i,\mathrm{U}}(\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})). \end{aligned} \tag{2.4.2}$$

Within the Scott–Barton framework, we refer to the unique solution $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ of (2.4.1) as *Scott–Barton–McCormick (SBM) relaxations* for (2.3.1) if the functions $(\mathbf{u}, \mathbf{o})$ in (2.4.1) are defined by (2.4.2) where $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ are the generalized McCormick relaxations [6] of $\mathbf{f}$. Prior to this thesis, SBM relaxations were the only actual state relaxations to be established within the Scott–Barton framework, since generalized McCormick relaxations were the only established way to generate Scott–Barton right-hand sides.

Intuitively, in order to be useful in global dynamic optimization algorithm, the state relaxations $\mathbf{x}^{\mathrm{cv}}$ and $\mathbf{x}^{\mathrm{cc}}$ must converge to $\mathbf{x}$ as P is subdivided. To achieve this, the functions $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{o}}$ in Assumption 2.4.11 are also expected to satisfy the following assumption.

Assumption 2.4.12 (from [79])**.** For a.e. $t \in I$, any $\mathbf{p} \in P$, and any $\boldsymbol{\xi}^{\mathrm{cv,A}}, \boldsymbol{\xi}^{\mathrm{cc,A}}, \boldsymbol{\xi}^{\mathrm{cv,B}}, \boldsymbol{\xi}^{\mathrm{cc,B}} \in X(t)$ such that $\boldsymbol{\xi}^{\mathrm{cv,A}} \leq \boldsymbol{\xi}^{\mathrm{cv,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,A}}$, suppose that

$$\tilde{\mathbf{u}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv,A}}, \boldsymbol{\xi}^{\mathrm{cc,A}}) \leq \tilde{\mathbf{u}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv,B}}, \boldsymbol{\xi}^{\mathrm{cc,B}}) \leq \tilde{\mathbf{o}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv,B}}, \boldsymbol{\xi}^{\mathrm{cc,B}}) \leq \tilde{\mathbf{o}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv,A}}, \boldsymbol{\xi}^{\mathrm{cc,A}}).$$

Note that $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{o}}$ constructed as generalized McCormick relaxations are guaranteed to satisfy Assumption 2.4.12 [6, 79].

Chapter 3

Comparing Solutions of Related ODEs Using New Differential Inequalities

This chapter, reproduced from the submitted journal article [91], presents new results for comparing the Carathéodory solutions of related ODE systems, under less stringent conditions than established results. A first result provides sufficient conditions under which one ODE system's solutions dominate another's. Unlike certain established results, this result does not require differentiability of ODE solutions or a quasi-monotonicity assumption [88, Definition 1.5.2] on right-hand side functions. A second result addresses ODEs that describe bounding trajectories for an original ODE system, and provides sufficient conditions under which one system of bounding trajectories is guaranteed to enclose another. By applying this result, it is shown for the first time that the Scott–Barton framework [2] (as summarized in Section 2.4) for constructing useful convex enclosures for the reachable sets of a parametric ODE system has the following property: if tighter enclosures of the original ODE right-hand side function are available, then these will translate into enclosures of the ODE solution that are at least as tight.

3.1 Introduction

This chapter focuses on inequalities between solutions of related ODE systems, based on differential inequalities and inequalities between the system's initial conditions. Such *ODE comparison results* are crucial in analysis methods for ODEs, such as approaches for investigating ODE solutions' uniqueness, stability, and continuous dependence on initial values [77, 87, 88, 103]. Such ODE comparison results also provide theoretical justification for reachable-set generation methods [69, 72, 74, 104]. Reachable-set methods are widely used in engineering applications such as fault detection [105], robust optimal control [106], safety analysis [20], and state and parameter estimation [104, 107, 108]. Recently, comparison results have also been used to compare competing reachable-set generation methods in terms of tightness [79]. Our primary motivation concerns algorithms for deterministic global dynamic optimization [19, 27], which typically construct convex enclosures of ODE reachable sets to obtain crucial bounding information. In this context, ODE comparison results may help with developing reachable-set enclosure methods that construct less conservative enclosures, which would intuitively reduce the number of iterations required in global dynamic optimization algorithms.

This work presents comparison results concerning ODE systems known to have *Carathéodory solutions* [77] (C-solutions). These solutions' derivatives are known to exist and are uniquely determined almost everywhere in the sense of Lebesgue measure. Such ODE systems are prevalent in applications such as hybrid dynamic systems [109, 110], Lebesgue-integrable control systems [12, 111], and nonsmooth ODE relaxation systems [2, 75]. Several established ODE comparison results [39,77,79,87–89] are concerned with bounding C-solutions in particular. We broadly group these results into three classes:

1. *one-sided bounding results* that bound C-solutions from either below or above,

2. *two-sided bounding results* that bound C-solutions from both below and above simultaneously, and

3. *bound comparison results* that consider two two-sided bounding pairs and provide sufficient conditions under which one bounding pair is guaranteed to enclose another.

Table 3.1 summarizes these established results, along with summaries of their assumptions enforced on the original ODE right-hand sides and solutions. As shown in Table 3.1, assumptions individually shared by several of these results include: whether the ODE solution is scalar-valued, whether the ODE solution is differentiable everywhere, whether the considered solution is the maximal solution among possible non-unique solutions, whether the ODE right-hand side (RHS) function has properties such as quasimonotonicity [88, Definition 1.5.2], satisfaction of a right-uniqueness condition [77, Equation 10.3], or a weakened variant [77, Condition 6.VII(γ)], and Lipschitz continuity [77, Example 11.IV]. The results also differ based on whether strict (with $<$ signs) or weak (with $\leq$ signs) inequalities between bounds and ODE solutions are obtained. As shown in Table 3.1, established one-sided bounding results for ODE systems require either the ODE solution to be differentiable, or require the RHS function to be quasimonotone increasing (c.f. [88, Definition 1.5.2]). However, C-solutions are typically not differentiable everywhere, and dynamic systems in applications do not exhibit quasimonotonicity in general [19]. Schaber et al. [79] propose the main established bound comparison result [79, Theorem 4.14], which is applied to show that the reachable-set enclosure method in [2] is guaranteed to yield tighter enclosures for solutions of parametric ODE systems than the method in [75]. However, this bound comparison result requires certain stringent assumptions of the compared bounding systems. This result requires the outer bounding system's right-hand side to be locally Lipschitz continuous, requires openness of important sets, and imposes certain stringent differential inequality assumptions, as will be detailed in Section 3.4 below. Thus, this result is not easily extended to compare other reachable-set generation methods such as [2, 3, 70–72].

Table 3.1: Summary of established comparison results for bounding ODEs with C-solutions.

Classes	Results	Properties of solutions			Properties of RHS			$<$	Additional requirements
		$n_x = 1$	Diff.	Max.	QMI	RU	Lip.		
One-sided	[77, Theorem 12.III]		✓			✓			
	[77, Theorem 5.III]	✓				✓		✓	
	[88, Theorem 1.10.2]	✓		✓					
	[77, Theorem 6.VII]				✓			✓	weakened RU
	[87, Theorem 16.2]			✓	✓				
Two-sided	[89, Theorem 13]		✓				✓		
	[77, Theorem 6.XI]							✓	weakened RU
	[39, Theorem 3.3.2]						✓		
Bound comparison	[79, Theorem 4.14]								stringent outer system

Abbreviations:

"$<$": strict inequality (vs. weak inequality)
Diff.: differentiable solutions required
Max.: maximal solutions required
QMI: quasimonotone increasing (c.f. [88, Definition 1.5.2])
RU: right-uniqueness condition (c.f. [77, Equation 10.3])
Lip.: Lipschitz continuity (c.f. [77, Example 11.IV])
"✓": this property is assumed to hold
RHS: ODE right-hand side function
n_x: dimension of ODE solution

State relaxations (as described by Scott and Barton [2]) are lower and upper bounds for solutions of parametric ODE systems, whose components are respectively convex and concave with respect to the parameters. State relaxations are useful in global dynamic optimization to construct lower bounds for globally optimal objective values in deterministic branch-and-bound approaches [27, 29, 31], and can also be used to construct convex polyhedral enclosures of reachable sets via finite combinations of their subtangents [44, 47]. Tighter state relaxations will intuitively translate into tighter lower bounds supplied to branch-and-bound, which would then reduce the number of branch-and-bound iterations required, and thus may speed up an overarching global optimization method.

As summarized in Section 2.4, Scott and Barton [2] propose a general framework for generating state relaxations. This framework describes state relaxations as the C-solutions of an auxiliary ODE system, which requires furnishing certain convex and concave relaxations for the original ODE initial condition and RHS functions. The generalized McCormick relaxation method [6] is then used to construct these relaxations, and has been shown to efficiently generate useful convex and concave relaxations for composite functions [6, 35]. As in [3], we denote these ODE relaxations derived from the generalized McCormick relaxations as *Scott–Barton–McCormick (SBM) relaxations*. However, in this Scott–Barton framework, it is thus far unknown whether the tightness of RHS relaxations translates into tightness of state relaxations. Since several recent advances [7, 34, 85] have been proposed for constructing tight convex relaxations for closed-form functions, new methods for constructing tighter state relaxations could significantly benefit from a greater understanding of the tightness properties of the Scott–Barton framework. Establishing these tightness properties requires a new bound comparison result, because the only established bound comparison result [79, Theorem 4.14] in Table 3.1 requires the outer bounding system's right-hand side to be locally Lipschitz continuous, which is not satisfied by relaxation systems in the Scott–Barton framework. Moreover, the stringent differential inequality assumptions and set requirements of [79, Theorem 4.14] must be relaxed to compare the SBM relaxations [2] and the optimization-based relaxations [3] in the Scott–Barton framework that will be introduced in the next chapter, as will be discussed later.

This work proposes a new one-sided bounding result that establishes weak inequalities between bounds and C-solutions of ODE systems. Unlike established one-sided bounding results, our new result does not require the ODE solutions to be differentiable everywhere, and does not require the original RHS function to be quasimonotone increasing. Thus, this new result is applicable to a broad class of dynamic systems in applications. Based on this new one-sided bounding result, we then propose a new bound comparison result for an original ODE system with C-solutions, under significantly less stringent assumptions than the only established bound comparison result [79, Theorem 4.14]. Our new result requires only a weakened right-uniqueness condition of the bounding systems' right-hand side that is weaker than the Lipschitz continuity required by [79, Theorem 4.14]. This new result also provides more moderate differential inequality assumptions and set requirements than [79, Theorem 4.14]. A detailed comparison between these two results will be given in Section 3.4. By applying this new bound comparison result, we show that the Scott–Barton ODE relaxation framework [2] has the following tightness property: if tighter convex and concave relaxations of the original ODE initial condition and RHS functions are available, then these tighter relaxations will necessarily translate into state relaxations that are at least as tight, and may thereby provide tighter bounding information for deterministic global dynamic optimization. While plausible, this tightness property was apparently previously unknown and is revealed by our new bound comparison result. This tightness result in turn paves the way towards using tighter non-McCormick relaxations for constructing tighter state relaxations in the Scott–Barton framework [2]. Moreover, this result also indicates that it is worthwhile to seek tighter enclosure methods for closed-form functions and models from the standpoint of reachability analysis or dynamic optimization in general, since doing so necessarily translates into superior descriptions of reachable-set enclosures for dynamic systems. The next chapter will propose a new ODE state relaxation approach [3] in the Scott–Barton framework, which allows using non-McCormick relaxations of the original RHS function to construct state relaxations. As an application of Theorem 3.5.1 of this chapter, it will be shown that if McCormick relaxations [5] of the original RHS are applied, then the new ODE relaxation approach [3] is guaranteed to yield state relaxations that lie within the SBM relaxations [2]. Moreover, embedding tighter non-McCormick relaxations in the new method will necessarily lead to state relaxations that are at least as tight. Example 4.6 in the next chapter shows that such tightness of the new state relaxations does indeed translate into fewer branch-and-bound iterations, which may ultimately reduce computational effort for deterministic global dynamic optimization.

The remainder of this chapter is organized as follows. Section 3.2 formalizes Carathéodory solutions for ODE systems. Section 3.3 presents the new one-sided bounding result. Section 3.4 presents the new bound comparison result, which is then shown to supersede to an established bound comparison result [79, Theorem 4.14]. As an application of these results, Section 3.5 presents new tightness results for the Scott–Barton ODE relaxation framework [2].

3.2 Background: Carathéodory solutions of ODE systems

Definition 3.2.1 (from [102])**.** Given $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$, a *set-valued mapping* $X : A \rightrightarrows B$ is a function that maps each element of A to a subset of B.

Definition 3.2.2 (adapted from [77])**.** Let $I := [t_0, t_f] \subsetneq \mathbb{R}$. Consider a set-valued mapping $X : I \rightrightarrows \mathbb{R}^n$, define a set $U := \{(t, \boldsymbol{\phi}) \in I \times \mathbb{R}^n : \boldsymbol{\phi} \in X(t)\}$, and let $\mathbf{x}_0 \in X(t_0)$. Consider a function $\mathbf{f} : U \to \mathbb{R}^n$ and an ODE system:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}(t, \mathbf{x}(t)), \quad \forall t \in (t_0, t_f], \\ \mathbf{x}(t_0) &= \mathbf{x}_0. \end{aligned} \tag{3.2.1}$$

A function $\mathbf{x} : I \to \mathbb{R}^n$ is called a *Carathéodory solution* (C-solution) of (3.2.1) on I if the following conditions hold:

1. the initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$ is satisfied,
2. $\mathbf{x}$ is absolutely continuous on I, and
3. for a.e. $t \in I$, $\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t))$.

3.3 One-sided bounding

The following theorem provides new sufficient conditions under which one ODE system's solutions dominate another's. The subsequent results in this chapter build upon this result.

Theorem 3.3.1. Let $I := [t_0, t_f] \subsetneq \mathbb{R}$. Consider set-valued mappings $X^{\mathrm{A}}, X^{\mathrm{B}} : I \rightrightarrows \mathbb{R}^n$. Define sets $U^{\mathrm{A}} := \{(t, \boldsymbol{\xi}) \in I \times \mathbb{R}^n : \boldsymbol{\xi} \in X^{\mathrm{A}}(t)\}$ and $U^{\mathrm{B}} := \{(t, \boldsymbol{\xi}) \in I \times \mathbb{R}^n : \boldsymbol{\xi} \in X^{\mathrm{B}}(t)\}$. Consider functions $\mathbf{x}^{\mathrm{A}}, \mathbf{x}^{\mathrm{B}} : I \to \mathbb{R}^n$. Define a function $\tilde{\mathbf{x}} : I \to \mathbb{R}^n$ so that for each $t \in I$ and each $i \in \{1,...,n\}$, $\tilde{x}_i(t) := \min(x_i^{\mathrm{A}}(t), x_i^{\mathrm{B}}(t))$. Consider vectors $\mathbf{x}_0^{\mathrm{A}} \in X^{\mathrm{A}}(t_0)$, $\mathbf{x}_0^{\mathrm{B}} \in X^{\mathrm{B}}(t_0)$ such that $\mathbf{x}_0^{\mathrm{A}} \leq \mathbf{x}_0^{\mathrm{B}}$. Consider functions $\mathbf{f}^{\mathrm{A}} : U^{\mathrm{A}} \to \mathbb{R}^n$ and $\mathbf{f}^{\mathrm{B}} : U^{\mathrm{B}} \to \mathbb{R}^n$. Suppose that the following conditions hold:

I.1 For each $t \in I$, $\mathbf{x}^{\mathrm{A}}(t), \tilde{\mathbf{x}}(t) \in X^{\mathrm{A}}(t)$ and $\mathbf{x}^{\mathrm{B}}(t) \in X^{\mathrm{B}}(t)$.

I.2 There exists a Lebesgue integrable function $k : I \to \mathbb{R}_+ \cup \{+\infty\}$ so that for any $i \in \{1,...,n\}$, a.e. $t \in I$, and any $\boldsymbol{\xi}, \boldsymbol{\xi}' \in X^{\mathrm{A}}(t)$ for which $\boldsymbol{\xi} \leq \boldsymbol{\xi}'$,

$$f_i^{\mathrm{A}}(t, \boldsymbol{\xi}') - f_i^{\mathrm{A}}(t, \boldsymbol{\xi}) \leq k(t) \|\boldsymbol{\xi}' - \boldsymbol{\xi}\|_\infty.$$

I.3 For any $i \in \{1,...,n\}$, a.e. $t \in I$, any $\boldsymbol{\xi}^{\mathrm{A}} \in X^{\mathrm{A}}(t)$, and any $\boldsymbol{\xi}^{\mathrm{B}} \in X^{\mathrm{B}}(t)$ such that $\xi_i^{\mathrm{A}} = \xi_i^{\mathrm{B}}$ and $\boldsymbol{\xi}^{\mathrm{A}} \leq \boldsymbol{\xi}^{\mathrm{B}}$,

$$f_i^{\mathrm{A}}(t, \boldsymbol{\xi}^{\mathrm{A}}) \leq f_i^{\mathrm{B}}(t, \boldsymbol{\xi}^{\mathrm{B}}).$$

I.4 The functions $\mathbf{x}^{\mathrm{A}}$ and $\mathbf{x}^{\mathrm{B}}$ are C-solutions on I of the following ODEs:

$$\dot{\mathbf{x}}^{\mathrm{A}}(t) = \mathbf{f}^{\mathrm{A}}(t, \mathbf{x}^{\mathrm{A}}(t)), \quad \mathbf{x}^{\mathrm{A}}(t_0) = \mathbf{x}_0^{\mathrm{A}},$$
$$\dot{\mathbf{x}}^{\mathrm{B}}(t) = \mathbf{f}^{\mathrm{B}}(t, \mathbf{x}^{\mathrm{B}}(t)), \quad \mathbf{x}^{\mathrm{B}}(t_0) = \mathbf{x}_0^{\mathrm{B}}.$$

Then,

$$\mathbf{x}^{\mathrm{A}}(t) \leq \mathbf{x}^{\mathrm{B}}(t), \quad \forall t \in I. \tag{3.3.1}$$

Proof. Since $\mathbf{x}_0^{\mathrm{A}} \leq \mathbf{x}_0^{\mathrm{B}}$ by construction, (3.3.1) holds at $t := t_0$. We proceed by showing that $\mathbf{x}^{\mathrm{A}}(t) \leq \mathbf{x}^{\mathrm{B}}(t)$, for all $t \in (t_0, t_f]$.

To arrive a contradiction, suppose that there exists $\tilde{t} \in (t_0, t_f]$ for which $x_i^{\mathrm{B}}(\tilde{t}) < x_i^{\mathrm{A}}(\tilde{t})$

for some $i \in \{1,...,n\}$, and thus define

$$t_1 := \inf\left\{t \in (t_0,t_f] : \exists \kappa \in \{1,...,n\} \text{ for which } x^{\mathrm{B}}_{\kappa}(t) < x^{\mathrm{A}}_{\kappa}(t)\right\} \in I. \tag{3.3.2}$$

Define a function $\boldsymbol{\delta} : I \to \mathbb{R}^n$ for which

$$\boldsymbol{\delta}(t) := \mathbf{x}^{\mathrm{A}}(t) - \mathbf{x}^{\mathrm{B}}(t). \tag{3.3.3}$$

According to Condition I.4, $\mathbf{x}^{\mathrm{A}}$ and $\mathbf{x}^{\mathrm{B}}$ are absolutely continuous on I, and thus, $\boldsymbol{\delta}$ is continuous on I. Applying Lemma 3.3.4 and 3.3.5 in [39] to $\boldsymbol{\delta}$ and t_1, we obtain the following.

Let $\mathbf{1} \in \mathbb{R}^n$ be a vector whose components are all equal to 1. Consider the function k in Condition I.2. It holds that $t_1 < t_f$, and, for any $t_4 \in (t_1,t_f]$, there exist $j \in \{1,...,n\}$, $\varepsilon \in \mathbb{R}_+$, an absolutely continuous and non-decreasing function $\rho : [t_1,t_4] \to \mathbb{R}$ whose derivative a.e. on $[t_1,t_4]$ is denoted as $\dot{\rho}$, and scalars $t_2,t_3 \in [t_1,t_4]$ with $t_2 < t_3$ such that

$$0 < \rho(t) \le \varepsilon, \quad \forall t \in [t_1,t_4], \tag{3.3.4}$$

$$\dot{\rho}(t) > k(t)\rho(t), \quad \text{a.e. } t \in [t_1,t_4], \tag{3.3.5}$$

$$\mathbf{x}^{\mathrm{A}}(t) - \rho(t)\mathbf{1} < \mathbf{x}^{\mathrm{B}}(t), \quad \forall t \in [t_2,t_3), \tag{3.3.6}$$

$$x^{\mathrm{B}}_j(t_2) = x^{\mathrm{A}}_j(t_2), \tag{3.3.7}$$

$$x^{\mathrm{B}}_j(t_3) = x^{\mathrm{A}}_j(t_3) - \rho(t_3), \tag{3.3.8}$$

$$x^{\mathrm{B}}_j(t) < x^{\mathrm{A}}_j(t), \quad \forall t \in (t_2,t_3). \tag{3.3.9}$$

According to (3.3.7) and (3.3.9), for all $t \in [t_2,t_3)$, $x^{\mathrm{B}}_j(t) \le x^{\mathrm{A}}_j(t)$, and thus

$$\tilde{x}_j(t) = x^{\mathrm{B}}_j(t), \quad \forall t \in [t_2,t_3). \tag{3.3.10}$$

By construction of $\tilde{\mathbf{x}}$, for all $t \in [t_2, t_3)$, we also have

$$\tilde{\mathbf{x}}(t) \leq \mathbf{x}^{\mathrm{B}}(t). \tag{3.3.11}$$

Then, since (3.3.10), (3.3.11), Condition I.1, and Condition I.3 hold, it follows that

$$f_j^{\mathrm{B}}(t, \mathbf{x}^{\mathrm{B}}(t)) \geq f_j^{\mathrm{A}}(t, \tilde{\mathbf{x}}(t)), \quad \text{a.e. } t \in [t_2, t_3). \tag{3.3.12}$$

Now, for each $\kappa \in \{1, ..., n\}$ and each $t \in [t_2, t_3)$, one of the following cases will occur.

1. If $x_\kappa^{\mathrm{B}}(t) \geq x_\kappa^{\mathrm{A}}(t)$, then $\tilde{x}_\kappa(t) = x_\kappa^{\mathrm{A}}(t)$ and

$$\tilde{x}_\kappa(t) - x_\kappa^{\mathrm{A}}(t) = 0, \tag{3.3.13}$$

2. If $x_\kappa^{\mathrm{B}}(t) < x_\kappa^{\mathrm{A}}(t)$, then $\tilde{x}_\kappa(t) = x_\kappa^{\mathrm{B}}(t)$; moreover, since (3.3.6) holds,

$$0 < x_\kappa^{\mathrm{A}}(t) - \tilde{x}_\kappa(t) < \rho(t). \tag{3.3.14}$$

The following inequality follows from (3.3.13) and (3.3.14):

$$\|\tilde{\mathbf{x}}(t) - \mathbf{x}^{\mathrm{A}}(t)\|_\infty < \rho(t), \quad \forall t \in [t_2, t_3). \tag{3.3.15}$$

Now, for each $t \in [t_2, t_3)$, since $\tilde{\mathbf{x}}(t) \leq \mathbf{x}^{\mathrm{A}}(t)$ according to the construction of $\tilde{\mathbf{x}}$, the following inequality holds according to Condition I.2,

$$f_j^{\mathrm{A}}(t, \tilde{\mathbf{x}}(t)) \geq f_j^{\mathrm{A}}(t, \mathbf{x}^{\mathrm{A}}(t)) - k(t)\|\mathbf{x}^{\mathrm{A}}(t) - \tilde{\mathbf{x}}(t)\|_\infty, \quad \text{a.e. } t \in [t_2, t_3).$$

Combining this with (3.3.12) yields

$$f_j^{\mathrm{B}}(t,\mathbf{x}^{\mathrm{B}}(t)) \geq f_j^{\mathrm{A}}(t,\mathbf{x}^{\mathrm{A}}(t)) - k(t)\|\mathbf{x}^{\mathrm{A}}(t) - \tilde{\mathbf{x}}(t)\|_\infty, \quad \text{a.e. } t \in [t_2, t_3).$$

Applying (3.3.15) yields

$$f_j^{\mathrm{B}}(t,\mathbf{x}^{\mathrm{B}}(t)) > f_j^{\mathrm{A}}(t,\mathbf{x}^{\mathrm{A}}(t)) - k(t)\rho(t), \quad \text{a.e. } t \in [t_2, t_3).$$

Since $\dot{\rho}(t) > k(t)\rho(t)$ for a.e. $t \in [t_2, t_3]$ according to (3.3.5), rearranging the above inequality yields,

$$f_j^{\mathrm{A}}(t,\mathbf{x}^{\mathrm{A}}(t)) - f_j^{\mathrm{B}}(t,\mathbf{x}^{\mathrm{B}}(t)) - \dot{\rho}(t) < 0, \quad \text{a.e. } t \in [t_2, t_3].$$

Since Condition I.4 holds, Theorem 3.1 in [87] implies that $\left(x_j^{\mathrm{A}}(t) - x_j^{\mathrm{B}}(t) - \rho(t)\right)$ is decreasing with respect to t on $[t_2, t_3]$, which in turn implies

$$x_j^{\mathrm{A}}(t_3) - x_j^{\mathrm{B}}(t_3) - \rho(t_3) < x_j^{\mathrm{A}}(t_2) - x_j^{\mathrm{B}}(t_2) - \rho(t_2). \tag{3.3.16}$$

However, according to (3.3.7) and (3.3.8), $x_j^{\mathrm{A}}(t_3) - x_j^{\mathrm{B}}(t_3) - \rho(t_3) = 0$ and $x_j^{\mathrm{A}}(t_2) - x_j^{\mathrm{B}}(t_2) = 0$. Then, (3.3.16) becomes $\rho(t_2) < 0$, which contradicts (3.3.4). Thus, $\tilde{t}$ cannot exist. □

Remark 3.3.2. Condition I.2 is a mild assumption that is implied by $\mathbf{f}$ with right-uniqueness conditions (such as [77, Theorem 6.IX]). Unlike the established one-sided bounding results for Carathéodory ODE systems summarized in Table 3.1, Theorem 3.3.1 requires neither the quasi-monotonicity assumption on $\mathbf{f}^{\mathrm{A}}$ or $\mathbf{f}^{\mathrm{B}}$ nor differentiability of $\mathbf{x}^{\mathrm{A}}$ and $\mathbf{x}^{\mathrm{B}}$. On the other hand, Theorem 3.3.1 has a new domain requirement formulated as Condition I.1. From our experience, this requirement is typically satisfied by dynamic models of engineering processes.

3.4 Bound comparison

Based on Theorem 3.3.1, the following theorem provides new sufficient conditions under which one dynamic bounding pair is guaranteed to enclose another.

Theorem 3.4.1. Let $I := [t_0, t_f] \subsetneq \mathbb{R}$. Consider set-valued mappings $C^{\mathrm{A}}, C^{\mathrm{B}} : I \rightrightarrows \mathbb{R}^n$. Define sets $D^{\mathrm{A}} := \{(t, \boldsymbol{\phi}, \boldsymbol{\psi}) \in I \times \mathbb{R}^n \times \mathbb{R}^n : \boldsymbol{\phi}, \boldsymbol{\psi} \in C^{\mathrm{A}}(t)\}$ and $D^{\mathrm{B}} := \{(t, \boldsymbol{\phi}, \boldsymbol{\psi}) \in I \times \mathbb{R}^n \times \mathbb{R}^n : \boldsymbol{\phi}, \boldsymbol{\psi} \in C^{\mathrm{B}}(t)\}$. Consider functions $\mathbf{v}^{\mathrm{A}}, \mathbf{w}^{\mathrm{A}}, \mathbf{v}^{\mathrm{B}}, \mathbf{w}^{\mathrm{B}} : I \to \mathbb{R}^n$. Define functions $\tilde{\mathbf{v}}, \tilde{\mathbf{w}} : I \to \mathbb{R}^n$ so that for each $i \in \{1, \ldots, n\}$ and each $t \in I$, $\tilde{v}_i := \min(v_i^{\mathrm{A}}(t), v_i^{\mathrm{B}}(t))$ and $\tilde{w}_i := \max(w_i^{\mathrm{A}}(t), w_i^{\mathrm{B}}(t))$. Consider vectors $\mathbf{v}_0^{\mathrm{A}}, \mathbf{w}_0^{\mathrm{A}} \in C^{\mathrm{A}}(t_0)$ and $\mathbf{v}_0^{\mathrm{B}}, \mathbf{v}_0^{\mathrm{B}} \in C^{\mathrm{B}}(t_0)$ so that $\mathbf{v}_0^{\mathrm{A}} \leq \mathbf{v}_0^{\mathrm{B}} \leq \mathbf{w}_0^{\mathrm{B}} \leq \mathbf{w}_0^{\mathrm{A}}$. Consider functions $\mathbf{d}^{\mathrm{L,A}}, \mathbf{d}^{\mathrm{U,A}} : D^{\mathrm{A}} \to \mathbb{R}^n$ and $\mathbf{d}^{\mathrm{L,B}}, \mathbf{d}^{\mathrm{U,B}} : D^{\mathrm{B}} \to \mathbb{R}^n$. Suppose that the following conditions hold:

II.1 For each $t \in I$, it follows that $\mathbf{v}^{\mathrm{A}}(t), \mathbf{w}^{\mathrm{A}}(t), \tilde{\mathbf{v}}(t), \tilde{\mathbf{w}}(t) \in C^{\mathrm{A}}(t)$ and $\mathbf{v}^{\mathrm{B}}(t), \mathbf{w}^{\mathrm{B}}(t) \in C^{\mathrm{B}}(t)$.

II.2 There exists a Lebesgue integrable function $k : I \to \mathbb{R}_+ \cup \{+\infty\}$ so that for any $i \in \{1, \ldots, n\}$, a.e. $t \in I$, and any $\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\phi}', \boldsymbol{\psi}' \in C^{\mathrm{A}}(t)$ for which $\boldsymbol{\phi} \leq \boldsymbol{\phi}' \leq \boldsymbol{\psi}' \leq \boldsymbol{\psi}$,

$$\begin{aligned} & d_i^{\mathrm{L,A}}(t, \boldsymbol{\phi}', \boldsymbol{\psi}') - d_i^{\mathrm{L,A}}(t, \boldsymbol{\phi}, \boldsymbol{\psi}) \leq k(t)(\|\boldsymbol{\phi}' - \boldsymbol{\phi}\|_\infty + \|\boldsymbol{\psi} - \boldsymbol{\psi}'\|_\infty), \\ \text{and} \quad & d_i^{\mathrm{U,A}}(t, \boldsymbol{\phi}, \boldsymbol{\psi}) - d_i^{\mathrm{U,A}}(t, \boldsymbol{\phi}', \boldsymbol{\psi}') \leq k(t)(\|\boldsymbol{\phi}' - \boldsymbol{\phi}\|_\infty + \|\boldsymbol{\psi} - \boldsymbol{\psi}'\|_\infty). \end{aligned}$$

II.3 For any $i \in \{1, \ldots, n\}$, a.e. $t \in I$, any $\boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}} \in C^{\mathrm{A}}(t)$, and any $\boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}} \in C^{\mathrm{B}}(t)$ such that $\boldsymbol{\phi}^{\mathrm{A}} \leq \boldsymbol{\phi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{A}}$,

(a) if $\phi_i^{\mathrm{A}} = \phi_i^{\mathrm{B}}$, then $d_i^{\mathrm{L,B}}(t, \boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}}) \geq d_i^{\mathrm{L,A}}(t, \boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}})$, and

(b) if $\psi_i^{\mathrm{A}} = \psi_i^{\mathrm{B}}$, then $d_i^{\mathrm{U,B}}(t, \boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}}) \leq d_i^{\mathrm{U,A}}(t, \boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}})$.

II.4 For all $t \in I$,

$$\mathbf{v}^{\mathrm{A}}(t) \leq \mathbf{w}^{\mathrm{A}}(t) \quad \text{and} \quad \mathbf{v}^{\mathrm{B}}(t) \leq \mathbf{w}^{\mathrm{B}}(t).$$

II.5 The functions $\mathbf{v}^{\mathrm{A}}, \mathbf{v}^{\mathrm{B}}, \mathbf{w}^{\mathrm{A}}, \mathbf{w}^{\mathrm{B}}$ are C-solutions on I of the following ODEs:

$$\begin{aligned}
\dot{\mathbf{v}}^{\mathrm{A}}(t) &= \mathbf{d}^{\mathrm{L,A}}(t, \mathbf{v}^{\mathrm{A}}(t), \mathbf{w}^{\mathrm{A}}(t)), \quad \mathbf{v}^{\mathrm{A}}(t_0) = \mathbf{v}_0^{\mathrm{A}}, \\
\dot{\mathbf{w}}^{\mathrm{A}}(t) &= \mathbf{d}^{\mathrm{U,A}}(t, \mathbf{v}^{\mathrm{A}}(t), \mathbf{w}^{\mathrm{A}}(t)), \quad \mathbf{w}^{\mathrm{A}}(t_0) = \mathbf{w}_0^{\mathrm{A}}, \\
\dot{\mathbf{v}}^{\mathrm{B}}(t) &= \mathbf{d}^{\mathrm{L,B}}(t, \mathbf{v}^{\mathrm{B}}(t), \mathbf{w}^{\mathrm{B}}(t)), \quad \mathbf{v}^{\mathrm{B}}(t_0) = \mathbf{v}_0^{\mathrm{B}}, \\
\dot{\mathbf{w}}^{\mathrm{B}}(t) &= \mathbf{d}^{\mathrm{U,B}}(t, \mathbf{v}^{\mathrm{B}}(t), \mathbf{w}^{\mathrm{B}}(t)), \quad \mathbf{w}^{\mathrm{B}}(t_0) = \mathbf{w}_0^{\mathrm{B}}.
\end{aligned} \tag{3.4.1}$$

Then, the following inequalities hold:

$$\mathbf{v}^{\mathrm{A}}(t) \leq \mathbf{v}^{\mathrm{B}}(t) \leq \mathbf{w}^{\mathrm{B}}(t) \leq \mathbf{w}^{\mathrm{A}}(t), \quad \forall t \in I. \tag{3.4.2}$$

Proof. Since Condition II.4 holds, it only remains to be shown that for all $t \in I$,

$$\mathbf{v}^{\mathrm{A}}(t) \leq \mathbf{v}^{\mathrm{B}}(t) \quad \text{and} \quad \mathbf{w}^{\mathrm{A}}(t) \geq \mathbf{w}^{\mathrm{B}}(t). \tag{3.4.3}$$

Define functions $\mathbf{x}^{\mathrm{A}}, \mathbf{x}^{\mathrm{B}} : I \to \mathbb{R}^{2n}$ so that for all $t \in I$,

$$\mathbf{x}^{\mathrm{A}}(t) := (\mathbf{v}^{\mathrm{A}}(t), -\mathbf{w}^{\mathrm{A}}(t)) \quad \text{and} \quad \mathbf{x}^{\mathrm{B}}(t) := (\mathbf{v}^{\mathrm{B}}(t), -\mathbf{w}^{\mathrm{B}}(t)).$$

To prove (3.4.3), it suffices to show that

$$\mathbf{x}^{\mathrm{A}}(t) \leq \mathbf{x}^{\mathrm{B}}(t), \quad \forall t \in I. \tag{3.4.4}$$

Define set-valued mappings $X^{\mathrm{A}}, X^{\mathrm{B}} : I \rightrightarrows \mathbb{R}^{2n}$ so that for each $t \in I$,

$$X^{\mathrm{A}}(t) := \{(\boldsymbol{\phi}, -\boldsymbol{\psi}) \in \mathbb{R}^{2n} : \boldsymbol{\phi}, \boldsymbol{\psi} \in C^{\mathrm{A}}(t), \text{ and } \boldsymbol{\phi} \leq \boldsymbol{\psi}\},$$
$$\text{and} \quad X^{\mathrm{B}}(t) := \{(\boldsymbol{\phi}, -\boldsymbol{\psi}) \in \mathbb{R}^{2n} : \boldsymbol{\phi}, \boldsymbol{\psi} \in C^{\mathrm{B}}(t), \text{ and } \boldsymbol{\phi} \leq \boldsymbol{\psi}\}.$$

Define sets $U^{\mathrm{A}} := \{(t, \boldsymbol{\xi}) \in I \times \mathbb{R}^{2n} : \boldsymbol{\xi} \in X^{\mathrm{A}}(t)\}$ and $U^{\mathrm{B}} := \{(t, \boldsymbol{\xi}) \in I \times \mathbb{R}^{2n} : \boldsymbol{\xi} \in X^{\mathrm{B}}(t)\}$. Define functions $\mathbf{f}^{\mathrm{A}} : U^{\mathrm{A}} \to \mathbb{R}^{2n}$ and $\mathbf{f}^{\mathrm{B}} : U^{\mathrm{B}} \to \mathbb{R}^{2n}$ so that, with $\boldsymbol{\xi}^{\mathrm{A}} := (\boldsymbol{\phi}^{\mathrm{A}}, -\boldsymbol{\psi}^{\mathrm{A}})$ and $\boldsymbol{\xi}^{\mathrm{B}} := (\boldsymbol{\phi}^{\mathrm{B}}, -\boldsymbol{\psi}^{\mathrm{B}})$, for each $(t, \boldsymbol{\xi}^{\mathrm{A}}) \in U^{\mathrm{A}}$ and each $(t, \boldsymbol{\xi}^{\mathrm{B}}) \in U^{\mathrm{B}}$,

$$\mathbf{f}^{\mathrm{A}}(t, \boldsymbol{\xi}^{\mathrm{A}}) := (\mathbf{d}^{\mathrm{L,A}}(t, \boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}}), -\mathbf{d}^{\mathrm{U,A}}(t, \boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}})),$$
$$\text{and} \quad \mathbf{f}^{\mathrm{B}}(t, \boldsymbol{\xi}^{\mathrm{B}}) := (\mathbf{d}^{\mathrm{L,B}}(t, \boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}}), -\mathbf{d}^{\mathrm{U,B}}(t, \boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}})).$$

Define a function $\tilde{\mathbf{x}}(t) : I \to \mathbb{R}^{2n}$ so that for each $i \in \{1, ..., 2n\}$, $\tilde{x}_i := \min(x_i^{\mathrm{A}}(t), x_i^{\mathrm{B}}(t))$. To prove (3.4.4), we proceed by showing that all conditions of Theorem 3.3.1 are satisfied with $(\mathbf{x}^{\mathrm{A}}, \mathbf{x}^{\mathrm{B}}, \mathbf{f}^{\mathrm{A}}, \mathbf{f}^{\mathrm{B}})$ as defined above.

Since Condition II.1 and II.4 hold, for each $t \in I$, $\mathbf{x}^{\mathrm{A}}(t) \in X^{\mathrm{A}}(t)$ and $\mathbf{x}^{\mathrm{B}}(t) \in X^{\mathrm{B}}(t)$. Note that by construction of $\tilde{\mathbf{x}}$,

$$\tilde{\mathbf{x}}(t) \equiv (\tilde{\mathbf{v}}(t), -\tilde{\mathbf{w}}(t)).$$

For each $t \in I$, since Condition II.4 holds, it follows that $\tilde{\mathbf{v}}(t) \leq \tilde{\mathbf{w}}(t)$. Furthermore, since $\tilde{\mathbf{v}}(t), \tilde{\mathbf{w}}(t) \in C^{\mathrm{A}}(t)$ according to Condition II.1, $\tilde{\mathbf{x}}(t) \in X^{\mathrm{A}}(t)$. Thus, Condition I.1 is satisfied.

Consider any $i \in \{1, ..., 2n\}$, a.e. $t \in I$, and any $\boldsymbol{\xi}, \boldsymbol{\xi}' \in X^{\mathrm{A}}(t)$ so that $\boldsymbol{\xi}' \geq \boldsymbol{\xi}$. Define $\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\phi}', \boldsymbol{\psi}' \in C^{\mathrm{A}}(t)$ so that $\boldsymbol{\xi} := (\boldsymbol{\phi}, -\boldsymbol{\psi})$ and $\boldsymbol{\xi}' := (\boldsymbol{\phi}', -\boldsymbol{\psi}')$. According to the definition of the l-infinity norm,

$$\|\boldsymbol{\xi}' - \boldsymbol{\xi}\|_\infty = \max(\|\boldsymbol{\phi}' - \boldsymbol{\phi}\|_\infty, \|\boldsymbol{\psi} - \boldsymbol{\psi}'\|_\infty),$$

and thus

$$2k(t)\|\boldsymbol{\xi}' - \boldsymbol{\xi}\|_\infty \geq k(t)\big(\|\boldsymbol{\phi}' - \boldsymbol{\phi}\|_\infty + \|\boldsymbol{\psi} - \boldsymbol{\psi}'\|_\infty\big). \tag{3.4.5}$$

Since $\boldsymbol{\xi}, \boldsymbol{\xi}' \in X^{\mathrm{A}}(t)$ and $\boldsymbol{\xi}' \geq \boldsymbol{\xi}$, we obtain $\boldsymbol{\phi} \leq \boldsymbol{\phi}' \leq \boldsymbol{\psi}' \leq \boldsymbol{\psi}$. Combining Condition II.2 with (3.4.5) then yields

$$f_i^{\mathrm{A}}(t, \boldsymbol{\xi}') - f_i^{\mathrm{A}}(t, \boldsymbol{\xi}) \leq 2k(t)\|\boldsymbol{\xi}' - \boldsymbol{\xi}\|_\infty.$$

Thus, Condition I.2 is satisfied.

Consider any $i \in \{1, ..., 2n\}$, a.e. $t \in I$, any $\boldsymbol{\xi}^{\mathrm{A}} \in X^{\mathrm{A}}(t)$, and any $\boldsymbol{\xi}^{\mathrm{B}} \in X^{\mathrm{B}}(t)$ for which $\xi_i^{\mathrm{A}} = \xi_i^{\mathrm{B}}$ and $\boldsymbol{\xi}^{\mathrm{B}} \geq \boldsymbol{\xi}^{\mathrm{A}}$. Define $\boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}} \in C^{\mathrm{A}}(t)$ and $\boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}} \in C^{\mathrm{B}}(t)$ for which $\boldsymbol{\xi}^{\mathrm{A}} \equiv (\boldsymbol{\phi}^{\mathrm{A}}, -\boldsymbol{\psi}^{\mathrm{A}})$ and $\boldsymbol{\xi}^{\mathrm{B}} \equiv (\boldsymbol{\phi}^{\mathrm{B}}, -\boldsymbol{\psi}^{\mathrm{B}})$. It follows that $\boldsymbol{\phi}^{\mathrm{A}} \leq \boldsymbol{\phi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{A}}$. Moreover, since Condition II.3 holds,

$$f_i^{\mathrm{B}}(t, \boldsymbol{\xi}^{\mathrm{B}}) \geq f_i^{\mathrm{A}}(t, \boldsymbol{\xi}^{\mathrm{A}}).$$

Thus, Condition I.3 is satisfied.

Let $\mathbf{x}_0^{\mathrm{A}} := (\mathbf{v}_0^{\mathrm{A}}, -\mathbf{w}_0^{\mathrm{A}})$ and $\mathbf{x}_0^{\mathrm{B}} := (\mathbf{v}_0^{\mathrm{B}}, -\mathbf{w}_0^{\mathrm{B}})$. Since Condition II.5 holds, it follows that $\mathbf{x}_0^{\mathrm{A}} \in X^{\mathrm{A}}(t_0)$, $\mathbf{x}_0^{\mathrm{B}} \in X^{\mathrm{B}}(t_0)$, and $\mathbf{x}_0^{\mathrm{A}} \leq \mathbf{x}_0^{\mathrm{B}}$. Thus, the functions $\mathbf{x}^{\mathrm{A}}$ and $\mathbf{x}^{\mathrm{B}}$ are C-solutions on I of the following ODEs:

$$\dot{\mathbf{x}}^{\mathrm{A}}(t) = \mathbf{f}^{\mathrm{A}}(t, \mathbf{x}^{\mathrm{A}}(t)), \quad \mathbf{x}^{\mathrm{A}}(t_0) = \mathbf{x}_0^{\mathrm{A}},$$

$$\dot{\mathbf{x}}^{\mathrm{B}}(t) = \mathbf{f}^{\mathrm{B}}(t, \mathbf{x}^{\mathrm{B}}(t)), \quad \mathbf{x}^{\mathrm{B}}(t_0) = \mathbf{x}_0^{\mathrm{B}}.$$

Thus, Condition I.4 is satisfied. Hence, Theorem 3.3.1 yields (3.4.4). □

Remark 3.4.2. Condition II.1 is a mild assumption that is satisfied if, for each t, $C^{\mathrm{A}}(t)$ is an interval and $C^{\mathrm{B}}(t)$ is a subset of $C^{\mathrm{A}}(t)$.

To the authors' knowledge, the only previous bound comparison result was proposed by Schaber et al. ([79, Theorem 4.14]). Theorem 3.4.1 has less stringent assumptions than [79, Theorem 4.14] in the following respects:

1. [79, Theorem 4.14] requires the right-hand side functions $(\mathbf{d}^{\mathrm{L,A}}, \mathbf{d}^{\mathrm{U,A}})$ of the ODEs in $(\mathbf{v}^{\mathrm{A}}, \mathbf{w}^{\mathrm{A}})$ in (3.4.1) to be locally Lipschitz continuous. Theorem 3.4.1 instead requires Condition II.2, which is satisfied if $(\mathbf{d}^{\mathrm{L,A}}, \mathbf{d}^{\mathrm{U,A}})$ obey a less stringent right-uniqueness condition [77, Theorem 6.IX].

2. [79, Theorem 4.14] requires $\|\mathbf{w}^{\mathrm{A}}(t) - \mathbf{v}^{\mathrm{A}}(t)\|_\infty$ to be increasing with respect to t, and requires $(\dot{\mathbf{v}}^{\mathrm{B}}(t) \geq \dot{\mathbf{v}}^{\mathrm{A}}(t))$, and $(\dot{\mathbf{w}}^{\mathrm{A}}(t) \geq \dot{\mathbf{w}}^{\mathrm{B}}(t))$. Instead, Theorem 3.4.1 has no requirement on $\|\mathbf{w}^{\mathrm{A}}(t) - \mathbf{v}^{\mathrm{A}}(t)\|_\infty$, only requires $\dot{v}_i^{\mathrm{B}}(t) \geq \dot{v}_i^{\mathrm{A}}(t)$ if $v_i^{\mathrm{B}}(t) = v_i^{\mathrm{A}}(t)$, and only requires $\dot{w}_i^{\mathrm{A}}(t) \geq \dot{w}_i^{\mathrm{B}}(t)$ if $w_i^{\mathrm{A}}(t) = w_i^{\mathrm{B}}(t)$, as shown in Condition II.3.

3. [79, Theorem 4.14] assumes the sets C^{A} and C^{B} in Theorem 3.4.1 to be open, equal, and independent of t. On the other hand, Theorem 3.4.1 permits C^{A} and C^{B} to be distinct and not necessarily open.

The next section shows that, due to these less stringent assumptions, Theorem 3.4.1 applies to the ODE relaxations given by [2].

3.5 Tightness results for ODE relaxations

Based on the new bound comparison theorem (Theorem 3.4.1), this section presents new tightness results for a state-of-the-art ODE relaxation framework by Scott and Barton [2] summarized in Section 2.4. The Scott–Barton framework constructs convex and concave relaxations for solutions of parametric ODE systems with respect to parameters, to ultimately provide bounding information to algorithms for deterministic global dynamic optimization [19, 27]. In this context, tighter ODE relaxations translate into tighter bounds for globally optimal objective values, and would thus reduce the number of iterations required in these algorithms. Such results could not be obtained using Schaber et al.'s established bound comparison result [79, Theorem 4.14]. The reason is that [79, Theorem 4.14] requires the outer bounding system's right-hand side to be locally Lipschitz continuous, while the right-hand side of the framework (2.4.1), as shown in [2, 3], only satisfies a right-uniqueness condition ([102, §10, Theorem 1]) that is weaker than the Lipschitz continuity. Moreover, as discussed in Section 3.4, Theorem 3.4.1 has less stringent differential inequality conditions and set requirements than [79, Theorem 4.14]. These less stringent conditions do indeed help compare the established state relaxations (c.f. [3, Section 5.5]) in the framework. However, this comparison cannot be done based on the differential inequality conditions and set requirements of [79, Theorem 4.14].

Thus, the following new tightness result of the Scott–Barton framework (2.4.1) is developed based on Theorem 3.4.1. Suppose that there are two considered choices for $(\mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}}, \mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}, \mathbf{u}, \mathbf{o})$ in (2.4.1); call these $(\mathbf{x}_0^{\mathrm{cv,A}}, \mathbf{x}_0^{\mathrm{cc,A}}, \mathbf{x}^{\mathrm{L,A}}, \mathbf{x}^{\mathrm{U,A}}, \mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}})$ and $(\mathbf{x}_0^{\mathrm{cv,B}}, \mathbf{x}_0^{\mathrm{cc,B}}, \mathbf{x}^{\mathrm{L,B}}, \mathbf{x}^{\mathrm{U,B}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$, and denote the resulting state relaxations as $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$, respectively. Theorem 3.5.1 below shows that if the state bounds $(\mathbf{x}^{\mathrm{L,B}}, \mathbf{x}^{\mathrm{U,B}})$ lie within $(\mathbf{x}^{\mathrm{L,A}}, \mathbf{x}^{\mathrm{U,A}})$, if the initial relaxations $(\mathbf{x}_0^{\mathrm{cv,B}}, \mathbf{x}_0^{\mathrm{cc,B}})$ lie within $(\mathbf{x}_0^{\mathrm{cv,A}}, \mathbf{x}_0^{\mathrm{cc,A}})$, and if the functions $(\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$

satisfy a condition that resembles Condition II.3 in Theorem 3.4.1, then the state relaxations $(\mathbf{x}^{\mathrm{cv,B}},\mathbf{x}^{\mathrm{cc,B}})$ are guaranteed to lie within the state relaxations $(\mathbf{x}^{\mathrm{cv,A}},\mathbf{x}^{\mathrm{cc,A}})$.

Theorem 3.5.1. Consider state lower bounds $\mathbf{x}^{\mathrm{L,A}},\mathbf{x}^{\mathrm{L,B}}: I \to \mathbb{R}^{n_x}$ and state upper bounds $\mathbf{x}^{\mathrm{U,A}},\mathbf{x}^{\mathrm{U,B}}: I \to \mathbb{R}^{n_x}$ for (2.3.1) on $I \times P$ that are absolutely continuous, and suppose for all $t \in I$ that $\mathbf{x}^{\mathrm{L,A}}(t) \leq \mathbf{x}^{\mathrm{L,B}}(t) \leq \mathbf{x}^{\mathrm{U,B}}(t) \leq \mathbf{x}^{\mathrm{U,A}}(t)$. For each $t \in I$, denote the intervals $[\mathbf{x}^{\mathrm{L,A}}(t),\mathbf{x}^{\mathrm{U,A}}(t)]$ and $[\mathbf{x}^{\mathrm{L,B}}(t),\mathbf{x}^{\mathrm{U,B}}(t)]$ as $X^{\mathrm{A}}(t)$ and $X^{\mathrm{B}}(t)$, respectively. Consider initial convex relaxations $\mathbf{x}_0^{\mathrm{cv,A}},\mathbf{x}_0^{\mathrm{cv,B}}: P \to \mathbb{R}^{n_x}$ and initial concave relaxations $\mathbf{x}_0^{\mathrm{cc,A}},\mathbf{x}_0^{\mathrm{cc,B}}: P \to \mathbb{R}^{n_x}$ for (2.3.1), and suppose for all $\mathbf{p} \in P$ that $\mathbf{x}_0^{\mathrm{cv,A}}(\mathbf{p}) \leq \mathbf{x}_0^{\mathrm{cv,B}}(\mathbf{p}) \leq \mathbf{x}_0^{\mathrm{cc,B}}(\mathbf{p}) \leq \mathbf{x}_0^{\mathrm{cc,A}}(\mathbf{p})$. Consider functions $\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}},\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}}: I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$, and suppose that the following conditions hold.

III.1 The functions $(\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}})$ and $(\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}})$ are Scott–Barton right-hand side functions for (2.3.1) as in Definition 2.4.9 based on bounds $X^{\mathrm{A}}(t)$ and $X^{\mathrm{B}}(t)$, respectively.

III.2 For any $i \in \{1,...,n_x\}$, any $\mathbf{p} \in P$, a.e. $t \in I$, any $\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}} \in X^{\mathrm{A}}(t)$, and any $\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}} \in X^{\mathrm{B}}(t)$ such that $\boldsymbol{\phi}^{\mathrm{A}} \leq \boldsymbol{\phi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{A}}$,

   (a) if $\phi_i^{\mathrm{A}} = \phi_i^{\mathrm{B}}$, then $u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}}) \leq u_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}})$,

   (b) if $\psi_i^{\mathrm{A}} = \psi_i^{\mathrm{B}}$, then $o_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}}) \geq o_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}})$.

Let $(\mathbf{x}^{\mathrm{cv,A}},\mathbf{x}^{\mathrm{cc,A}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}},\mathbf{x}_0^{\mathrm{cv}},\mathbf{x}_0^{\mathrm{cc}},\mathbf{u},\mathbf{o}) := (\mathbf{x}^{\mathrm{L,A}},\mathbf{x}^{\mathrm{U,A}},\mathbf{x}_0^{\mathrm{cv,A}},\mathbf{x}_0^{\mathrm{cc,A}},\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}}).$$

Let $(\mathbf{x}^{\mathrm{cv,B}},\mathbf{x}^{\mathrm{cc,B}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}},\mathbf{x}_0^{\mathrm{cv}},\mathbf{x}_0^{\mathrm{cc}},\mathbf{u},\mathbf{o}) := (\mathbf{x}^{\mathrm{L,B}},\mathbf{x}^{\mathrm{U,B}},\mathbf{x}_0^{\mathrm{cv,B}},\mathbf{x}_0^{\mathrm{cc,B}},\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}}).$$

Then, $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ are both valid state relaxations for (2.3.1) on $I \times P$. Moreover, for any $(t, \mathbf{p}) \in I \times P$,

$$\mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p}). \tag{3.5.1}$$

Proof. Since Condition III.1 holds, [2, Corollary 1 and Theorem 3] imply that both $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ are valid state relaxations for (2.3.1) on $I \times P$. Throughout the remainder of this proof, consider any fixed $\mathbf{p} \in P$. The inequality (3.5.1) will be demonstrated by verifying all conditions in Theorem 3.4.1 with the following substitutions: for all $t \in I$,

$$\begin{aligned} \mathbf{v}^{\mathrm{A}}(t) &:= \mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}), \quad \mathbf{w}^{\mathrm{A}}(t) := \mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p}), \\ \mathbf{v}^{\mathrm{B}}(t) &:= \mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}), \quad \mathbf{w}^{\mathrm{B}}(t) := \mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p}). \end{aligned} \tag{3.5.2}$$

Since Condition III.1 holds, [2, Lemma 1] implies that for all $t \in I$,

$$\begin{aligned} & \mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}), \mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p}) \in X^{\mathrm{A}}(t), \\ \text{and} \quad & \mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}), \mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p}) \in X^{\mathrm{B}}(t). \end{aligned} \tag{3.5.3}$$

Furthermore, since for each $t \in I$, $X^{\mathrm{A}}(t)$ is an interval and $X^{\mathrm{B}}(t) \subseteq X^{\mathrm{A}}(t)$, Remark 3.4.2 implies that Condition II.1 is satisfied with $(X^{\mathrm{A}}, X^{\mathrm{B}})$ in place of $(C^{\mathrm{A}}, C^{\mathrm{B}})$.

Since $(\mathbf{x}^{\mathrm{L,A}}, \mathbf{x}^{\mathrm{U,A}}, \mathbf{x}^{\mathrm{L,B}}, \mathbf{x}^{\mathrm{U,B}})$ are absolutely continuous on I, they are differentiable almost everywhere. Thus, there is a zero-measure subset $\tilde{I} \subsetneq I$ for which the derivatives $\dot{\mathbf{x}}^{\mathrm{L,A}}, \dot{\mathbf{x}}^{\mathrm{U,A}}, \dot{\mathbf{x}}^{\mathrm{L,B}}, \dot{\mathbf{x}}^{\mathrm{U,B}}$ are well-defined on $I \backslash \tilde{I}$ and for which $t_f \in \tilde{I}$. Define functions

$\mathbf{u}_{\mathrm{r}}^{\mathrm{A}}, \mathbf{o}_{\mathrm{r}}^{\mathrm{A}} : I \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ so that for each $i \in \{1, ..., n_x\}$ and $(t, \boldsymbol{\phi}, \boldsymbol{\psi}) \in I \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$,

$$
\begin{aligned}
u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi},\boldsymbol{\psi}) &:= \begin{cases} u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi}), & \text{if} \quad \phi_i > x_i^{\mathrm{L,A}}(t) \text{ and } t \in I \backslash \tilde{I}, \\ \max\left(\dot{x}_i^{\mathrm{L,A}}(t), u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi})\right), & \text{if} \quad \phi_i \leq x_i^{\mathrm{L,A}}(t) \text{ and } t \in I \backslash \tilde{I}, \\ 0, & \text{if} \quad t \in \tilde{I}, \end{cases} \\
o_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi},\boldsymbol{\psi}) &:= \begin{cases} o_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi}), & \text{if} \quad \psi_i < x_i^{\mathrm{U,A}}(t) \text{ and } t \in I \backslash \tilde{I}, \\ \min\left(\dot{x}_i^{\mathrm{U,A}}(t), o_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi})\right), & \text{if} \quad \psi_i \geq x_i^{\mathrm{U,A}}(t) \text{ and } t \in I \backslash \tilde{I}, \\ 0 & \text{if} \quad t \in \tilde{I}. \end{cases}
\end{aligned} \tag{3.5.4}
$$

Define functions $\mathbf{u}_{\mathrm{r}}^{\mathrm{B}}, \mathbf{o}_{\mathrm{r}}^{\mathrm{B}} : I \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ so that for each $i \in \{1, ..., n_x\}$ and $(t, \boldsymbol{\phi}, \boldsymbol{\psi}) \in I \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, $u_{\mathrm{r},i}^{\mathrm{B}}(t, \boldsymbol{\phi}, \boldsymbol{\psi})$ and $o_{\mathrm{r},i}^{\mathrm{B}}(t, \boldsymbol{\phi}, \boldsymbol{\psi})$ are defined by (3.5.4) except with $(u_{\mathrm{r},i}^{\mathrm{B}}, o_{\mathrm{r},i}^{\mathrm{B}}, u_i^{\mathrm{B}}, o_i^{\mathrm{B}}, x_i^{\mathrm{L,B}}, x_i^{\mathrm{U,B}})$ in place of $(u_{\mathrm{r},i}^{\mathrm{A}}, o_{\mathrm{r},i}^{\mathrm{A}}, u_i^{\mathrm{A}}, o_i^{\mathrm{A}}, x_i^{\mathrm{L,A}}, x_i^{\mathrm{U,A}})$.

Now, we show that the functions $(\mathbf{u}_{\mathrm{r}}^{\mathrm{A}}, \mathbf{o}_{\mathrm{r}}^{\mathrm{A}})$ satisfy Condition II.2 of Theorem 3.4.1 in place of $(\mathbf{d}^{\mathrm{L,A}}, \mathbf{d}^{\mathrm{U,A}})$. Consider any $i \in \{1, ..., n_x\}$, any $t \in I \backslash \tilde{I}$, and any $\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\phi}', \boldsymbol{\psi}' \in X^{\mathrm{A}}(t)$ such that $\phi_i' \geq \phi_i$. Since $\mathbf{u}^{\mathrm{A}}(t, \mathbf{p}, \cdot, \cdot)$ is Lipschitz continuous on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly over $(t, \mathbf{p}) \in I \times P$, there exists $k \geq 0$ so that the following conditions are satisfied.

1. If either $\phi_i = x_i^{\mathrm{L,A}}(t)$ and $\phi_i' = x_i^{\mathrm{L,A}}(t)$, or $\phi_i > x_i^{\mathrm{L,A}}(t)$ and $\phi_i' > x_i^{\mathrm{L,A}}(t)$, then

$$
|u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi},\boldsymbol{\psi}) - u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}',\boldsymbol{\psi}')| \leq k\big(\|\boldsymbol{\phi} - \boldsymbol{\phi}'\|_\infty + \|\boldsymbol{\psi} - \boldsymbol{\psi}'\|_\infty\big). \tag{3.5.5}
$$

2. If $\phi_i = x_i^{\mathrm{L,A}}(t)$ and $\phi_i' > x_i^{\mathrm{L,A}}(t)$, then

$$
u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}',\boldsymbol{\psi}') - u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi},\boldsymbol{\psi}) \equiv u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}') - \max(\dot{x}_i^{\mathrm{L,A}}(t), u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi})).
$$

Thus,

(a) if $\dot{x}_i^{\mathrm{L,A}}(t) \leq u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi})$, then

$$\begin{aligned} u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}',\boldsymbol{\psi}') - u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi},\boldsymbol{\psi}) &\equiv u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}') - u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi}) \\ &\leq k\big(\|\boldsymbol{\phi}-\boldsymbol{\phi}'\|_\infty + \|\boldsymbol{\psi}-\boldsymbol{\psi}'\|_\infty\big), \end{aligned} \tag{3.5.6}$$

(b) if $\dot{x}_i^{\mathrm{L,A}}(t) > u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi})$, then

$$\begin{aligned} u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}',\boldsymbol{\psi}') - u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi},\boldsymbol{\psi}) &\equiv u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}') - \dot{x}_i^{\mathrm{L,A}}(t) \\ &\leq u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}') - u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi}) \\ &\leq k\big(\|\boldsymbol{\phi}-\boldsymbol{\phi}'\|_\infty + \|\boldsymbol{\psi}-\boldsymbol{\psi}'\|_\infty\big). \end{aligned} \tag{3.5.7}$$

Combining (3.5.5), (3.5.6), and (3.5.7) shows that, for all $\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\phi}',\boldsymbol{\psi}' \in X^{\mathrm{A}}(t)$ such that $\phi_i' \geq \phi_i$,

$$u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}',\boldsymbol{\psi}') - u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi},\boldsymbol{\psi}) \leq k\big(\|\boldsymbol{\phi}-\boldsymbol{\phi}'\|_\infty + \|\boldsymbol{\psi}-\boldsymbol{\psi}'\|_\infty\big). \tag{3.5.8}$$

A similar argument shows that, for each $i \in \{1,...,n\}$, $t \in I\backslash\tilde{I}$, and $\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\phi}',\boldsymbol{\psi}' \in X^{\mathrm{A}}(t)$ such that $\psi_i \geq \psi_i'$,

$$o_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi},\boldsymbol{\psi}) - o_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}',\boldsymbol{\psi}') \leq k\big(\|\boldsymbol{\phi}-\boldsymbol{\phi}'\|_\infty + \|\boldsymbol{\psi}-\boldsymbol{\psi}'\|_\infty\big). \tag{3.5.9}$$

Since (3.5.8) and (3.5.9) hold, the functions $(\mathbf{u}_{\mathrm{r}}^{\mathrm{A}},\mathbf{o}_{\mathrm{r}}^{\mathrm{A}})$ (with $\mathbf{p}$ fixed) satisfy Condition II.2 in place of $(\mathbf{d}^{\mathrm{L,A}},\mathbf{d}^{\mathrm{U,A}})$.

Now, we show that the functions $(\mathbf{u}_{\mathrm{r}}^{\mathrm{A}},\mathbf{o}_{\mathrm{r}}^{\mathrm{A}},\mathbf{u}_{\mathrm{r}}^{\mathrm{B}},\mathbf{o}_{\mathrm{r}}^{\mathrm{B}})$ satisfy Condition II.3 in place of $(\mathbf{d}^{\mathrm{L,A}},\mathbf{d}^{\mathrm{U,A}},\mathbf{d}^{\mathrm{L,B}},\mathbf{d}^{\mathrm{U,B}})$. Consider any $i \in \{1,...,n_x\}$, any $t \in I\backslash\tilde{I}$, any $\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}} \in X^{\mathrm{A}}(t)$, and any $\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}} \in X^{\mathrm{B}}(t)$ such that $\boldsymbol{\phi}^{\mathrm{A}} \leq \boldsymbol{\phi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{A}}$ and $\phi_i^{\mathrm{A}} = \phi_i^{\mathrm{B}}$. We now consider several cases separately.

1. If $\phi_i^{\mathrm{A}} > x_i^{\mathrm{L,A}}(t)$ and $\phi_i^{\mathrm{B}} > x_i^{\mathrm{L,B}}(t)$, then

$$\begin{aligned} u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}}) &\equiv u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}}), \\ \text{and} \quad u_{\mathrm{r},i}^{\mathrm{B}}(t,\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}}) &\equiv u_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}}). \end{aligned}$$

   Since Condition III.2 holds, it follows that

$$u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}}) \leq u_{\mathrm{r},i}^{\mathrm{B}}(t,\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}}). \tag{3.5.10}$$

2. If $\phi_i^{\mathrm{A}} = x_i^{\mathrm{L,A}}(t)$ and $\phi_i^{\mathrm{B}} > x_i^{\mathrm{L,B}}(t)$, this implies that $x_i^{\mathrm{L,A}}(t) > x_i^{\mathrm{L,B}}(t)$ which contradicts the assumption $\mathbf{x}^{\mathrm{L,A}}(t) \leq \mathbf{x}^{\mathrm{L,B}}(t) \leq \mathbf{x}^{\mathrm{U,B}}(t) \leq \mathbf{x}^{\mathrm{U,A}}(t)$. Hence, this case does not occur.

3. If $\phi_i^{\mathrm{A}} > x_i^{\mathrm{L,A}}(t)$ and $\phi_i^{\mathrm{B}} = x_i^{\mathrm{L,B}}(t)$, then

$$\begin{aligned} u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}}) &\equiv u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}}), \\ \text{and} \quad u_{\mathrm{r},i}^{\mathrm{B}}(t,\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}}) &\equiv \max(\dot{x}_i^{\mathrm{L,B}}(t), u_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}})). \end{aligned}$$

   Since Condition III.2 holds, it follows that

$$u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}}) \leq u_{\mathrm{r},i}^{\mathrm{B}}(t,\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}}). \tag{3.5.11}$$

4. If $\phi_i^{\mathrm{A}} = x_i^{\mathrm{L,A}}(t)$ and $\phi_i^{\mathrm{B}} = x_i^{\mathrm{L,B}}(t)$, then

$$\begin{aligned} u_{\mathrm{r},i}^{\mathrm{A}}(t,\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}}) &\equiv \max(\dot{x}_i^{\mathrm{L,A}}(t), u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}^{\mathrm{A}},\boldsymbol{\psi}^{\mathrm{A}})), \\ \text{and} \quad u_{\mathrm{r},i}^{\mathrm{B}}(t,\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}}) &\equiv \max(\dot{x}_i^{\mathrm{L,B}}(t), u_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}^{\mathrm{B}},\boldsymbol{\psi}^{\mathrm{B}})), \end{aligned}$$

Since $\phi_i^{\mathrm{A}} = \phi_i^{\mathrm{B}}$ by assumption, it follows that $x_i^{\mathrm{L,A}}(t) = x_i^{\mathrm{L,B}}(t)$. Moreover, since $\mathbf{x}^{\mathrm{L,A}}(\tau) \leq \mathbf{x}^{\mathrm{L,B}}(\tau)$ for each $\tau \in I$ by assumption and since $t_f \in \tilde{I}$ so $t \neq t_f$, it follows that $\dot{x}_i^{\mathrm{L,A}}(t) \leq \dot{x}_i^{\mathrm{L,B}}(t)$. Combining this with Condition III.2 yields

$$u_{\mathrm{r},i}^{\mathrm{A}}(t, \boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}}) \leq u_{\mathrm{r},i}^{\mathrm{B}}(t, \boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}}). \tag{3.5.12}$$

Combining (3.5.10), (3.5.11), and (3.5.12) shows that, for all $\boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}} \in X^{\mathrm{A}}(t)$ and $\boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}} \in X^{\mathrm{B}}(t)$ such that $\boldsymbol{\phi}^{\mathrm{A}} \leq \boldsymbol{\phi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{A}}$ and $\phi_i^{\mathrm{A}} = \phi_i^{\mathrm{B}}$,

$$u_{\mathrm{r},i}^{\mathrm{A}}(t, \boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}}) \leq u_{\mathrm{r},i}^{\mathrm{B}}(t, \boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}}).$$

A similar argument shows that, for all $\boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}} \in X^{\mathrm{A}}(t)$ and $\boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}} \in X^{\mathrm{B}}(t)$ such that $\boldsymbol{\phi}^{\mathrm{A}} \leq \boldsymbol{\phi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{A}}$ and $\psi_i^{\mathrm{A}} = \psi_i^{\mathrm{B}}$,

$$o_{\mathrm{r},i}^{\mathrm{A}}(t, \boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}}) \geq o_{\mathrm{r},i}^{\mathrm{B}}(t, \boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}}).$$

Thus, the functions $(\mathbf{u}_{\mathrm{r}}^{\mathrm{A}}, \mathbf{o}_{\mathrm{r}}^{\mathrm{A}}, \mathbf{u}_{\mathrm{r}}^{\mathrm{B}}, \mathbf{o}_{\mathrm{r}}^{\mathrm{B}})$ (with $\mathbf{p}$ fixed) satisfy Condition II.3 in place of $(\mathbf{d}^{\mathrm{L,A}}, \mathbf{d}^{\mathrm{U,A}}, \mathbf{d}^{\mathrm{L,B}}, \mathbf{d}^{\mathrm{U,B}})$.

Next, since it has been shown at the beginning that both $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ are valid state relaxations for (2.4.1) on $I \times P$, it follows that for all $t \in I$,

$$\begin{aligned} &\mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}) \leq \mathbf{x}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p}) \\ \text{and} \quad &\mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}) \leq \mathbf{x}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p}). \end{aligned}$$

Thus, $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}}, \mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ satisfy Condition II.4 in place of $(\mathbf{v}^{\mathrm{A}}, \mathbf{w}^{\mathrm{A}}, \mathbf{v}^{\mathrm{B}}, \mathbf{w}^{\mathrm{B}})$.

Finally, it is observed that

$$\max(x_i^{\mathrm{L,A}}(t_0), x_{0,i}^{\mathrm{cv,A}}(\mathbf{p})) \leq \max(x_i^{\mathrm{L,B}}(t_0), x_{0,i}^{\mathrm{cv,B}}(\mathbf{p})) \leq x_{0,i}(\mathbf{p})$$
$$\leq \min(x_i^{\mathrm{U,B}}(t_0), x_{0,i}^{\mathrm{cc,B}}(\mathbf{p})) \leq \min(x_i^{\mathrm{U,A}}(t_0), x_{0,i}^{\mathrm{cc,A}}(\mathbf{p})),$$

and observe that $(\mathbf{x}^{\mathrm{cv,A}}(\cdot,\mathbf{p}), \mathbf{x}^{\mathrm{cc,A}}(\cdot,\mathbf{p}), \mathbf{x}^{\mathrm{cv,B}}(\cdot,\mathbf{p}), \mathbf{x}^{\mathrm{cc,B}}(\cdot,\mathbf{p}))$ are the C-solutions on I of the following ODEs. For each $i \in \{1,...,n_x\}$, for all $t \in (t_0, t_f]$,

$$\begin{aligned}
\dot{x}_i^{\mathrm{cv,A}}(t,\mathbf{p}) &= u_{\mathrm{r},i}^{\mathrm{A}}(t,\mathbf{p},\mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p})), \quad & x_i^{\mathrm{cv,A}}(t_0,\mathbf{p}) &= \max(x_i^{\mathrm{L,A}}(t_0), x_{0,i}^{\mathrm{cv,A}}(\mathbf{p})),\\
\dot{x}_i^{\mathrm{cc,A}}(t,\mathbf{p}) &= o_{\mathrm{r},i}^{\mathrm{A}}(t,\mathbf{p},\mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p})), \quad & x_i^{\mathrm{cc,A}}(t_0,\mathbf{p}) &= \min(x_i^{\mathrm{U,A}}(t_0), x_{0,i}^{\mathrm{cc,A}}(\mathbf{p})),\\
\dot{x}_i^{\mathrm{cv,B}}(t,\mathbf{p}) &= u_{\mathrm{r},i}^{\mathrm{B}}(t,\mathbf{p},\mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p})), \quad & x_i^{\mathrm{cv,B}}(t_0,\mathbf{p}) &= \max(x_i^{\mathrm{L,B}}(t_0), x_{0,i}^{\mathrm{cv,B}}(\mathbf{p})),\\
\dot{x}_i^{\mathrm{cc,B}}(t,\mathbf{p}) &= o_{\mathrm{r},i}^{\mathrm{B}}(t,\mathbf{p},\mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p})), \quad & x_i^{\mathrm{cc,B}}(t_0,\mathbf{p}) &= \min(x_i^{\mathrm{U,B}}(t_0), x_{0,i}^{\mathrm{cc,B}}(\mathbf{p})),
\end{aligned}$$

Thus, Condition II.5 is satisfied with $(\mathbf{u}_{\mathrm{r}}^{\mathrm{A}}, \mathbf{o}_{\mathrm{r}}^{\mathrm{A}}, \mathbf{u}_{\mathrm{r}}^{\mathrm{B}}, \mathbf{o}_{\mathrm{r}}^{\mathrm{B}})$ with $\mathbf{p}$ fixed in place of $(\mathbf{d}^{\mathrm{L,A}}, \mathbf{d}^{\mathrm{U,A}}, \mathbf{d}^{\mathrm{L,B}}, \mathbf{d}^{\mathrm{U,B}})$. Hence, Theorem 3.4.1 applies with the substitutions (3.5.2). □

Remark 3.5.2. The theorem above will be applied to show that the state relaxation method in [3] that will be introduced in the next chapter has the following tightness properties:

1. If McCormick relaxations [5] of $\mathbf{f}$ are applied in this method, then the resulting state relaxations are guaranteed to lie within the SBM relaxations [2].

2. Embedding tighter relaxations for $\mathbf{f}$ in this method will necessarily lead to state relaxations that are at least as tight.

As mentioned in Section 2.4, the Scott–Barton right-hand side functions $(\mathbf{u}, \mathbf{o})$ in (2.4.1) may be constructed via (2.4.2) where the functions $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ satisfy Assumption 2.4.11. The following corollary provides sufficient conditions under which one method for furnishing

$(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ ultimately leads to state relaxations $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$ that lie within the relaxations by another method, through (2.4.2) and (2.4.1).

Corollary 3.5.3. Consider functions $(\mathbf{x}^{\mathrm{L,A}},\mathbf{x}^{\mathrm{L,B}},\mathbf{x}^{\mathrm{U,A}},\mathbf{x}^{\mathrm{U,B}})$ and $(\mathbf{x}_0^{\mathrm{cv,A}},\mathbf{x}_0^{\mathrm{cv,B}},\mathbf{x}_0^{\mathrm{cc,A}},\mathbf{x}_0^{\mathrm{cc,B}})$ and intervals $X^{\mathrm{A}}(t)$ and $X^{\mathrm{B}}(t)$ as in Theorem 3.5.1. Consider functions $\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}},\tilde{\mathbf{u}}^{\mathrm{A}},\tilde{\mathbf{o}}^{\mathrm{A}},\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}},\tilde{\mathbf{u}}^{\mathrm{B}},\tilde{\mathbf{o}}^{\mathrm{B}}:$ $I\times P\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$ for which the following conditions hold.

IV.1 Assumption 2.4.11 is satisfied with $(\tilde{\mathbf{u}},\tilde{\mathbf{o}}):=(\tilde{\mathbf{u}}^{\mathrm{A}},\tilde{\mathbf{o}}^{\mathrm{A}})$ and $X:=X^{\mathrm{A}}$, and with $(\tilde{\mathbf{u}},\tilde{\mathbf{o}}):=(\tilde{\mathbf{u}}^{\mathrm{B}},\tilde{\mathbf{o}}^{\mathrm{B}})$ and $X:=X^{\mathrm{B}}$.

IV.2 For any $i\in\{1,\ldots,n_x\}$, a.e. $t\in I$, any $\mathbf{p}\in P$, any $\boldsymbol{\phi},\boldsymbol{\psi}\in X^{\mathrm{A}}(t)$, and any $\boldsymbol{\phi}',\boldsymbol{\psi}'\in X^{\mathrm{B}}(t)$ such that $\boldsymbol{\phi}\le\boldsymbol{\phi}'\le\boldsymbol{\psi}'\le\boldsymbol{\psi}$ and $\phi_i=\phi_i'=\psi_i'=\psi_i$,

$$\begin{aligned} &\tilde{u}_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi})\le\tilde{u}_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}'),\\ \text{and}\quad &\tilde{o}_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi})\ge\tilde{o}_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}'). \end{aligned}\tag{3.5.13}$$

IV.3 (2.4.2) holds with $(\mathbf{u},\mathbf{o},\tilde{\mathbf{u}},\tilde{\mathbf{o}}):=(\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}},\tilde{\mathbf{u}}^{\mathrm{A}},\tilde{\mathbf{o}}^{\mathrm{A}})$ and with $(\mathbf{u},\mathbf{o},\tilde{\mathbf{u}},\tilde{\mathbf{o}}):=(\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}},\tilde{\mathbf{u}}^{\mathrm{B}},\tilde{\mathbf{o}}^{\mathrm{B}})$.

Let $(\mathbf{x}^{\mathrm{cv,A}},\mathbf{x}^{\mathrm{cc,A}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}},\mathbf{x}_0^{\mathrm{cv}},\mathbf{x}_0^{\mathrm{cc}},\mathbf{u},\mathbf{o}):=(\mathbf{x}^{\mathrm{L,A}},\mathbf{x}^{\mathrm{U,A}},\mathbf{x}_0^{\mathrm{cv,A}},\mathbf{x}_0^{\mathrm{cc,A}},\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}}).$$

Let $(\mathbf{x}^{\mathrm{cv,B}},\mathbf{x}^{\mathrm{cc,B}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}},\mathbf{x}_0^{\mathrm{cv}},\mathbf{x}_0^{\mathrm{cc}},\mathbf{u},\mathbf{o}):=(\mathbf{x}^{\mathrm{L,B}},\mathbf{x}^{\mathrm{U,B}},\mathbf{x}_0^{\mathrm{cv,B}},\mathbf{x}_0^{\mathrm{cc,B}},\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}}).$$

Then, the functions $(\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}})$ and $(\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}})$ are Scott–Barton right-hand side functions, and satisfy Condition III.2 in Theorem 3.5.1. The solutions $(\mathbf{x}^{\mathrm{cv,A}},\mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}},\mathbf{x}^{\mathrm{cc,B}})$ are both valid state relaxations for (2.3.1). Moreover, for any $(t, \mathbf{p}) \in I \times P$,

$$\mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p}).$$

Proof. We proceed by showing that the functions $(\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$ satisfy Conditions III.1 and III.2 of Theorem 3.5.1. Then, the claimed results follow from Theorem 3.5.1.

Consider any $i \in \{1,...,n_x\}$, any $\mathbf{p} \in P$, and a.e. $t \in I$. Consider any $\boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}} \in X^{\mathrm{A}}(t)$ and $\boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}} \in X^{\mathrm{B}}(t)$ for which $\boldsymbol{\phi}^{\mathrm{A}} \leq \boldsymbol{\phi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{A}}$ and $\phi_i^{\mathrm{A}} = \phi_i^{\mathrm{B}}$. Let $(\boldsymbol{\phi}, \boldsymbol{\psi}) := \mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}})$ and $(\boldsymbol{\phi}', \boldsymbol{\psi}') := \mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}})$. By construction of $\mathbf{r}^{i,\mathrm{L}}$, $\phi_i = \psi_i = \phi_i^{\mathrm{A}}$, and $\phi_i' = \psi_i' = \phi_i^{\mathrm{B}}$. Moreover, since $\phi_i^{\mathrm{A}} = \phi_i^{\mathrm{B}}$, it follows that

$$\phi_i = \phi_i' = \psi_i' = \psi_i.$$

Also by construction of $\mathbf{r}^{i,\mathrm{L}}$, for all $\kappa \in \{1,...,n\}$ and $\kappa \neq i$, $\phi_\kappa = \phi_\kappa^{\mathrm{A}}$, $\psi_\kappa = \psi_\kappa^{\mathrm{A}}$, $\phi_\kappa' = \phi_\kappa^{\mathrm{B}}$, and $\psi_\kappa' = \psi_\kappa^{\mathrm{B}}$, and thus, $\phi_\kappa \leq \phi_\kappa' \leq \psi_\kappa' \leq \psi_\kappa$, $\boldsymbol{\phi}, \boldsymbol{\psi} \in X^{\mathrm{A}}(t)$, and $\boldsymbol{\phi}', \boldsymbol{\psi}' \in X^{\mathrm{B}}(t)$. Then, since (3.5.13) holds, it follows that

$$\tilde{u}_i^{\mathrm{A}}(t, \mathbf{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \leq \tilde{u}_i^{\mathrm{B}}(t, \mathbf{p}, \boldsymbol{\phi}', \boldsymbol{\psi}'),$$

which implies that

$$\tilde{u}_i^{\mathrm{A}}(t, \mathbf{p}, \mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}})) \leq \tilde{u}_i^{\mathrm{B}}(t, \mathbf{p}, \mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}})).$$

Combining this with Condition IV.3 yields

$$u_i^{\mathrm{A}}(t, \mathbf{p}, \boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}}) \leq u_i^{\mathrm{B}}(t, \mathbf{p}, \boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}}).$$

A similar argument shows that, for any $\boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}} \in X^{\mathrm{A}}(t)$ and any $\boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}} \in X^{\mathrm{B}}(t)$ such that $\boldsymbol{\phi}^{\mathrm{A}} \leq \boldsymbol{\phi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{B}} \leq \boldsymbol{\psi}^{\mathrm{A}}$ and $\psi_i^{\mathrm{A}} = \psi_i^{\mathrm{B}}$,

$$o_i^{\mathrm{A}}(t, \mathbf{p}, \boldsymbol{\phi}^{\mathrm{A}}, \boldsymbol{\psi}^{\mathrm{A}}) \geq o_i^{\mathrm{B}}(t, \mathbf{p}, \boldsymbol{\phi}^{\mathrm{B}}, \boldsymbol{\psi}^{\mathrm{B}}).$$

Thus, the functions $(\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$ satisfy Condition III.2 in Theorem 3.5.1. Moreover, since Condition IV.1 holds, [2, Lemma 10 and Lemma 11] imply that the functions $(\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}})$ and $(\mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$ are valid Scott–Barton right-hand side functions based on $X^{\mathrm{A}}(t)$ and $X^{\mathrm{B}}(t)$, respectively. Thus, Condition III.1 in Theorem 3.5.1 holds, and the claimed results follow from Theorem 3.5.1. □

As shown in [2], if valid state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ for (2.3.1) are considered, and if we have functions $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ as in Assumption 2.4.11, then for a.e. $t \in I$, the mappings $\tilde{\mathbf{u}}(t, \cdot, \mathbf{x}^{\mathrm{cv}}(t, \cdot), \mathbf{x}^{\mathrm{cc}}(t, \cdot))$ and $\tilde{\mathbf{o}}(t, \cdot, \mathbf{x}^{\mathrm{cv}}(t, \cdot), \mathbf{x}^{\mathrm{cc}}(t, \cdot))$ are respectively convex and concave relaxations of the composition $\mathbf{f}(t, \cdot, \mathbf{x}(t, \cdot))$ in (2.3.1) on P. Such functions $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ are called *relaxation functions* for $\mathbf{f}$ in [79]. The following theorem shows that under both Assumptions 2.4.11 and 2.4.12, tighter relaxation functions $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ will necessarily ultimately translate into state relaxations that are at least as tight through (2.4.2) and (2.4.1).

Theorem 3.5.4. Consider functions $(\mathbf{x}^{\mathrm{L,A}}, \mathbf{x}^{\mathrm{L,B}}, \mathbf{x}^{\mathrm{U,A}}, \mathbf{x}^{\mathrm{U,B}})$ and $(\mathbf{x}_0^{\mathrm{cv,A}}, \mathbf{x}_0^{\mathrm{cv,B}}, \mathbf{x}_0^{\mathrm{cc,A}}, \mathbf{x}_0^{\mathrm{cc,B}})$ and intervals $X^{\mathrm{A}}(t)$ and $X^{\mathrm{B}}(t)$ as in Theorem 3.5.1. Consider functions $\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}, \tilde{\mathbf{u}}^{\mathrm{A}}, \tilde{\mathbf{o}}^{\mathrm{A}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}}, \tilde{\mathbf{u}}^{\mathrm{B}}, \tilde{\mathbf{o}}^{\mathrm{B}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ for which the following conditions hold:

V.1 Assumption 2.4.11 is satisfied with $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}}) := (\tilde{\mathbf{u}}^{\mathrm{A}}, \tilde{\mathbf{o}}^{\mathrm{A}})$ and $X := X^{\mathrm{A}}$, and with $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}}) := (\tilde{\mathbf{u}}^{\mathrm{B}}, \tilde{\mathbf{o}}^{\mathrm{B}})$ and $X := X^{\mathrm{B}}$,

V.2 Assumption 2.4.12 is satisfied with $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}}) := (\tilde{\mathbf{u}}^{\mathrm{A}}, \tilde{\mathbf{o}}^{\mathrm{A}})$ and $X := X^{\mathrm{A}}$, V.3 for a.e. $t \in I$, any $\mathbf{p} \in P$, and any $\boldsymbol{\phi}', \boldsymbol{\psi}' \in X^{\mathrm{B}}(t)$,

$$\begin{aligned} \tilde{\mathbf{u}}^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}') &\leq \tilde{\mathbf{u}}^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}'), \\ \text{and} \quad \tilde{\mathbf{o}}^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}') &\geq \tilde{\mathbf{o}}^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}'), \end{aligned} \tag{3.5.14}$$

V.4 (2.4.2) holds with $(\mathbf{u},\mathbf{o},\tilde{\mathbf{u}},\tilde{\mathbf{o}}) := (\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}},\tilde{\mathbf{u}}^{\mathrm{A}},\tilde{\mathbf{o}}^{\mathrm{A}})$ and with $(\mathbf{u},\mathbf{o},\tilde{\mathbf{u}},\tilde{\mathbf{o}}) := (\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}},\tilde{\mathbf{u}}^{\mathrm{B}},\tilde{\mathbf{o}}^{\mathrm{B}})$.

Then, the functions $(\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}})$ and $(\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}})$ are Scott–Barton right-hand side functions, and satisfy Condition III.2 in Theorem 3.5.1. The solutions $(\mathbf{x}^{\mathrm{cv,A}},\mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}},\mathbf{x}^{\mathrm{cc,B}})$ are both valid state relaxations for (2.3.1). Moreover, for any $(t,\mathbf{p}) \in I \times P$,

$$\mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p}).$$

Proof. Noting that Condition IV.1 of Corollary 3.5.3 is satisfied, we proceed by showing that under Conditions V.2 and V.3, the functions $(\tilde{\mathbf{u}}^{\mathrm{A}},\tilde{\mathbf{o}}^{\mathrm{A}},\tilde{\mathbf{u}}^{\mathrm{B}},\tilde{\mathbf{o}}^{\mathrm{B}})$ satisfy Condition IV.2 of Corollary 3.5.3. Then, the claimed results follow from Corollary 3.5.3.

Consider a.e. $t \in I$, any $\mathbf{p} \in P$, any $\boldsymbol{\phi},\boldsymbol{\psi} \in X^{\mathrm{A}}(t)$, and any $\boldsymbol{\phi}',\boldsymbol{\psi}' \in X^{\mathrm{B}}(t)$ such that $\boldsymbol{\phi} \leq \boldsymbol{\phi}' \leq \boldsymbol{\psi}' \leq \boldsymbol{\psi}$. Since Condition V.2 holds and $X^{\mathrm{B}}(t) \subseteq X^{\mathrm{A}}(t)$,

$$\tilde{\mathbf{u}}^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi}) \leq \tilde{\mathbf{u}}^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}') \quad \text{and} \quad \tilde{\mathbf{o}}^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi}) \geq \tilde{\mathbf{o}}^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}').$$

Moreover, since (3.5.14) holds,

$$\begin{aligned} \tilde{\mathbf{u}}^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi}) &\leq \tilde{\mathbf{u}}^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}') \leq \tilde{\mathbf{u}}^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}'), \\ \text{and} \quad \tilde{\mathbf{o}}^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi},\boldsymbol{\psi}) &\geq \tilde{\mathbf{o}}^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}') \geq \tilde{\mathbf{o}}^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\phi}',\boldsymbol{\psi}'). \end{aligned}$$

Thus, Condition IV.2 holds. □

3.6 Conclusions and future work

New comparison results have been provided for ODE systems with Carathéodory solutions. These new results have less stringent requirements than established results and are thus more broadly applicable. Theorem 3.3.1 provides sufficient conditions under which one ODE system's solutions dominate another's. Unlike certain established results, Theorem 3.3.1 does not require differentiability of ODE solutions or the quasi-monotonicity assumption on right-hand side functions, and thus is desirable for dynamic models of engineering processes. Based on Theorem 3.3.1, Theorem 3.4.1 provides sufficient conditions under which one dynamic bounding pair necessarily encloses another. This result has less stringent assumptions than the only established bound comparison result [79, Theorem 4.14], and is useful for comparing competing reachable-set generation methods in terms of tightness. By applying Theorem 3.4.1, it was shown in Section 3.5 that a state-of-the-art framework (2.4.1) [2] for generating convex relaxations for solutions of a nonconvex parametric ODE system (2.3.1) has the following tightness property: if tighter initial relaxations $(\mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}})$, tighter state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$, and tighter right-hand side relaxations $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ are available, then these relaxations in (2.4.1) and (2.4.2) will necessarily translate into state relaxations that are at least as tight for the original ODE solutions. This result is beneficial for developing new methods for formulating tighter ODE relaxations, aiding in furnishing tighter bounding for deterministic global dynamic optimization and constructing tighter convex enclosures of reachable sets. In particular, Theorem 3.5.1 is applied to show the desirable tightness properties of a new ODE relaxation formulation proposed in [3], as will be seen in the next chapter.

Future work may involve seeking new relaxation methods for furnishing tighter $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ in Assumption 2.4.11, since doing so will always yield tighter ODE relaxations through

(2.4.2) and (2.4.1), as shown in Theorem 3.5.4. Extending the ODE-based results of this chapter to systems of differential-algebraic equations may also be possible.

Chapter 4

Optimization-Based Convex Relaxations for Nonconvex Parametric Systems of ODEs

This chapter, reproduced from the published journal article [3], proposes novel convex and concave relaxations for the solutions of parametric ODE systems in the Scott–Barton framework [2], to aid in furnishing bounding information for deterministic global dynamic optimization methods. These relaxations are constructed as the solutions of auxiliary ODE systems with embedded convex optimization problems, whose objective functions employ convex and concave relaxations of the original ODE right-hand side. Unlike established relaxation methods, any valid convex and concave relaxations for the original right-hand side are permitted in the approach, including the McCormick relaxations [5] and the αBB relaxations [9, 49]. By applying the tightness results of the Scott–Barton framework developed in Chapter 3, it is shown that tighter such relaxations will necessarily translate into at least as tight relaxations for the ODE solutions, thus providing tighter bounding information for an overarching global dynamic optimization method. Notably, if McCormick relaxations are employed in the new approach, then the obtained relaxations are guaranteed to be at least as tight as the state-of-the-art SBM relaxations [2] based on generalized McCormick relaxations. The new relaxations converge rapidly to the original system as the considered parametric subdomain shrinks. Several examples are presented for comparison with established ODE relaxations, based on a proof-of-concept implementation in MATLAB. In a global optimization case study, the new ODE relaxations are shown to lead to fewer branch-and-bound global optimization iterations than the SBM relaxations.

4.1 Introduction

This chapter considers the generic nonconvex dynamic optimization problem (1.1.1) with embedded the underlying ODE system (1.1.2) that is formalized in Section 2.3. As introduced in Section 1.1, state relaxations (as in Definition 2.4.1) are fundamental in deterministic algorithms of global dynamic optimization. Tighter state relaxations will necessarily translate into tighter convex relaxations of the objective function c in (1.1.1), thus providing tighter lower bounds in global optimization. Such tightness may improve computational efficiency of an overarching global optimization method, by reducing the number of iterations required by deterministic branch-and-bound algorithms [27–29, 31]. Since current deterministic algorithms for global dynamic optimization can only solve problems of modest size, improved techniques must be sought for computing tighter state relaxations efficiently, to ultimately extend the scope of these algorithms to problems of practical interest.

Established approaches for constructing either state relaxations or **p**-invariant *state bounds* for **x** in (1.1.2) may be classified into two broad categories: *discretize-then-relax* approaches and *relax-then-discretize* approaches. The discretize-then-relax approaches [35, 57–59, 62–64, 66, 67] first discretize the original ODE system (1.1.2) using methods such as Euler method, orthogonal collocation, and Taylor expansion, and then the resulting equation system is bounded by different approaches. The relax-then-discretize approaches [1, 2, 69–71, 73–76, 78, 108, 112], on the other hand, directly handle the original dynamic system, and typically compute valid state relaxations by constructing and solving an auxiliary ODE system, whose right-hand side is derived from various bounds or relaxations of the original right-hand side **f**. Compared to the discretize-then-relax approaches, the relax-then-discretize approaches have the advantage that they are able to exploit the adaptive time-stepping and error control of numerical ODE solvers. Refer to Section 1.3 for a thorough review of established state relaxation approaches.

This chapter proposes a new relax-then-discretize approach for relaxing the ODEs (1.1.2), in which the right-hand side of the relaxations' ODE system includes embedded convex optimization problems. These optimization problems employ bound constraints and convex and concave relaxations for the right-hand side **f**. This approach is based on the Scott–Barton ODE relaxation framework [2] as summarized in Section 2.4, but constructs very different auxiliary right-hand side functions from [2] to satisfy the framework's requirements. While the Scott–Barton method [2] is based on generalized McCormick relaxations of the original right-hand side **f**, our new formulation is compatible with any valid convex and concave relaxations for **f**, including affine relaxations, the αBB relaxations, the McCormick relaxations, or even the pointwise tightest among multiple relaxations. Moreover, tighter such relaxations of **f** will always yield tighter state relaxations for (1.1.2), which incentivizes seeking tighter relaxations for closed-form functions to help relax dynamic systems. Furthermore, if McCormick relaxations of $\mathbf{f}$ are applied, then the new state relaxations are at least as tight as the SBM relaxations [2]. Numerical examples also suggest that when αBB relaxations [9] are used for $\mathbf{f}$ in the new approach, significantly tighter relaxations may be obtained compared to the primary established method for generating αBB state relaxations [1]. Moreover, the new state relaxations inherit second-order pointwise convergence from the supplied relaxations of $\mathbf{f}$, and thus help to avoid the cluster effect in branch-and-bound-based algorithms.

We note that several established dynamic bounding methods also employ embedded optimization problems, such as the methods in [72, 76], which are based on earlier theoretical results [74, 77, 89] involving differential inequalities. In particular, our formulation becomes similar to a formulation by Singer and Barton [76] in the special case where we adopt affine relaxations of convex relaxations of $\mathbf{f}$; their method's justification relies heavily on these relaxations being affine. The right-hand side convex optimization problems in our new approach also resemble those in an approach for constructing state bounds by Harwood et al. [72]. Both our approach and the approach of [72] require furnishing convex relaxations of $\mathbf{f}$ as objective functions, but have different decision variables and different goals (computing state relaxations versus computing state bounds). Neither approach appears to be a special case of the other.

We also note that the right-hand sides of our new auxiliary ODE system superficially resemble multivariate McCormick relaxations [7, Theorem 2] of the composition $\mathbf{f}(t,\cdot,\mathbf{x}(t,\cdot))$ combined with interval flattening operations [2, Definition 11]. However, as will be discussed in Section 4.2, our result is not a direct consequence of [7, Theorem 2]. Since we do not know *a priori* that our relaxations are valid and convex, we cannot satisfy the assumptions of [7, Theorem 2] when proving our relaxations' validity.

The remainder of this chapter is organized as follows. Section 4.2 formulates the new approach based on the Scott–Barton framework summarized in Section 2.4. Section 4.3 establishes its useful theoretical properties, including continuity of the new auxiliary right-hand side, existence and uniqueness of solutions, bounding, convexity, tightness, and convergence properties. The approach of this chapter is also compared to the following established approaches: the SBM approach [2], the primary dynamic αBB relaxation approach [1], and the Auxiliary Variable Method [50]. In Section 4.4, a proof-of-concept implementation of the new approach in MATLAB [113] is outlined, and several numerical examples are described for comparison with the established approaches. A problem instance of (1.1.1) is solved to global optimality in Julia [95], both with the new state relaxations and the SBM relaxations. In this instance, our new ODE relaxations require significantly fewer iterations in branch-and-bound.

4.2 New state relaxation formulation

This section presents a new approach for constructing useful state relaxations for the ODE process model (2.3.1) formalized in Section 2.3, to ultimately furnish bounding information to help solve the nonconvex optimization problem (1.1.1) to global optimality. Beneficial properties of this formulation will then be established in Section 4.3. This approach utilizes the Scott–Barton framework (2.4.1), but constructs new $\mathbf{u}$ and $\mathbf{o}$ functions as optimal-value functions (in the sense of e.g. [114]) for parametric convex optimization problems. The new approach requires functions $\mathbf{f}^{\mathrm{cv}}$ and $\mathbf{f}^{\mathrm{cc}}$ that satisfy the following assumption. Constructive methods to satisfy this assumption are discussed subsequently.

Assumption 4.2.1. Suppose that functions $\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}} : I \times P \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ satisfy all of the following conditions:

1. $\mathbf{f}^{\mathrm{cv}}$ and $\mathbf{f}^{\mathrm{cc}}$ are continuous,

2. $\mathbf{f}^{\mathrm{cv}}(t, \mathbf{p}, \cdot)$ and $\mathbf{f}^{\mathrm{cc}}(t, \mathbf{p}, \cdot)$ are Lipschitz continuous on $\mathbb{R}^{n_x}$, uniformly over $(t, \mathbf{p}) \in I \times P$, and

3. for a.e. $t \in I$, the functions $\mathbf{f}^{\mathrm{cv}}(t, \cdot, \cdot)$ and $\mathbf{f}^{\mathrm{cc}}(t, \cdot, \cdot)$ are, respectively, convex and concave relaxations of $\mathbf{f}(t, \cdot, \cdot)$ in (2.3.1) on $P \times X(t)$.

Remark 4.2.2. Several established convex relaxation approaches produce functions $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ that satisfy Assumption 4.2.1, including certain affine relaxations, the αBB relaxations [9], the McCormick relaxations [5], or the pointwise tightest among certain multiple relaxations defined as follows. Suppose that multiple convex relaxations $\mathbf{f}^{\mathrm{cv},1}, \mathbf{f}^{\mathrm{cv},2}, ..., \mathbf{f}^{\mathrm{cv},k}$ and concave relaxations $\mathbf{f}^{\mathrm{cc},1}, \mathbf{f}^{\mathrm{cc},2}, ..., \mathbf{f}^{\mathrm{cc},\kappa}$ that satisfy Assumption 4.2.1 are available. For each $i \in \{1, ..., n_x\}$, pointwise tightest relaxations $(f_i^{\mathrm{cv,multi}}, f_i^{\mathrm{cc,multi}})$ are then defined so that for each $t \in I$, $\mathbf{p} \in P$, and $\boldsymbol{\xi} \in X(t)$,

$$\begin{aligned} f_i^{\mathrm{cv,multi}}(t, \mathbf{p}, \boldsymbol{\xi}) &:= \max(f_i^{\mathrm{cv},1}(t, \mathbf{p}, \boldsymbol{\xi}), f_i^{\mathrm{cv},2}(t, \mathbf{p}, \boldsymbol{\xi}), ..., f_i^{\mathrm{cv},k}(t, \mathbf{p}, \boldsymbol{\xi})) \\ \text{and} \quad f_i^{\mathrm{cc,multi}}(t, \mathbf{p}, \boldsymbol{\xi}) &:= \min(f_i^{\mathrm{cc},1}(t, \mathbf{p}, \boldsymbol{\xi}), f_i^{\mathrm{cc},2}(t, \mathbf{p}, \boldsymbol{\xi}), ..., f_i^{\mathrm{cc},\kappa}(t, \mathbf{p}, \boldsymbol{\xi})). \end{aligned} \tag{4.2.1}$$

These relaxations are readily confirmed to satisfy Assumption 4.2.1. Lipschitz continuity on the full space $\mathbb{R}^{n_x}$ may be enforced by passing to an appropriate Lipschitz extension. Although convex/concave envelopes (the tightest possible relaxations) of $\mathbf{f}(t, \cdot, \cdot)$ are not typically available, they may be used here as well in principle. This approach will be illustrated in Example 4.5.

With Assumption 4.2.1 satisfied, define a function $\mathbf{v} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ so that for all $i \in \{1, ..., n_x\}$ and $\boldsymbol{\alpha}, \boldsymbol{\xi}^{\text{cv}}, \boldsymbol{\xi}^{\text{cc}} \in \mathbb{R}^{n_x}$,

$$v_i(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\text{cv}}, \boldsymbol{\xi}^{\text{cc}}) := \tfrac{1}{2}[(\alpha_i + 1)\xi_i^{\text{cc}} - (\alpha_i - 1)\xi_i^{\text{cv}}]. \tag{4.2.2}$$

Intuitively, $v_i(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\text{cv}}, \boldsymbol{\xi}^{\text{cc}})$ is a linear combination of $\boldsymbol{\xi}^{\text{cv}}$ and $\boldsymbol{\xi}^{\text{cc}}$, weighted in a particular way based on the value of $\boldsymbol{\alpha}$. The key step of our formulation is the following. Construct functions $\mathbf{u}$ and $\mathbf{o}$, for use in the Scott–Barton framework (2.4.1), so that for all $i \in \{1, ..., n_x\}$ and $(t, \mathbf{p}, \boldsymbol{\xi}^{\text{cv}}, \boldsymbol{\xi}^{\text{cc}}) \in I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$,

$$\begin{aligned} & u_i(t, \mathbf{p}, \boldsymbol{\xi}^{\text{cv}}, \boldsymbol{\xi}^{\text{cc}}) := \min_{\boldsymbol{\alpha} \in [-1,1]^{n_x}} f_i^{\text{cv}}(t, \mathbf{p}, \mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\text{cv}}, \boldsymbol{\xi}^{\text{cc}})) \text{ subject to } \alpha_i = -1, \\ \text{and} \quad & o_i(t, \mathbf{p}, \boldsymbol{\xi}^{\text{cv}}, \boldsymbol{\xi}^{\text{cc}}) := \max_{\boldsymbol{\alpha} \in [-1,1]^{n_x}} f_i^{\text{cc}}(t, \mathbf{p}, \mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\text{cv}}, \boldsymbol{\xi}^{\text{cc}})) \text{ subject to } \alpha_i = +1. \end{aligned} \tag{4.2.3}$$

We will show in Section 4.3 below that this new choice of $(\mathbf{u}, \mathbf{o})$ always yields valid state relaxations, and that these are tighter than the SBM relaxations when McCormick relaxations of $\mathbf{f}$ are used to construct $(\mathbf{f}^{\text{cv}}, \mathbf{f}^{\text{cc}})$. Observe that the nonlinear programs defining u_i and o_i above are, respectively, bound-constrained convex minimization and concave maximization problems that are always feasible. Overall, our new dynamic relaxation approach is to solve the auxiliary ODE system (2.4.1) with $\mathbf{u}$ and $\mathbf{o}$ defined in (4.2.3) to yield valid state relaxations for (2.3.1); this validity will be established in Section 4.3.

Remark 4.2.3. If the pointwise tightest relaxations $(\mathbf{f}^{\text{cv,multi}}, \mathbf{f}^{\text{cc,multi}})$ considered in Remark 4.2.2 are applied in (4.2.3), then the nonsmooth "max" and "min" functions in (4.2.1) may be eliminated in (4.2.3) by introducing an extra decision variable, as follows:

$$
\begin{aligned}
u_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \min_{s\in\mathbb{R},\boldsymbol{\alpha}\in[-1,1]^{n_x}} \quad & s \\
\text{subject to} \quad & f_i^{\mathrm{cv},m}(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})) \le s, \quad \forall m\in\{1,\dots,k\}, \\
& \alpha_i = -1, \\
o_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \max_{s\in\mathbb{R},\boldsymbol{\alpha}\in[-1,1]^{n_x}} \quad & s \\
\text{subject to} \quad & f_i^{\mathrm{cc},m}(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})) \ge s, \quad \forall m\in\{1,\dots,\kappa\}, \\
& \alpha_i = +1.
\end{aligned}
$$

In the special case where $\boldsymbol{\xi}^{\mathrm{cv}} \le \boldsymbol{\xi}^{\mathrm{cc}}$ in (4.2.3), observe that (4.2.3) may be reformulated as

$$
\begin{aligned}
& u_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \min_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cv}}, \\
\text{and} \quad & o_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \max_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cc}}.
\end{aligned}
\tag{4.2.4}
$$

However, since numerical ODE solvers typically assume that the domain of the right-hand side function is an open set [72], these solvers may attempt to evaluate infeasible perturbations of (4.2.4) during integration of (2.4.1). On the other hand, the optimization problems in (4.2.3) are always feasible by construction. Thus, the formulation (4.2.3) may be beneficial during numerical implementation. The following example illustrates this point.

Example 4.1. Under Assumption 4.2.1, let $\mathbf{x}_0(\mathbf{p}) = \mathbf{p}$ for each $\mathbf{p}\in P$. Then, to construct Scott–Barton ODEs (2.4.1), it is natural to define $\mathbf{x}_0^{\mathrm{cv}}(\mathbf{p}) \equiv \mathbf{x}_0^{\mathrm{cc}}(\mathbf{p}) := \mathbf{p}$. Thus, the exact solution of (2.4.1) with (4.2.4) embedded is well-defined according to Section 4.3 below, and will have $\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}) \approx \mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})$ when $t \approx t_0$. As a result, during numerical integration of the relaxation ODEs (2.4.1) at $t = t_0$, an implicit ODE solver may attempt to evaluate the right-hand side functions (u_i, o_i) at $\xi_j^{\mathrm{cv}} > \xi_j^{\mathrm{cc}}$ for some $j \neq i$, and thus cause infeasibility of the optimization problems in (4.2.4). On the other hand, the formulation (4.2.3) remains feasible.

Observe that if valid state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ are employed in (4.2.3) in place of the dummy variables $(\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})$, the optimization problems in (4.2.4) would become multivariate McCormick relaxations [7, Theorem 2] of the composition $\mathbf{f}(t, \cdot, \mathbf{x}(t, \cdot))$, composed with the flattening operations of Definition 2.4.3. However, it is not obvious from [2] or [7] that this approach would yield valid state relaxations for (2.3.1). In [2], generalized McCormick relaxation method is the only method that is established to furnish valid Scott–Barton right-hand-side functions, and thus construct valid state relaxations in the Scott–Barton framework (2.4.1). In [7], multivariate McCormick relaxations are only applied to build relaxations for closed-form composite functions, and their correctness does not translate directly into correct state relaxations for (2.3.1).

Remark 4.2.4. By inspection of (4.2.4), in the special case where $n_x = 1$, $\mathbf{u}$ and $\mathbf{o}$ in (4.2.3) reduce to closed-form functions:

$$u(t, \mathbf{p}, \xi^{\mathrm{cv}}, \xi^{\mathrm{cc}}) \equiv f^{\mathrm{cv}}(t, \mathbf{p}, \xi^{\mathrm{cv}}) \qquad \text{and} \qquad o(t, \mathbf{p}, \xi^{\mathrm{cv}}, \xi^{\mathrm{cc}}) \equiv f^{\mathrm{cc}}(t, \mathbf{p}, \xi^{\mathrm{cc}}). \tag{4.2.5}$$

This formulation shows that if there is only one state variable, unlike established methods [2, 64, 66, 75] that propagate x^{cv} and x^{cc} in a coupled system, our new approach constructs x^{cv} and x^{cc} independently of each other. Thus, if only x^{cv} is of interest, then only x_0^{cv} and f^{cv} are required, and there is no need to construct x_0^{cc} or f^{cc} to compute x^{cv}.

Remark 4.2.5. Consider the special case where $n_x = 1$. In light of [79, Remark 5.4], generalized McCormick relaxations $\tilde{u}^{\mathrm{gMC}}, \tilde{o}^{\mathrm{gMC}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ for f may be applied to construct $(f^{\mathrm{cv}}, f^{\mathrm{cc}})$ in Assumption 4.2.1 by setting

$$f^{\mathrm{cv}}(t, \mathbf{p}, \xi) := \tilde{u}^{\mathrm{gMC}}(t, \mathbf{p}, \xi, \xi) \quad \text{and} \quad f^{\mathrm{cc}}(t, \mathbf{p}, \xi) := \tilde{o}^{\mathrm{gMC}}(t, \mathbf{p}, \xi, \xi)$$

(with notation as in [79]). With this choice of $(f^{\mathrm{cv}}, f^{\mathrm{cc}})$, our new approach (2.4.1) with (4.2.5) embedded reduces to the SBM method described in Section 2.4. Nonetheless, it is not apparent from [2] that x^{cv} and x^{cc} are decoupled in this case.

Remark 4.2.6. For $n_x > 1$, evaluating $(\mathbf{u}, \mathbf{o})$ in (4.2.3) ostensibly involves handling convex optimization problems by NLP-based methods. However, if the original right-hand side function $\mathbf{f}(t, \cdot, \cdot)$ is quadratic for each $t \in I$, and if αBB relaxations [9] are applied for $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$, then the optimization problems in (4.2.3) become bound-constrained convex quadratic programming problems, which have closed-form solutions that could in principle be computed *a priori*.

4.3 Properties of new state relaxations

This section establishes the following useful properties of the auxiliary ODE system (2.4.1) with (4.2.3) newly embedded, under Assumption 4.2.1. Each of the following properties essentially follows from a corresponding condition in Assumption 4.2.1.

- The right-hand side functions $\mathbf{u}$ and $\mathbf{o}$ defined in (4.2.3) are continuous, and $\mathbf{u}(t, \mathbf{p}, \cdot, \cdot)$ and $\mathbf{o}(t, \mathbf{p}, \cdot, \cdot)$ are Lipschitz continuous on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly over $(t, \mathbf{p}) \in I \times P$.
- The auxiliary ODE system (2.4.1) with (4.2.3) embedded has exactly one solution $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$, which is guaranteed to lie within the predefined state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$.

This unique solution is also a valid state relaxation for $\mathbf{x}$ in (2.3.1).

- In this approach, tighter state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ and tighter relaxations $(\mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}})$ and $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ necessarily translate into tighter state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$. If McCormick relaxations are used to construct $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$, then the new state relaxations are guaranteed to be at least as tight as the SBM relaxations [2].
- The new state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ inherit second-order pointwise convergence from $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$, and thereby avoid the cluster effect [37, 38] in branch-and-bound methods for deterministic nonconvex optimization.

This section then concludes by discussing how the approach of this chapter compares to established state relaxation methods.

4.3.1 Continuity

In this subsection, Proposition 4.3.1 will show that the functions $\mathbf{u}$ and $\mathbf{o}$ given by (4.2.3) are continuous. Proposition 4.3.2 will then show that the functions $\mathbf{u}(t, \mathbf{p}, \cdot, \cdot)$ and $\mathbf{o}(t, \mathbf{p}, \cdot, \cdot)$ are Lipschitz continuous on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly over $(t, \mathbf{p}) \in I \times P$.

Proposition 4.3.1. Under Assumption 4.2.1, the functions $\mathbf{u}$ and $\mathbf{o}$ defined in (4.2.3) are continuous.

Proof. Consider any fixed $i \in \{1, ..., n_x\}$. Since f_i^{cv} is continuous according to Assumption 4.2.1 and since $\mathbf{v}$ is also continuous as shown in (4.2.2), the mapping $(t, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \mapsto f_i^{\mathrm{cv}}(t, \mathbf{p}, \mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}))$ is continuous. Moreover, since the feasible set of the optimization problem of u_i in (4.2.3) is nonempty and independent of $(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})$, u_i is continuous according to Berge's Maximum Theorem [115, p116]. That the function $\mathbf{o}$ is continuous can be proved similarly. $\square$

Proposition 4.3.2. Under Assumption 4.2.1, the functions $\mathbf{u}(t,\mathbf{p},\cdot,\cdot)$ and $\mathbf{o}(t,\mathbf{p},\cdot,\cdot)$ defined in (4.2.3) are Lipschitz continuous on $\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$, uniformly over $(t,\mathbf{p})\in I\times P$.

Proof. Consider any fixed $i\in\{1,...,n_x\}$ and $(t,\mathbf{p})\in I\times P$. According to Assumption 4.2.1, there exists $l>0$ independent of i, t, and $\mathbf{p}$, so that for any $\boldsymbol{\xi}^{\mathrm{A}},\boldsymbol{\xi}^{\mathrm{B}}\in\mathbb{R}^{n_x}$,

$$|f_i^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{A}})-f_i^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{B}})|\leq l\|\boldsymbol{\xi}^{\mathrm{A}}-\boldsymbol{\xi}^{\mathrm{B}}\|_\infty. \tag{4.3.1}$$

By construction of $\mathbf{v}$ in (4.2.2), for each $m\in\{1,...,n_x\}$, $\boldsymbol{\alpha}\in[-1,1]^{n_x}$, and $\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}},\boldsymbol{\xi}^{\mathrm{cc,A}}\in\mathbb{R}^{n_x}$, we have:

$$\|\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}})-\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}})\|_\infty\leq\|\boldsymbol{\xi}^{\mathrm{cv,A}}-\boldsymbol{\xi}^{\mathrm{cv,B}}\|_\infty+\|\boldsymbol{\xi}^{\mathrm{cc,A}}-\boldsymbol{\xi}^{\mathrm{cc,B}}\|_\infty. \tag{4.3.2}$$

Combining (4.3.1) and (4.3.2), for any $\boldsymbol{\alpha}\in[-1,1]^{n_x}$ and $\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}},\boldsymbol{\xi}^{\mathrm{cc,A}}\in\mathbb{R}^{n_x}$,

$$\begin{aligned}&|f_i^{\mathrm{cv}}(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}))-f_i^{\mathrm{cv}}(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}}))|\\&\qquad\leq l\Big(\|\boldsymbol{\xi}^{\mathrm{cv,A}}-\boldsymbol{\xi}^{\mathrm{cv,B}}\|_\infty+\|\boldsymbol{\xi}^{\mathrm{cc,A}}-\boldsymbol{\xi}^{\mathrm{cc,B}}\|_\infty\Big).\end{aligned}$$

Moreover, since the function $f_i^{\mathrm{cv}}(t,\mathbf{p},\mathbf{v}(\cdot,\cdot,\cdot))$ is continuous, Theorem 2.1 in [116] shows that for any $\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}},\boldsymbol{\xi}^{\mathrm{cc,A}}\in\mathbb{R}^{n_x}$,

$$|u_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}})-u_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}})|\leq l\Big(\|\boldsymbol{\xi}^{\mathrm{cv,A}}-\boldsymbol{\xi}^{\mathrm{cv,B}}\|_\infty+\|\boldsymbol{\xi}^{\mathrm{cc,A}}-\boldsymbol{\xi}^{\mathrm{cc,B}}\|_\infty\Big). \tag{4.3.3}$$

Since the choice of $(t,\mathbf{p})$ was arbitrary, (4.3.3) holds for any $(t,\mathbf{p})\in I\times P$.

That $\mathbf{o}(t,\mathbf{p},\cdot,\cdot)$ and the other $u_j(t,\mathbf{p},\cdot,\cdot)$s are Lipschitz continuous on $\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$ uniformly over $(t,\mathbf{p})\in I\times P$ can be demonstrated similarly.

□

4.3.2 Existence and right-uniqueness

In this subsection, Theorem 4.3.3 establishes existence on $I \times P$ of solutions of (2.4.1) with $(\mathbf{u}, \mathbf{o})$ defined by (4.2.3), and these solutions are guaranteed to lie within the predefined state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$. Theorem 4.3.4 subsequently establishes right-uniqueness for these solutions.

Theorem 4.3.3. Under Assumption 4.2.1, for each $\mathbf{p} \in P$, the ODE system (2.4.1) with $\mathbf{p}$ fixed and with $(\mathbf{u}, \mathbf{o})$ defined by (4.2.3) has at least one solution $(\mathbf{x}^{\mathrm{cv}}(\cdot, \mathbf{p}), \mathbf{x}^{\mathrm{cc}}(\cdot, \mathbf{p}))$ on I, which satisfies $\mathbf{x}^{\mathrm{cv}}(t, \mathbf{p}), \mathbf{x}^{\mathrm{cc}}(t, \mathbf{p}) \in X(t)$ for each $t \in I$.

Proof. Consider a sufficiently large interval $\tilde{X} := \{\boldsymbol{\xi} \in \mathbb{R}^{n_x} : \underline{\mathbf{x}} \leq \boldsymbol{\xi} \leq \overline{\mathbf{x}}\}$, for which $X(t) \subseteq \tilde{X}$ for each $t \in I$. Consider a function $\mathbf{m} : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ for which, for any $\boldsymbol{\gamma} \in \mathbb{R}^{n_x}$, $\mathbf{m}(\boldsymbol{\gamma})$ returns the componentwise median value of the collection $\{\underline{\mathbf{x}}, \overline{\mathbf{x}}, \boldsymbol{\gamma}\}$. Thus, the function $\mathbf{m}$ maps $\mathbb{R}^{n_x}$ into $\tilde{X}$. Define functions $\hat{\mathbf{u}}, \hat{\mathbf{o}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ as the compositions:

$$
\begin{aligned}
& \hat{\mathbf{u}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \equiv \mathbf{u}(t, \mathbf{p}, \mathbf{m}(\boldsymbol{\xi}^{\mathrm{cv}}), \mathbf{m}(\boldsymbol{\xi}^{\mathrm{cc}})) \\
\text{and} \quad & \hat{\mathbf{o}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \equiv \mathbf{o}(t, \mathbf{p}, \mathbf{m}(\boldsymbol{\xi}^{\mathrm{cv}}), \mathbf{m}(\boldsymbol{\xi}^{\mathrm{cc}})),
\end{aligned}
$$

Now, we establish global existence of solutions of (2.4.1) with the substitution $(\mathbf{u}, \mathbf{o}) \leftarrow (\hat{\mathbf{u}}, \hat{\mathbf{o}})$. It will be shown subsequently that the solutions shown to exist for $(\hat{\mathbf{u}}, \hat{\mathbf{o}})$ must also solve (2.4.1) directly.

Consider any fixed $\mathbf{p} \in P$ throughout. Since $\mathbf{x}^{\mathrm{L}}$ and $\mathbf{x}^{\mathrm{U}}$ are LR-analytic, the derivatives $\dot{\mathbf{x}}^{\mathrm{L}}$ and $\dot{\mathbf{x}}^{\mathrm{U}}$ exist and are continuous except at finitely many discontinuities on I, according to Theorem 3.12 in [101]. Moreover, since the function $\mathbf{m}$ is Lipschitz continuous and maps $\mathbb{R}^{n_x}$ into $\tilde{X}$, $\hat{\mathbf{u}}$ and $\hat{\mathbf{o}}$ inherit the continuity of $\mathbf{u}$ and $\mathbf{o}$ established in Proposition 4.3.1, and are globally bounded. Thus, $(\hat{\mathbf{u}}, \hat{\mathbf{o}})$ are piecewise continuous in the sense of Filippov [102], which implies that (2.4.1) with $(\mathbf{u},\mathbf{o}) \leftarrow (\hat{\mathbf{u}},\hat{\mathbf{o}})$ is a differential inclusion system whose solutions may be alternatively described following the approach of §4, [102]. Since $\hat{\mathbf{u}}$ and $\hat{\mathbf{o}}$ are globally bounded, global existence then follows from applying [102, §7, Theorem 1] on a sufficiently large neighborhood of the initial value $(\mathbf{x}_0^{\mathrm{cv}}(\mathbf{p}),\mathbf{x}_0^{\mathrm{cc}}(\mathbf{p}))$.

Since the functions $(\hat{\mathbf{u}},\hat{\mathbf{o}})$ are continuous, [2, Lemma 1] implies that any solution $(\mathbf{x}^{\mathrm{cv}}(\cdot,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(\cdot,\mathbf{p}))$ of (2.4.1) with the substitution $(\mathbf{u},\mathbf{o}) \leftarrow (\hat{\mathbf{u}},\hat{\mathbf{o}})$ must satisfy $\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}) \in X(t)$ for each $t \in I$. Furthermore, by construction of the $\mathbf{m}$ function, these solutions must also be solutions of (2.4.1) with $(\mathbf{u},\mathbf{o})$ as defined in (4.2.3). □

Theorem 4.3.4. Under Assumption 4.2.1, for each $\mathbf{p} \in P$, the ODE system (2.4.1) with $\mathbf{p}$ fixed and with $(\mathbf{u},\mathbf{o})$ defined by (4.2.3) has right-uniqueness of solutions on I.

Proof. Consider any fixed $\mathbf{p} \in P$. Since the functions $\mathbf{u}$ and $\mathbf{o}$ have the uniform Lipschitz continuity established in Proposition 4.3.2, it is readily verified that the overall right-hand side functions of (2.4.1) satisfy the sufficient condition for right-uniqueness established in [102, §10, Theorem 1]. □

Remark 4.3.5. Along a solution trajectory of the ODE system (2.4.1), the derivatives $\dot{x}_i^{\mathrm{cv}}$ and $\dot{x}_i^{\mathrm{cc}}$ may respectively switch between the u_i and "max" cases and between the o_i and "min" cases depending on whether the solutions $(x_i^{\mathrm{cv}},x_i^{\mathrm{cc}})$ are on or strictly within the state bounds $(x_i^{\mathrm{L}},x_i^{\mathrm{U}})$. With these discontinuities present, (2.4.1) may only be expected to have right-uniqueness, even though the functions $\mathbf{u}$ and $\mathbf{o}$ are Lipschitz continuous. We note that if these switches do not occur, then (2.4.1) will indeed have standard "both-sides" uniqueness according to classical uniqueness results for ODEs such as [99, Theorem 1.1].

4.3.3 Bounding of original ODE solution

Theorem 4.3.7 below shows that the solutions $\mathbf{x}^{\mathrm{cv}}$ and $\mathbf{x}^{\mathrm{cc}}$ of (2.4.1) with $(\mathbf{u}, \mathbf{o})$ defined by (4.2.3) are, respectively, valid underestimators and overestimators of the solution $\mathbf{x}$ of (2.3.1). Recall that bound preserving dynamics were defined in Definition 2.4.7.

Lemma 4.3.6. Under Assumption 4.2.1, the functions $\mathbf{u}$ and $\mathbf{o}$ defined by (4.2.3) describe bound preserving dynamics for (2.3.1).

Proof. Consider any $i \in \{1, ..., n_x\}$, any functions $\boldsymbol{\xi}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}} : I \times P \to \mathbb{R}^{n_x}$, a.e. $t \in I$, and any $\mathbf{p} \in P$ satisfying the conditions of Definition 2.4.7. First, assume that $\xi_i(t, \mathbf{p}) = \xi_i^{\mathrm{cv}}(t, \mathbf{p})$. Since $\boldsymbol{\xi}^{\mathrm{cv}}(t, \mathbf{p}) \leq \boldsymbol{\xi}^{\mathrm{cc}}(t, \mathbf{p})$, (4.2.3) may be reformulated as (4.2.4), which yields

$$u_i(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}(t, \mathbf{p}), \boldsymbol{\xi}^{\mathrm{cc}}(t, \mathbf{p})) = \min_{\boldsymbol{\zeta} \in [\boldsymbol{\xi}^{\mathrm{cv}}(t, \mathbf{p}), \boldsymbol{\xi}^{\mathrm{cc}}(t, \mathbf{p})]} f_i^{\mathrm{cv}}(t, \mathbf{p}, \boldsymbol{\zeta}) \text{ subject to } \zeta_i = \xi_i^{\mathrm{cv}}(t, \mathbf{p}). \tag{4.3.4}$$

Since $\xi_i(t, \mathbf{p}) = \xi_i^{\mathrm{cv}}(t, \mathbf{p})$ and $\boldsymbol{\xi}^{\mathrm{cv}}(t, \mathbf{p}) \leq \boldsymbol{\xi}(t, \mathbf{p}) \leq \boldsymbol{\xi}^{\mathrm{cc}}(t, \mathbf{p})$, it follows that $\boldsymbol{\xi}(t, \mathbf{p})$ is feasible in the optimization problem in (4.3.4), and thus $u_i(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}(t, \mathbf{p}), \boldsymbol{\xi}^{\mathrm{cc}}(t, \mathbf{p})) \leq f_i^{\mathrm{cv}}(t, \mathbf{p}, \boldsymbol{\xi}(t, \mathbf{p}))$. Furthermore, since the function $f_i^{\mathrm{cv}}(t, \mathbf{p}, \cdot)$ is a convex relaxation of $f_i(t, \mathbf{p}, \cdot)$ on $X(t)$ and since $\boldsymbol{\xi}(t, \mathbf{p}) \in [\boldsymbol{\xi}^{\mathrm{cv}}(t, \mathbf{p}), \boldsymbol{\xi}^{\mathrm{cc}}(t, \mathbf{p})] \subseteq X(t)$,

$$u_i(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}(t, \mathbf{p}), \boldsymbol{\xi}^{\mathrm{cc}}(t, \mathbf{p})) \leq f_i^{\mathrm{cv}}(t, \mathbf{p}, \boldsymbol{\xi}(t, \mathbf{p})) \leq f_i(t, \mathbf{p}, \boldsymbol{\xi}(t, \mathbf{p})).$$

If we instead assume that $\xi_i(t, \mathbf{p}) = \xi_i^{\mathrm{cc}}(t, \mathbf{p})$, it can be proved similarly that

$$o_i(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}(t, \mathbf{p}), \boldsymbol{\xi}^{\mathrm{cc}}(t, \mathbf{p})) \geq f_i(t, \mathbf{p}, \boldsymbol{\xi}(t, \mathbf{p})).$$

Thus, the functions $\mathbf{u}$ and $\mathbf{o}$ describe bound preserving dynamics for (2.3.1). $\square$

Theorem 4.3.7. Suppose that Assumption 4.2.1 holds. Let $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ be a solution of (2.4.1) with $\mathbf{u}$ and $\mathbf{o}$ defined by (4.2.3). Let $\mathbf{x}$ be a solution of (2.3.1). Then, for all $(t, \mathbf{p}) \in I \times P$, $\mathbf{x}^{\mathrm{cv}}(t, \mathbf{p}) \leq \mathbf{x}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cc}}(t, \mathbf{p})$.

Proof. According to Lemma 4.3.6, the functions $\mathbf{u}$ and $\mathbf{o}$ describe bound preserving dynamics for (2.3.1). Since, moreover, Proposition 4.3.2 holds and the right-hand side $\mathbf{f}$ of (2.3.1) is Lipschitz continuous, the claimed result then follows from [2, Corollary 1]. □

Note that according to Theorem 4.3.7 above, the optimization problems in (4.2.4) are always feasible along a solution trajectory of (2.4.1) with (4.2.3) embedded.

4.3.4 Convexity

This subsection shows that the solutions $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ of (2.4.1) with (4.2.3) embedded are, respectively, convex and concave with respect to $\mathbf{p}$. Combined with the results of Sections 4.3.2 and 4.3.3, this shows that candidate relaxations deriving from (2.4.1) with $(\mathbf{u}, \mathbf{o})$ defined by (4.2.3) are indeed valid state relaxations for the original ODE system (2.3.1). Recall that convexity preserving dynamics were defined in Definition 2.4.8.

Lemma 4.3.8. Under Assumption 4.2.1, the functions $\mathbf{u}$ and $\mathbf{o}$ defined by (4.2.3) describe convexity preserving dynamics for (2.3.1).

Proof. Consider any $i \in \{1, ..., n_x\}$, a.e. $t \in I$, any $\mathbf{p}^{\mathrm{A}}, \mathbf{p}^{\mathrm{B}} \in P$, any $\lambda \in (0,1)$, $\bar{\mathbf{p}} := \lambda \mathbf{p}^{\mathrm{A}} + (1-\lambda)\mathbf{p}^{\mathrm{B}}$, and any functions $\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}} : I \times P \to \mathbb{R}^{n_x}$ as satisfying the conditions in Definition 2.4.8. Firstly, assume that

$$\xi_i^{\mathrm{cv}}(t, \bar{\mathbf{p}}) = \lambda \xi_i^{\mathrm{cv}}(t, \mathbf{p}^{\mathrm{A}}) + (1-\lambda)\xi_i^{\mathrm{cv}}(t, \mathbf{p}^{\mathrm{B}}).$$

Since $\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{q}) \leq \boldsymbol{\xi}^{\text{cc}}(t,\mathbf{q})$, for all $\mathbf{q} \in \{\mathbf{p}^{\text{A}}, \mathbf{p}^{\text{B}}, \bar{\mathbf{p}}\}$, the reformulation (4.2.4) of (4.2.3) yields

$$u_i(t,\mathbf{q},\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{q}),\boldsymbol{\xi}^{\text{cc}}(t,\mathbf{q})) = \min_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{q}),\boldsymbol{\xi}^{\text{cc}}(t,\mathbf{q})]} f_i^{\text{cv}}(t,\mathbf{q},\boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\text{cv}}(t,\mathbf{q}). \tag{4.3.5}$$

Let $\boldsymbol{\xi}^{*,\text{A}}, \boldsymbol{\xi}^{*,\text{B}} \in \mathbb{R}^{n_x}$ be optimal solutions of the optimization problems in (4.3.5) at $\mathbf{q} := \mathbf{p}^{\text{A}}$ and $\mathbf{p}^{\text{B}}$, respectively, and define $\boldsymbol{\zeta} := \lambda\boldsymbol{\xi}^{*,\text{A}} + (1-\lambda)\boldsymbol{\xi}^{*,\text{B}}$. Then, we have

$$\begin{aligned} &\lambda\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}^{\text{A}}) + (1-\lambda)\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}^{\text{B}}) \leq \boldsymbol{\zeta} \\ \text{and}\quad &\lambda\xi_i^{\text{cv}}(t,\mathbf{p}^{\text{A}}) + (1-\lambda)\xi_i^{\text{cv}}(t,\mathbf{p}^{\text{B}}) = \zeta_i. \end{aligned}$$

Moreover, since

$$\begin{aligned} &\boldsymbol{\xi}^{\text{cv}}(t,\bar{\mathbf{p}}) \leq \lambda\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}^{\text{A}}) + (1-\lambda)\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{p}^{\text{B}}) \\ \text{and}\quad &\xi_i^{\text{cv}}(t,\bar{\mathbf{p}}) = \lambda\xi_i^{\text{cv}}(t,\mathbf{p}^{\text{A}}) + (1-\lambda)\xi_i^{\text{cv}}(t,\mathbf{p}^{\text{B}}) \end{aligned}$$

by assumption, it follows that

$$\boldsymbol{\xi}^{\text{cv}}(t,\bar{\mathbf{p}}) \leq \boldsymbol{\zeta} \quad \text{and} \quad \xi_i^{\text{cv}}(t,\bar{\mathbf{p}}) = \zeta_i.$$

Similarly, it can be shown that $\boldsymbol{\zeta} \leq \boldsymbol{\xi}^{\text{cc}}(t,\bar{\mathbf{p}})$. Thus, $\boldsymbol{\zeta}$ is feasible in the optimization problem in (4.3.5) at $\mathbf{q} := \bar{\mathbf{p}}$, which implies that

$$u_i(t,\bar{\mathbf{p}},\boldsymbol{\xi}^{\text{cv}}(t,\bar{\mathbf{p}}),\boldsymbol{\xi}^{\text{cc}}(t,\bar{\mathbf{p}})) \leq f_i^{\text{cv}}(t,\bar{\mathbf{p}},\boldsymbol{\zeta}).$$

Since for a.e. $t \in I$, $f_i^{\text{cv}}(t,\cdot,\cdot)$ is convex on $P \times X(t)$ and $[\boldsymbol{\xi}^{\text{cv}}(t,\mathbf{q}),\boldsymbol{\xi}^{\text{cc}}(t,\mathbf{q})] \subseteq X(t)$, for each $\mathbf{q} \in \{\mathbf{p}^{\mathrm{A}}, \mathbf{p}^{\mathrm{B}}, \bar{\mathbf{p}}\}$ and a.e. $t \in I$, it follows that

$$\begin{aligned} u_i(t, \bar{\mathbf{p}}, \boldsymbol{\xi}^{\mathrm{cv}}(t, \bar{\mathbf{p}}), \boldsymbol{\xi}^{\mathrm{cc}}(t, \bar{\mathbf{p}})) &\leq f_i^{\mathrm{cv}}(t, \bar{\mathbf{p}}, \boldsymbol{\zeta}), \\ &\leq \lambda f_i^{\mathrm{cv}}(t, \mathbf{p}^{\mathrm{A}}, \boldsymbol{\xi}^{*,\mathrm{A}}) + (1-\lambda) f_i^{\mathrm{cv}}(t, \mathbf{p}^{\mathrm{B}}, \boldsymbol{\xi}^{*,\mathrm{B}}) \\ &= \lambda u_i(t, \mathbf{p}^{\mathrm{A}}, \boldsymbol{\xi}^{\mathrm{cv}}(t, \mathbf{p}^{\mathrm{A}}), \boldsymbol{\xi}^{\mathrm{cc}}(t, \mathbf{p}^{\mathrm{A}})) \\ &\quad + (1-\lambda) u_i(t, \mathbf{p}^{\mathrm{B}}, \boldsymbol{\xi}^{\mathrm{cv}}(t, \mathbf{p}^{\mathrm{B}}), \boldsymbol{\xi}^{\mathrm{cc}}(t, \mathbf{p}^{\mathrm{B}})), \end{aligned}$$

as required.

If we instead assume that $\xi_i^{\mathrm{cc}}(t, \bar{\mathbf{p}}) = \lambda \xi_i^{\mathrm{cc}}(t, \mathbf{p}^{\mathrm{A}}) + (1-\lambda) \xi_i^{\mathrm{cc}}(t, \mathbf{p}^{\mathrm{B}})$, then an analogous concavity is established for $o_i(t, \cdot, \boldsymbol{\xi}^{\mathrm{cv}}(t, \cdot), \boldsymbol{\xi}^{\mathrm{cc}}(t, \cdot))$. Thus, the functions $\mathbf{u}$ and $\mathbf{o}$ describe convexity preserving dynamics for (2.3.1). □

Theorem 4.3.9. Suppose that Assumption 4.2.1 holds. Let $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ be a solution of (2.4.1) with $\mathbf{u}$ and $\mathbf{o}$ defined by (4.2.3). Then, for each $i \in \{1, ..., n_x\}$ and each $t \in I$, $x_i^{\mathrm{cv}}(t, \cdot)$ is convex on P and $x_i^{\mathrm{cc}}(t, \cdot)$ is concave on P.

Proof. According to Lemmata 4.3.6 and 4.3.8, the functions $\mathbf{u}$ and $\mathbf{o}$ describe both bound preserving dynamics and convexity preserving dynamics for (2.3.1). The claimed result then follows from Proposition 4.3.2 and [2, Theorem 3]. □

4.3.5 Tightness

This subsection shows that if tighter relaxations of $\mathbf{f}$ are employed in (4.2.3), then the auxiliary ODE system (2.4.1) with $(\mathbf{u}, \mathbf{o})$ defined by (4.2.3) is guaranteed to yield tighter state relaxations for the ODE system (2.3.1). A procedure is presented for embedding any functions $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ satisfying Assumptions 2.4.11 and 2.4.12 into the new approach. Doing so is shown to yield state relaxations that are at least as tight as those obtained by solving

(2.4.1) with (2.4.2) embedded. Hence, if McCormick relaxations are used in (4.2.3), then the new state relaxations are at least as tight as the SBM relaxations [2].

To begin, we reproduce the following tightness result concerning the Scott–Barton ODE relaxation framework (2.4.1) from Theorem 3.5.1 in Chapter 3. Suppose that there are two considered choices for $(\mathbf{u},\mathbf{o})$ in (2.4.1); call these $(\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}})$ and $(\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}})$, and let us label quantities relating to each with "A" or "B" superscripts. Proposition 4.3.10 below provides sufficient conditions under which the resulting relaxations $(\mathbf{x}^{\mathrm{cv,B}},\mathbf{x}^{\mathrm{cc,B}})$ are guaranteed to be at least as tight as $(\mathbf{x}^{\mathrm{cv,A}},\mathbf{x}^{\mathrm{cc,A}})$. This result is then specialized to the particular choice (4.2.3) of $(\mathbf{u},\mathbf{o})$.

Proposition 4.3.10 (from [91])**.** Consider state lower bounds $\mathbf{x}^{\mathrm{L,A}},\mathbf{x}^{\mathrm{L,B}}: I\to\mathbb{R}^{n_x}$ and state upper bounds $\mathbf{x}^{\mathrm{U,A}},\mathbf{x}^{\mathrm{U,B}}: I\to\mathbb{R}^{n_x}$ for (2.3.1) that are absolutely continuous, and suppose for all $t\in I$ that $\mathbf{x}^{\mathrm{L,A}}(t)\leq\mathbf{x}^{\mathrm{L,B}}(t)\leq\mathbf{x}^{\mathrm{U,B}}(t)\leq\mathbf{x}^{\mathrm{U,A}}(t)$. For each $t\in I$, denote the intervals $[\mathbf{x}^{\mathrm{L,A}}(t),\mathbf{x}^{\mathrm{U,A}}(t)]$ and $[\mathbf{x}^{\mathrm{L,B}}(t),\mathbf{x}^{\mathrm{U,B}}(t)]$ as $X^{\mathrm{A}}(t)$ and $X^{\mathrm{B}}(t)$, respectively. Consider convex relaxations $\mathbf{x}_0^{\mathrm{cv,A}},\mathbf{x}_0^{\mathrm{cv,B}}: P\to\mathbb{R}^{n_x}$ and concave relaxations $\mathbf{x}_0^{\mathrm{cc,A}},\mathbf{x}_0^{\mathrm{cc,B}}: P\to\mathbb{R}^{n_x}$ for the initial-value function $\mathbf{x}_0$ in (2.3.1), and suppose for all $\mathbf{p}\in P$ that $\mathbf{x}_0^{\mathrm{cv,A}}(\mathbf{p})\leq\mathbf{x}_0^{\mathrm{cv,B}}(\mathbf{p})\leq\mathbf{x}_0^{\mathrm{cc,B}}(\mathbf{p})\leq\mathbf{x}_0^{\mathrm{cc,A}}(\mathbf{p})$. Consider functions $\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}},\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}}: I\times P\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$, and suppose that the following conditions hold.

I. The functions $(\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}})$ and $(\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}})$ are Scott–Barton right-hand side functions as in Definition 2.4.9 based on bounds $X^{\mathrm{A}}(t)$ and $X^{\mathrm{B}}(t)$, respectively.

II. For any $i\in\{1,\ldots,n_x\}$, any $\mathbf{p}\in P$, a.e. $t\in I$, any $\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}\in X^{\mathrm{A}}(t)$, and any $\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}}\in X^{\mathrm{B}}(t)$ such that $\boldsymbol{\xi}^{\mathrm{cv,A}}\leq\boldsymbol{\xi}^{\mathrm{cv,B}}\leq\boldsymbol{\xi}^{\mathrm{cc,B}}\leq\boldsymbol{\xi}^{\mathrm{cc,A}}$,

   (a) if $\xi_i^{\mathrm{cv,A}}=\xi_i^{\mathrm{cv,B}}$, then $u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}})\leq u_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}})$,

   (b) if $\xi_i^{\mathrm{cc,A}}=\xi_i^{\mathrm{cc,B}}$, then $o_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}})\geq o_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}})$.

Let $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}, \mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}}, \mathbf{u}, \mathbf{o}) := (\mathbf{x}^{\mathrm{L,A}}, \mathbf{x}^{\mathrm{U,A}}, \mathbf{x}_0^{\mathrm{cv,A}}, \mathbf{x}_0^{\mathrm{cc,A}}, \mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}).$$

Let $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}, \mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}}, \mathbf{u}, \mathbf{o}) := (\mathbf{x}^{\mathrm{L,B}}, \mathbf{x}^{\mathrm{U,B}}, \mathbf{x}_0^{\mathrm{cv,B}}, \mathbf{x}_0^{\mathrm{cc,B}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}}).$$

Then, $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ are both valid state relaxations for (2.3.1). Moreover, for any $(t, \mathbf{p}) \in I \times P$,

$$\mathbf{x}^{\mathrm{cv,A}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cv,B}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,B}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,A}}(t, \mathbf{p}).$$

Based on this proposition, the following theorem provides an inclusion monotonicity result for the approach of this chapter, showing essentially that tighter bounds and relaxations for $\mathbf{f}$ and $\mathbf{x}_0$ in (2.3.1) translate into tighter state relaxations for (2.3.1).

Theorem 4.3.11. Consider functions $(\mathbf{x}^{\mathrm{L,A}}, \mathbf{x}^{\mathrm{L,B}}, \mathbf{x}^{\mathrm{U,A}}, \mathbf{x}^{\mathrm{U,B}})$ and $(\mathbf{x}_0^{\mathrm{cv,A}}, \mathbf{x}_0^{\mathrm{cv,B}}, \mathbf{x}_0^{\mathrm{cc,A}}, \mathbf{x}_0^{\mathrm{cc,B}})$ and intervals $X^{\mathrm{A}}(t)$ and $X^{\mathrm{B}}(t)$ as in Proposition 4.3.10. Consider functions $\mathbf{f}^{\mathrm{cv,A}}, \mathbf{f}^{\mathrm{cc,A}}, \mathbf{f}^{\mathrm{cv,B}}, \mathbf{f}^{\mathrm{cc,B}} : I \times P \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$, and suppose that the following conditions hold.

I. Assumption 4.2.1 is satisfied both with $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}}) := (\mathbf{f}^{\mathrm{cv,A}}, \mathbf{f}^{\mathrm{cc,A}})$ and $X(t) := X^{\mathrm{A}}(t)$, and with $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}}) := (\mathbf{f}^{\mathrm{cv,B}}, \mathbf{f}^{\mathrm{cc,B}})$ and $X(t) := X^{\mathrm{B}}(t)$.

II. For any $t \in I$, $\mathbf{p} \in P$, and $\boldsymbol{\xi} \in X^{\mathrm{B}}(t)$,

$$\mathbf{f}^{\mathrm{cv,A}}(t, \mathbf{p}, \boldsymbol{\xi}) \leq \mathbf{f}^{\mathrm{cv,B}}(t, \mathbf{p}, \boldsymbol{\xi}) \leq \mathbf{f}^{\mathrm{cc,B}}(t, \mathbf{p}, \boldsymbol{\xi}) \leq \mathbf{f}^{\mathrm{cc,A}}(t, \mathbf{p}, \boldsymbol{\xi}).$$

Consider the function $\mathbf{v}$ defined in (4.2.2). Define functions $\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ so that, for each $i \in \{1, ..., n_x\}$ and each $(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \in I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$,

$$\begin{aligned}
u_i^{\mathrm{A}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) &:= \min_{\boldsymbol{\alpha} \in [-1,1]^{n_x}} f_i^{\mathrm{cv,A}}(t, \mathbf{p}, \mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})) \text{ subject to } \alpha_i = -1, \\
o_i^{\mathrm{A}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) &:= \max_{\boldsymbol{\alpha} \in [-1,1]^{n_x}} f_i^{\mathrm{cc,A}}(t, \mathbf{p}, \mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})) \text{ subject to } \alpha_i = +1, \\
u_i^{\mathrm{B}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) &:= \min_{\boldsymbol{\alpha} \in [-1,1]^{n_x}} f_i^{\mathrm{cv,B}}(t, \mathbf{p}, \mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc})) \text{ subject to } \alpha_i = -1, \\
\text{and} \quad o_i^{\mathrm{B}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) &:= \max_{\boldsymbol{\alpha} \in [-1,1]^{n_x}} f_i^{\mathrm{cc,B}}(t, \mathbf{p}, \mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})) \text{ subject to } \alpha_i = +1.
\end{aligned}$$

Let $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}, \mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}}, \mathbf{u}, \mathbf{o}) := (\mathbf{x}^{\mathrm{L,A}}, \mathbf{x}^{\mathrm{U,A}}, \mathbf{x}_0^{\mathrm{cv,A}}, \mathbf{x}_0^{\mathrm{cc,A}}, \mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}).$$

Let $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}, \mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}}, \mathbf{u}, \mathbf{o}) := (\mathbf{x}^{\mathrm{L,B}}, \mathbf{x}^{\mathrm{U,B}}, \mathbf{x}_0^{\mathrm{cv,B}}, \mathbf{x}_0^{\mathrm{cc,B}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}}).$$

Then, $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ are both valid state relaxations for (2.3.1). Moreover, for any $(t, \mathbf{p}) \in I \times P$,

$$\mathbf{x}^{\mathrm{cv,A}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cv,B}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,B}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,A}}(t, \mathbf{p}).$$

Proof. We proceed by verifying all conditions of $(\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$ in Proposition 4.3.10. Since Assumption 4.2.1 applies to $(\mathbf{f}^{\mathrm{cv,A}}, \mathbf{f}^{\mathrm{cc,A}})$ and $(\mathbf{f}^{\mathrm{cv,B}}, \mathbf{f}^{\mathrm{cc,B}})$, since Propositions 4.3.1 and 4.3.2 hold, and since Lemmata 4.3.6 and 4.3.8 apply with $(\mathbf{u}, \mathbf{o}) := (\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}})$ or $(\mathbf{u}, \mathbf{o}) := (\mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$, Condition I in Proposition 4.3.10 is satisfied. Thus, $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ are valid state relaxations for (2.3.1).

Consider any $i \in \{1,...,n_x\}$, any $\mathbf{p} \in P$, and a.e. $t \in I$. Consider any $\boldsymbol{\xi}^{\mathrm{cv,A}}, \boldsymbol{\xi}^{\mathrm{cc,A}} \in X^{\mathrm{A}}(t)$ and $\boldsymbol{\xi}^{\mathrm{cv,B}}, \boldsymbol{\xi}^{\mathrm{cc,B}} \in X^{\mathrm{B}}(t)$ such that

$$\boldsymbol{\xi}^{\mathrm{cv,A}} \leq \boldsymbol{\xi}^{\mathrm{cv,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,A}} \quad \text{and} \quad \xi_i^{\mathrm{cv,A}} = \xi_i^{\mathrm{cv,B}}. \tag{4.3.6}$$

Then, according to the reformulation (4.2.4),

$$u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}) \equiv \min_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}]} f_i^{\mathrm{cv,A}}(t,\mathbf{p},\boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cv,A}}, \tag{4.3.7}$$

$$\text{and} \quad u_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}}) \equiv \min_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}}]} f_i^{\mathrm{cv,B}}(t,\mathbf{p},\boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cv,B}}. \tag{4.3.8}$$

Since (4.3.6) holds, the feasible set of the optimization problem in (4.3.8) is a subset of the feasible set of the optimization problem in (4.3.7). Furthermore, since $f_i^{\mathrm{cv,A}}(t,\mathbf{p},\boldsymbol{\xi}) \leq f_i^{\mathrm{cv,B}}(t,\mathbf{p},\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in X^{\mathrm{B}}(t)$, it holds that $u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}) \leq u_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}})$. Similarly, if we consider any $\boldsymbol{\xi}^{\mathrm{cv,A}}, \boldsymbol{\xi}^{\mathrm{cc,A}} \in X^{\mathrm{A}}(t)$ and $\boldsymbol{\xi}^{\mathrm{cv,B}}, \boldsymbol{\xi}^{\mathrm{cc,B}} \in X^{\mathrm{B}}(t)$ such that

$$\boldsymbol{\xi}^{\mathrm{cv,A}} \leq \boldsymbol{\xi}^{\mathrm{cv,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,A}} \text{ and } \xi_i^{\mathrm{cc,A}} = \xi_i^{\mathrm{cc,B}},$$

it is readily verified that

$$o_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}) \geq o_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}}).$$

Thus, Condition II in Proposition 4.3.10 is satisfied. □

Remark 4.3.12. Theorem 4.3.11 implies that if the convex/concave envelopes (i.e. the tightest possible relaxations) of $\mathbf{x}_0$ and $\mathbf{f}(t,\cdot,\cdot)$ for all $t \in I$ are available, then these envelopes will translate into the tightest possible state relaxations generated by the new approach in this chapter. Note that these envelopes for $\mathbf{x}_0$ and $\mathbf{f}$ will not necessarily translate into envelopes for $\mathbf{x}$.

The following corollary shows that any functions $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{o}}$ that satisfy Assumptions 2.4.11 and 2.4.12 can be embedded into the new ODE relaxation approach, and that doing so always yields state relaxations that are at least as tight as those obtained by the approach by Scott and Barton [2] of solving (2.4.1) with (2.4.2) embedded. Note that this corollary does not specify any particular choice of $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$. Corollary 4.3.15 then shows that our new relaxations are at least as tight as SBM relaxations when $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ are specialized to generalized McCormick relaxations.

Corollary 4.3.13. Consider functions $(\mathbf{x}_0^{\mathrm{cv,A}},\mathbf{x}_0^{\mathrm{cv,B}},\mathbf{x}_0^{\mathrm{cc,A}},\mathbf{x}_0^{\mathrm{cc,B}})$ as in Proposition 4.3.10, with $X^{\mathrm{A}} := X$ and $X^{\mathrm{B}} := X$ (where X was described by Definition 2.4.2). Consider functions $\tilde{\mathbf{u}},\tilde{\mathbf{o}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ that satisfy Assumptions 2.4.11 and 2.4.12. Consider the interval flattening functions $\mathbf{r}^{i,\mathrm{L}}$ and $\mathbf{r}^{i,\mathrm{U}}$ for each $i \in \{1,..,n_x\}$ from Definition 2.4.3. Consider the function $\mathbf{v}$ defined in (4.2.2). Define functions $\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}},\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ so that, for each $i \in \{1,...,n_x\}$ and each $(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \in I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$,

$$u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \tilde{u}_i(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})),$$

$$o_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \tilde{o}_i(t,\mathbf{p},\mathbf{r}^{i,\mathrm{U}}(\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})),$$

$$u_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \min_{\boldsymbol{\alpha}\in[-1,1]^{n_x}} \tilde{u}_i(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}),\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})) \text{ subject to } \alpha_i = -1,$$

and $$o_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \max_{\boldsymbol{\alpha}\in[-1,1]^{n_x}} \tilde{o}_i(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}),\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})) \text{ subject to } \alpha_i = +1.$$

Define functions $\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}} : I \times P \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ for which

$$\mathbf{f}^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi}) := \tilde{\mathbf{u}}(t,\mathbf{p},\boldsymbol{\xi},\boldsymbol{\xi}) \quad \text{and} \quad \mathbf{f}^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi}) := \tilde{\mathbf{o}}(t,\mathbf{p},\boldsymbol{\xi},\boldsymbol{\xi}). \tag{4.3.9}$$

Let $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ be a solution of (2.4.1) with $(\mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}}, \mathbf{u}, \mathbf{o}) := (\mathbf{x}_0^{\mathrm{cv,A}}, \mathbf{x}_0^{\mathrm{cc,A}}, \mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}})$. Let $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ be a solution of (2.4.1) with $(\mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}}, \mathbf{u}, \mathbf{o}) := (\mathbf{x}_0^{\mathrm{cv,B}}, \mathbf{x}_0^{\mathrm{cc,B}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$. Then, $(\mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$ are the special case of (4.2.3) with $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ given by (4.3.9), this choice of $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ satisfies Assumption 4.2.1, and $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ are both valid state relaxations for (2.3.1). Moreover, for any $(t,\mathbf{p}) \in I \times P$,

$$\mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p}).$$

Proof. We proceed by verifying all conditions of $(\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$ in Proposition 4.3.10 with $X^A(t) \equiv X^B(t) := X(t)$. Since the functions $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{o}}$ satisfy Assumptions 2.4.11 and 2.4.12, it follows that for a.e. $t \in I$, the mappings $\tilde{\mathbf{u}}(t,\cdot,\cdot,\cdot)$ and $\tilde{\mathbf{o}}(t,\cdot,\cdot,\cdot)$ are *relaxation functions* for $\mathbf{f}(t,\cdot,\cdot)$ on $P \times X(t)$ as defined in [79, Definition 5.3 and Remark 5.4]. Then, according to Lemma 2.4.11 in [39], $\mathbf{f}^{\mathrm{cv}}$ and $\mathbf{f}^{\mathrm{cc}}$ defined in (4.3.9) satisfy Assumption 4.2.1. Thus, $\mathbf{u}^{\mathrm{B}}$ and $\mathbf{o}^{\mathrm{B}}$ are indeed the special case of (4.2.3) with (4.3.9) embedded, and are Scott–Barton right-hand side functions as in Definition 2.4.9, according to Propositions 4.3.1 and 4.3.2 and Lemmata 4.3.6 and 4.3.8. According to Lemmata 10 and 11 in [2], $\mathbf{u}^{\mathrm{A}}$ and $\mathbf{o}^{\mathrm{A}}$ are also Scott–Barton right-hand side functions. Hence, Condition I in Proposition 4.3.10 is satisfied, and $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ are valid state relaxations.

Next, we verify Condition II in Proposition 4.3.10. Consider any $i \in \{1,\ldots,n_x\}$, any

$\mathbf{p} \in P$, and a.e. $t \in I$. Consider any $\boldsymbol{\xi}^{\mathrm{cv,A}}, \boldsymbol{\xi}^{\mathrm{cc,A}}, \boldsymbol{\xi}^{\mathrm{cv,B}}, \boldsymbol{\xi}^{\mathrm{cc,B}} \in X(t)$ such that

$$\boldsymbol{\xi}^{\mathrm{cv,A}} \leq \boldsymbol{\xi}^{\mathrm{cv,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,A}} \quad \text{and} \quad \xi_i^{\mathrm{cv,A}} = \xi_i^{\mathrm{cv,B}}. \tag{4.3.10}$$

Let $(\hat{\boldsymbol{\xi}}^{\mathrm{cv,A}}, \hat{\boldsymbol{\xi}}^{\mathrm{cc,A}}) := \mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\xi}^{\mathrm{cv,A}}, \boldsymbol{\xi}^{\mathrm{cc,A}})$. According to Definition 2.4.3 of $\mathbf{r}^{i,\mathrm{L}}$,

$$\begin{aligned}
&\hat{\xi}_i^{\mathrm{cv,A}} = \hat{\xi}_i^{\mathrm{cc,A}} = \xi_i^{\mathrm{cv,A}}, \\
&\hat{\xi}_j^{\mathrm{cv,A}} = \xi_j^{\mathrm{cv,A}}, \quad \text{and} \quad \hat{\xi}_j^{\mathrm{cc,A}} = \xi_j^{\mathrm{cc,A}}, \quad \forall j \in \{1, ..., n_x\} \text{ and } j \neq i.
\end{aligned} \tag{4.3.11}$$

Since $\boldsymbol{\xi}^{\mathrm{cv,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,B}}$, according to the reformulation (4.2.4),

$$u_i^{\mathrm{B}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv,B}}, \boldsymbol{\xi}^{\mathrm{cc,B}}) \equiv \min_{\boldsymbol{\xi} \in [\boldsymbol{\xi}^{\mathrm{cv,B}}, \boldsymbol{\xi}^{\mathrm{cc,B}}]} \tilde{u}_i(t, \mathbf{p}, \boldsymbol{\xi}, \boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cv,B}}.$$

Let $\boldsymbol{\xi}^{*,\mathrm{B}}$ be an optimal solution of the optimization problem above. It follows that $\boldsymbol{\xi}^{\mathrm{cv,B}} \leq \boldsymbol{\xi}^{*,\mathrm{B}} \leq \boldsymbol{\xi}^{\mathrm{cc,B}}$ and $\xi_i^{*,\mathrm{B}} = \xi_i^{\mathrm{cv,B}}$. Moreover, since (4.3.10) and (4.3.11) hold,

$$\begin{aligned}
&\xi_i^{*,\mathrm{B}} = \xi_i^{\mathrm{cv,B}} = \xi_i^{\mathrm{cv,A}} = \hat{\xi}_i^{\mathrm{cv,A}} = \hat{\xi}_i^{\mathrm{cc,A}}, \\
&\hat{\xi}_j^{\mathrm{cv,A}} = \xi_j^{\mathrm{cv,A}} \leq \xi_j^{\mathrm{cv,B}} \leq \xi_j^{*,\mathrm{B}} \leq \xi_j^{\mathrm{cc,B}} \leq \xi_j^{\mathrm{cc,A}} = \hat{\xi}_j^{\mathrm{cc,A}}, \quad \forall j \in \{1, ..., n_x\} \text{ and } j \neq i.
\end{aligned}$$

Thus, $\hat{\boldsymbol{\xi}}^{\mathrm{cv,A}} \leq \boldsymbol{\xi}^{*,\mathrm{B}} \leq \hat{\boldsymbol{\xi}}^{\mathrm{cc,A}}$. Since Assumption 2.4.12 holds for $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ and $\hat{\boldsymbol{\xi}}^{\mathrm{cv,A}}, \hat{\boldsymbol{\xi}}^{\mathrm{cc,A}} \in X(t)$,

$$\tilde{u}_i(t, \mathbf{p}, \hat{\boldsymbol{\xi}}^{\mathrm{cv,A}}, \hat{\boldsymbol{\xi}}^{\mathrm{cc,A}}) \leq \tilde{u}_i(t, \mathbf{p}, \boldsymbol{\xi}^{*,\mathrm{B}}, \boldsymbol{\xi}^{*,\mathrm{B}}),$$

Since $(\hat{\boldsymbol{\xi}}^{\mathrm{cv,A}}, \hat{\boldsymbol{\xi}}^{\mathrm{cc,A}}) := \mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\xi}^{\mathrm{cv,A}}, \boldsymbol{\xi}^{\mathrm{cc,A}})$ by definition, it follows that

$$\tilde{u}_i(t, \mathbf{p}, \mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\xi}^{\mathrm{cv,A}}, \boldsymbol{\xi}^{\mathrm{cc,A}})) \leq u_i^{\mathrm{B}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv,B}}, \boldsymbol{\xi}^{\mathrm{cc,B}}),$$

which implies that

$$u_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}) \leq u_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}}).$$

Similarly, if we instead assume $\xi_i^{\mathrm{cc,A}} = \xi_i^{\mathrm{cc,B}}$, it can be verified that

$$o_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}) \geq o_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}}).$$

Thus, the claimed result follows from Proposition 4.3.10. □

Remark 4.3.14. The corollary above applies the same state bounds $(\mathbf{x}^L,\mathbf{x}^U)$ to both relaxation systems “A” and “B”. We remark that, as a straightforward extension of Proposition 4.3.10 and Corollary 4.3.13, the following tightness result holds. Consider state bounds $(\mathbf{x}^{\mathrm{L,A}},\mathbf{x}^{\mathrm{L,B}},\mathbf{x}^{\mathrm{U,A}},\mathbf{x}^{\mathrm{U,B}})$ as in Proposition 4.3.10, and, informally, consider a method for constructing appropriate $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ that permits any choice of state bounds. Suppose that we construct $(\tilde{\mathbf{u}}^{\mathrm{B}},\tilde{\mathbf{o}}^{\mathrm{B}})$ based on $(\mathbf{x}^{\mathrm{L,B}},\mathbf{x}^{\mathrm{U,B}})$ and construct $(\tilde{\mathbf{u}}^{\mathrm{A}},\tilde{\mathbf{o}}^{\mathrm{A}})$ based on $(\mathbf{x}^{\mathrm{L,A}},\mathbf{x}^{\mathrm{U,A}})$. Then, using $(\tilde{\mathbf{u}}^{\mathrm{B}},\tilde{\mathbf{o}}^{\mathrm{B}})$ in our new ODE relaxation approach will necessarily lead to state relaxations that are at least as tight as those obtained by using $(\tilde{\mathbf{u}}^{\mathrm{A}},\tilde{\mathbf{o}}^{\mathrm{A}})$ in the Scott–Barton method [2].

The following corollary shows that if $\mathbf{f}^{\mathrm{cv}}$ and $\mathbf{f}^{\mathrm{cc}}$ in Assumption 4.2.1 are constructed using McCormick relaxations, then the state relaxations obtained by the new approach are guaranteed to be at least as tight as the SBM relaxations [2].

Corollary 4.3.15. Consider the setup of Corollary 4.3.13, except with $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ chosen to be the generalized McCormick relaxations [6] $(\tilde{\mathbf{u}}^{\mathrm{gMC}},\tilde{\mathbf{o}}^{\mathrm{gMC}})$ for $\mathbf{f}$. Then, $(\mathbf{x}^{\mathrm{cv,A}},\mathbf{x}^{\mathrm{cc,A}})$ and

$(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ are both valid state relaxations for (2.3.1). Moreover, for any $(t, \mathbf{p}) \in I \times P$,

$$\mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p}).$$

Proof. According to Section 4.2 in [2] and Theorem 2.4.32 in [39], Assumptions 2.4.11 and 2.4.12 are satisfied with $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}}) := (\tilde{\mathbf{u}}^{\mathrm{gMC}}, \tilde{\mathbf{o}}^{\mathrm{gMC}})$. The claimed result then follows from Corollary 4.3.13. □

Observe that $(\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}})$ in Corollary 4.3.15 come from the SBM method, while $(\mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$ come from our new relaxation method (4.2.3) with $\mathbf{f}^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi}) := \tilde{\mathbf{u}}^{\mathrm{gMC}}(t,\mathbf{p},\boldsymbol{\xi},\boldsymbol{\xi})$ and $\mathbf{f}^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi}) := \tilde{\mathbf{o}}^{\mathrm{gMC}}(t,\mathbf{p},\boldsymbol{\xi},\boldsymbol{\xi})$. Note that by this definition, $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ are standard McCormick relaxations [5] of $\mathbf{f}$ and satisfy Assumption 4.2.1. Moreover, the tightness result presented in Remark 4.3.14 also applies here, where the generalized McCormick relaxation method is used for constructing $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ based on any state bounds.

Remark 4.3.16. Theorem 4.3.11 and Corollary 4.3.15 together show that, if relaxations that are at least as tight as McCormick relaxations [5] of $\mathbf{f}$ are embedded into our new approach, then our new ODE relaxations are guaranteed to be at least as tight as the SBM relaxations [2]. Thus, our ODE relaxation approach may reduce the number of iterations required by an overarching global dynamic optimization method. Example 4.6 in Section 4.4.2 shows that this is indeed the case for a particular case study. We expect that such a reduction would benefit an overall deterministic global dynamic optimization algorithm.

4.3.6 Convergence as domain shrinks

In this subsection, we show that state relaxations obtained by the new approach converge pointwise quadratically (in the sense of Bompadre and Mitsos [68]) to the solution $\mathbf{x}$ of

(2.3.1) as the parametric domain P is subdivided, provided that the supplied relaxations of $\mathbf{x}_0$ and $\mathbf{f}$ converge analogously. In order to explicitly describe this dependence on parametric domain, we introduce the following extended notation for this subsection.

1. For any $Q \subseteq \mathbb{R}^n$, let $\mathbb{I}Q$ denote the set of all nonempty interval subsets of Q.

2. For any $\tilde{P} \in \mathbb{I}P$, let $\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}(t,\mathbf{p})$ and $\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}(t,\mathbf{p})$ denote state relaxations for (2.3.1) at any $(t,\mathbf{p})$ on $I \times \tilde{P}$, and let $\mathbf{x}^{\mathrm{cv}}_{0,\tilde{P}}(\mathbf{p})$ and $\mathbf{x}^{\mathrm{cc}}_{0,\tilde{P}}(\mathbf{p})$ denote convex and concave relaxations of the initial value function $\mathbf{x}_0$ on $\tilde{P}$, respectively.

3. For any $\tilde{P} \in \mathbb{I}P$, let $\mathbf{x}^{\mathrm{L}}_{\tilde{P}}(t)$ and $\mathbf{x}^{\mathrm{U}}_{\tilde{P}}(t)$ denote LR-analytic state bounds for (2.3.1) at any $t \in I$ based on $\tilde{P}$, and denote the interval $[\mathbf{x}^{\mathrm{L}}_{\tilde{P}}(t), \mathbf{x}^{\mathrm{U}}_{\tilde{P}}(t)]$ as $X(t;\tilde{P})$.

4. Assume for simplicity that D is sufficiently large and there exists a compact $\hat{D} \subsetneq D$ so that for all $t \in I$ and $\tilde{P} \in \mathbb{I}P$, $X(t;\tilde{P}) \subseteq \hat{D}$. Let $\mathbf{f}_W$ denote the original ODE right-hand side function $\mathbf{f}$ from (2.3.1) on the domain $I \times W$, where $W \in \mathbb{I}P \times \mathbb{I}\hat{D}$. Correspondingly, let $\mathbf{f}^{\mathrm{cv}}_W$ and $\mathbf{f}^{\mathrm{cc}}_W$ denote relaxations that satisfy Assumption 4.2.1 on W. Denote the lower and upper bounds of W as $\mathbf{w}^{\mathrm{L}} \in \mathbb{R}^{n_x+n_p}$ and $\mathbf{w}^{\mathrm{U}} \in \mathbb{R}^{n_x+n_p}$, respectively. Then, let $\mathrm{wid}(W) := \max\{w^{\mathrm{U}}_i - w^{\mathrm{L}}_i : 1 \le i \le n_x + n_p\}$ denote the width of W.

5. Let $\mathbf{u}_{\tilde{P}}$ and $\mathbf{o}_{\tilde{P}}$ denote the functions $\mathbf{u}$ and $\mathbf{o}$ with P replaced by $\tilde{P} \in \mathbb{I}P$.

In the language of [79], the following theorem shows that if the inclusion functions $\{\mathbf{x}^{\mathrm{L}}_{\tilde{P}}, \mathbf{x}^{\mathrm{U}}_{\tilde{P}}\}_{\tilde{P}\in\mathbb{I}P}$ have at least first-order convergence in P, uniformly on I, if the scheme of estimators $\{\mathbf{x}^{\mathrm{cv}}_{0,\tilde{P}}, \mathbf{x}^{\mathrm{cc}}_{0,\tilde{P}}\}_{\tilde{P}\in\mathbb{I}P}$ has second-order pointwise convergence in P, and if the scheme of right-hand side relaxations $\{\mathbf{f}^{\mathrm{cv}}_W, \mathbf{f}^{\mathrm{cc}}_W\}_{W\in\mathbb{I}P\times\mathbb{I}\hat{D}}$ has second-order pointwise convergence in $P \times \hat{D}$, uniformly on I, then the scheme of state relaxations $\{\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}, \mathbf{x}^{\mathrm{cc}}_{\tilde{P}}\}_{\tilde{P}\in\mathbb{I}P}$ has second-order pointwise convergence, uniformly on I.

Theorem 4.3.17. Suppose that Assumption 4.2.1 holds with each $\tilde{P} \in \mathbb{I}P$ in place of P, and suppose that all of the following conditions are satisfied.

I. For some $\kappa_0 > 0$, for all $\tilde{P} \in \mathbb{I}P$ and $\mathbf{p} \in \tilde{P}$,

$$\|\mathbf{x}^{\mathrm{cc}}_{0,\tilde{P}}(\mathbf{p}) - \mathbf{x}^{\mathrm{cv}}_{0,\tilde{P}}(\mathbf{p})\|_\infty \leq \kappa_0 \mathrm{wid}(\tilde{P})^2. \tag{4.3.12}$$

II. For some $\kappa_{\mathbf{f}} > 0$, for all $W \in \mathbb{I}P \times \mathbb{I}\hat{D}$ and $(t, \mathbf{p}, \boldsymbol{\xi}) \in I \times W$,

$$\|\mathbf{f}^{\mathrm{cc}}_W(t, \mathbf{p}, \boldsymbol{\xi}) - \mathbf{f}^{\mathrm{cv}}_W(t, \mathbf{p}, \boldsymbol{\xi})\|_\infty \leq \kappa_{\mathbf{f}} \mathrm{wid}(W)^2. \tag{4.3.13}$$

III. For some $\kappa_{\mathrm{B}} > 0$, for all $t \in I$ and $\tilde{P} \in \mathbb{I}P$,

$$\mathrm{wid}(X(t; \tilde{P})) \leq \kappa_{\mathrm{B}} \mathrm{wid}(\tilde{P}). \tag{4.3.14}$$

IV. For some $l > 0$, for all $t \in I$, $\mathbf{p} \in P$, and $\boldsymbol{\xi}^{\mathrm{A}}, \boldsymbol{\xi}^{\mathrm{B}} \in \hat{D}$,

$$\|\mathbf{f}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{A}}) - \mathbf{f}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{B}})\|_\infty \leq l \|\boldsymbol{\xi}^{\mathrm{A}} - \boldsymbol{\xi}^{\mathrm{B}}\|_\infty. \tag{4.3.15}$$

Consider any solution $(\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}, \mathbf{x}^{\mathrm{cc}}_{\tilde{P}})$ of the ODE system (2.4.1) with (4.2.3) embedded, with $\tilde{P}$ in place of P. Then, for all $t \in I$, $\tilde{P} \in \mathbb{I}P$, and $\mathbf{p} \in \tilde{P}$,

$$\|\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}(t, \mathbf{p}) - \mathbf{x}^{\mathrm{cv}}_{\tilde{P}}(t, \mathbf{p})\|_\infty \leq \left(\kappa_0 + 2(t_f - t_0)\kappa_{\mathbf{f}} \max(1, \kappa_{\mathrm{B}}^2)\right) \exp(l(t_f - t_0)) \mathrm{wid}(\tilde{P})^2. \tag{4.3.16}$$

Proof. Consider any fixed $i \in \{1, ..., n_x\}$, $t \in I$, $\tilde{P} \in \mathbb{I}P$, and $\mathbf{p} \in \tilde{P}$ throughout. Since (4.3.13) holds, for any $\tau \in [t_0, t]$ and $\boldsymbol{\xi} \in X(\tau; \tilde{P})$,

$$\begin{aligned} f^{\mathrm{cc}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}) - f^{\mathrm{cv}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}) &\leq \kappa_{\mathbf{f}} \mathrm{wid}(\tilde{P} \times X(\tau;\tilde{P}))^2 \\ &= \kappa_{\mathbf{f}} \max\left(\mathrm{wid}(\tilde{P}), \mathrm{wid}(X(\tau;\tilde{P}))\right)^2. \end{aligned}$$

Furthermore, since (4.3.14) holds, for any $\tau \in [t_0, t]$ and $\boldsymbol{\xi} \in X(\tau; \tilde{P})$,

$$\begin{aligned} f^{\mathrm{cc}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}) - f^{\mathrm{cv}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}) &\leq \kappa_{\mathbf{f}} \max\left(\mathrm{wid}(\tilde{P}), \mathrm{wid}(X(\tau;\tilde{P}))\right)^2 \\ &\leq \kappa_{\mathbf{f}} \max\left(\mathrm{wid}(\tilde{P}), \kappa_{\mathrm{B}}\mathrm{wid}(\tilde{P})\right)^2 \qquad (4.3.17) \\ &= \kappa_{\mathbf{f}} \max\left(1, \kappa_{\mathrm{B}}^2\right)\mathrm{wid}(\tilde{P})^2. \end{aligned}$$

Consider any nonempty interval $[\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}] \subseteq X(\tau; \tilde{P})$. Denote particular optimal solutions of the optimization problems of $u_{i,\tilde{P}}(\tau, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})$ and $o_{i,\tilde{P}}(\tau, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})$ in (4.2.4) as $\boldsymbol{\xi}^{*,u_i}$ and $\boldsymbol{\xi}^{*,o_i}$, respectively. Then,

$$\begin{aligned} o_{i,\tilde{P}}(\tau,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) - u_{i,\tilde{P}}(\tau,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &= f^{\mathrm{cc}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,o_i}) - f^{\mathrm{cv}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,u_i}) \\ &= f^{\mathrm{cc}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,o_i}) - f_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,o_i}) \\ &\quad + \left(f_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,u_i}) - f^{\mathrm{cv}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,u_i})\right) \\ &\quad + \left(f_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,o_i}) - f_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,u_i})\right). \end{aligned} \qquad (4.3.18)$$

Combining (4.3.15), (4.3.17), and (4.3.18), we have that for any $\tau \in [t_0, t]$ and any $[\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}] \subseteq$

$X(\tau;\tilde{P})$,

$$
\begin{aligned}
o_{i,\tilde{P}}(\tau,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})-u_{i,\tilde{P}}(\tau,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &\leq f^{\mathrm{cc}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,o_i})-f^{\mathrm{cv}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,o_i})\\
&\quad+\big(f^{\mathrm{cc}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,u_i})-f^{\mathrm{cv}}_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,u_i})\big)\\
&\quad+\big(f_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,o_i})-f_{i,\tilde{P}\times X(\tau;\tilde{P})}(\tau,\mathbf{p},\boldsymbol{\xi}^{*,u_i})\big)\\
&\leq 2\kappa_{\mathbf{f}}\max(1,\kappa_{\mathrm{B}}^2)\mathrm{wid}(\tilde{P})^2+l\|\boldsymbol{\xi}^{\mathrm{cc}}-\boldsymbol{\xi}^{\mathrm{cv}}\|_\infty.
\end{aligned}
\tag{4.3.19}
$$

Now, we obtain the convergence properties of $\{\mathbf{x}^{\mathrm{cv}}_{\tilde{P}},\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}\}_{\tilde{P}\in\mathbb{I}P}$ using the Gronwall-Bellman Inequality (as presented in [100, Lemma A.1]). Expressing (2.4.1) in integral form yields:

$$
\begin{aligned}
x^{\mathrm{cc}}_{i,\tilde{P}}(t,\mathbf{p})-x^{\mathrm{cv}}_{i,\tilde{P}}(t,\mathbf{p}) &= x^{\mathrm{cc}}_{0,i,\tilde{P}}(\mathbf{p})-x^{\mathrm{cv}}_{0,i,\tilde{P}}(\mathbf{p})+\int_{t_0}^{t}\big(\dot{x}^{\mathrm{cc}}_{i,\tilde{P}}(\tau,\mathbf{p})-\dot{x}^{\mathrm{cv}}_{i,\tilde{P}}(\tau,\mathbf{p})\big)\mathrm{d}\tau\\
&\leq x^{\mathrm{cc}}_{0,i,\tilde{P}}(\mathbf{p})-x^{\mathrm{cv}}_{0,i,\tilde{P}}(\mathbf{p})\\
&\quad+\int_{t_0}^{t}\big(o_{i,\tilde{P}}(\tau,\mathbf{p},\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}(\tau,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}(\tau,\mathbf{p}))-u_{i,\tilde{P}}(\tau,\mathbf{p},\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}(\tau,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}(\tau,\mathbf{p}))\big)\mathrm{d}\tau.
\end{aligned}
\tag{4.3.20}
$$

Since Theorem 4.3.3 shows that $[\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}(\tau,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}(\tau,\mathbf{p})]\subseteq X(\tau;\tilde{P})$ and since (4.3.19) holds, combining (4.3.19) and (4.3.20) yields

$$
\begin{aligned}
x^{\mathrm{cc}}_{i,\tilde{P}}(t,\mathbf{p})-x^{\mathrm{cv}}_{i,\tilde{P}}(t,\mathbf{p}) \leq x^{\mathrm{cc}}_{0,i,\tilde{P}}(\mathbf{p})-x^{\mathrm{cv}}_{0,i,\tilde{P}}(\mathbf{p})&+\int_{t_0}^{t}\big(2\kappa_{\mathbf{f}}\max\big(1,\kappa_{\mathrm{B}}^2\big)\mathrm{wid}(\tilde{P})^2\big)\mathrm{d}\tau\\
&+\int_{t_0}^{t}\big(l\|\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}(\tau,\mathbf{p})-\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}(\tau,\mathbf{p})\|_\infty\big)\mathrm{d}\tau.
\end{aligned}
$$

Furthermore, since (4.3.12) holds, since $\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}(t,\mathbf{p})\geq\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}(t,\mathbf{p})$ as shown in Theorem 4.3.7, and since the inequality above holds for any $i\in\{1,...,n_x\}$,

$$
\begin{aligned}
\|\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}(t,\mathbf{p})-\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}(t,\mathbf{p})\|_\infty &\leq \big(\kappa_0+2(t_f-t_0)\kappa_{\mathbf{f}}\max(1,\kappa_{\mathrm{B}}^2)\big)\mathrm{wid}(\tilde{P})^2\\
&\quad+\int_{t_0}^{t}\big(l\|\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}(\tau,\mathbf{p})-\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}(\tau,\mathbf{p})\|_\infty\big)\mathrm{d}\tau.
\end{aligned}
$$

Applying the Gronwall-Bellman Inequality yields

$$\begin{aligned}\|\mathbf{x}^{\mathrm{cc}}_{\tilde{P}}(t,\mathbf{p})-\mathbf{x}^{\mathrm{cv}}_{\tilde{P}}(t,\mathbf{p})\|_\infty &\leq \left(\kappa_0+2(t_f-t_0)\kappa_{\mathbf{f}}\max(1,\kappa_{\mathrm{B}}^2)\right)\exp(l(t-t_0))\mathrm{wid}(\tilde{P})^2\\ &\leq \left(\kappa_0+2(t_f-t_0)\kappa_{\mathbf{f}}\max(1,\kappa_{\mathrm{B}}^2)\right)\exp(l(t_f-t_0))\mathrm{wid}(\tilde{P})^2.\end{aligned}$$

□

This convergence property will be illustrated in Example 4.2 in Section 4.4.

Remark 4.3.18. The inequality (4.3.19) shows that the functions $(\mathbf{u},\mathbf{o})$ described by (4.2.4) have the *(1,2)-convergence* property of [79, Definition 5.15]. This property was originally established for the generalized McCormick relaxations [6]. Since this beneficial convergence property also holds for our new $(\mathbf{u},\mathbf{o})$, it follows immediately that all benefits of the SBM relaxations [2] established in [79] also hold for our new optimization-based state relaxations. These benefits include: (a) a potentially smaller convergence prefactor of $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$ in (4.3.16), which could aid branch-and-bound convergence, and (b) that the relaxations $(\mathbf{x}^{\mathrm{cv}}(t,\cdot),\mathbf{x}^{\mathrm{cc}}(t,\cdot))$ could get tighter over time t under appropriate conditions. These follow directly from proofs of analogous results in [79], except with the SBM $(\mathbf{u},\mathbf{o})$ functions replaced by our new $(\mathbf{u},\mathbf{o})$ defined in (4.2.4).

Remark 4.3.19. If McCormick relaxations are used for $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$ in Assumption 4.2.1, then the quadratic pointwise convergence of our new state relaxations is directly implied by combining Corollary 4.3.15 and [79, Corollary 5.21]. Corollary 4.3.15 shows that the new state relaxations are at least as tight as the SBM relaxations [2] whose quadratic pointwise convergence was established in [79, Corollary 5.21]. However, Theorem 4.3.17 applies even when non-McCormick relaxations are used for $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$ such as αBB relaxations. This theorem newly demonstrates that quadratic pointwise convergence of $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$ will translate into quadratic pointwise convergence of $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$.

4.3.7 Comparison to established relaxation methods

In this subsection, we compare our new ODE relaxation approach (namely (2.4.1) with (4.2.3) or (4.2.4) embedded) to established ODE relaxation approaches. Firstly, we compare the new approach to the SBM method [2] based on strength of assumptions, tightness, computational complexity, and difficulty of implementation. Then, we present a formulation that embeds αBB relaxations [9] of $\mathbf{f}$ into our new approach, which we compare to the dynamic αBB relaxations proposed by Papamichail and Adjiman [1]. We choose these particular comparisons since these established methods are also based on differential inequalities and interval bounds, and are the most similar to our new method. This similarity enables relatively straightforward apples-to-apples comparisons in which we may assume, for example, that all approaches have access to the same interval bounds $[\mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{U}}(t)]$ for (2.3.1). Lastly, we discuss how our new approach compares with the class of discretize-then-relax approaches.

Comparison to the SBM method

Our new approach ((2.4.1) with $(\mathbf{u}, \mathbf{o})$ defined in (4.2.3)) has less stringent assumptions on relaxations of the original right-hand side $\mathbf{f}$ than the SBM method [2]. While the SBM method only admits generalized McCormick relaxations [6] of $\mathbf{f}$ to construct Scott–Barton right-hand sides $(\mathbf{u}, \mathbf{o})$, our new $(\mathbf{u}, \mathbf{o})$ formulation (4.2.3) admits any convex and concave relaxations of $\mathbf{f}$ that satisfy Assumption 4.2.1, such as McCormick relaxations [5], αBB relaxations [9], affine relaxations, the pointwise tightest among multiple relaxations, and convex envelopes.

Corollary 4.3.15 shows that our new state relaxations are at least as tight as the SBM relaxations [2], when Assumption 4.2.1 is satisfied by applying McCormick relaxations [5] to $\mathbf{f}$. Thus, embedding our new state relaxations into a branch-and-bound procedure for global optimization will yield less conservative lower bounds of global optimal objective values of (1.1.1), which we expect will translate into fewer branch-and-bound iterations required to reach global optimality. Moreover, Theorem 4.3.11 shows that tighter relaxations of $\mathbf{f}$ will always translate into tighter relaxations for $\mathbf{x}$ in our new approach, which incentivizes seeking tighter relaxations for closed-form functions in order to relax dynamic systems. As will be shown in Example 4.6 in Section 4.4.2, embedding convex envelopes of the original right-hand side $\mathbf{f}$ in our new ODE relaxation approach may significantly reduce the number of branch-and-bound iterations, compared to the SBM relaxations, in a global dynamic optimization instance of (1.1.1).

Regarding computational complexity, the SBM method [2] constructs Scott–Barton right-hand sides $(\mathbf{u}, \mathbf{o})$ in (2.4.1) as closed-form functions, while our new approach constructs new $(\mathbf{u}, \mathbf{o})$ as optimal-value functions (4.2.3). When these are evaluated naively using numerical NLP solvers, evaluating relaxations in our new approach will generally be more computationally expensive than the SBM relaxation method, as will be seen in Example 4.4. However, the computational efficiency of our new approach may be further improved. For example, the optimal solutions in the formulation (4.2.3) (as opposed to (4.2.4)) at a time t may still be optimal in the near future. Thus, warm-started optimization could be particularly useful in this setting. Since the optimization problems in (4.2.3) and (4.2.4) are convex, the ODE system (2.4.1) with (4.2.3) or (4.2.4) embedded can be reformulated as an equivalent nonlinear complimentarity system (NCS), using a Karush-Kuhn-Tucker complementarity reformulation of (4.2.3) or (4.2.4). In principle, this NCS could be solved with efficient NCS solvers such as SICONOS [117]. We expect that, if active constraints and optimal solutions in (4.2.3) or (4.2.4) are managed during integration analogously to integrators such as DFBAlab [118], then an optimal implementation of our new approach would be roughly as efficient as the SBM method for evaluating state relaxations. Furthermore, since our new relaxations are tighter in general, we expect that our new approach would ultimately lead to less computational time in deterministic global dynamic optimization.

Implementation of the SBM method requires ODE solvers and a generalized McCormick relaxation package such as the `C++` library `MC++` or EAGO [119] in Julia [95]. A naive implementation of our new approach that solves the optimization problems in (4.2.3) numerically will additionally require a convex NLP solver such as IPOPT [120], `fmincon` in MATLAB for smooth $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$, or Nesterov's level method [40] for nonsmooth $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$. This would be a significant computational expense. Currently, EAGO in Julia is the only open-source branch-and-bound global optimization library in which custom relaxations may be used, but a stripped-down version of IPOPT is currently the only NLP solver available in JuMP [121] in Julia. An active-set tracking approach as discussed in the previous paragraph would remove the need and computational expense of the NLP solver.

Comparison to dynamic αBB relaxations

In this subsection, we summarize an established αBB dynamic relaxation method by Papamichail and Adjiman [1], and then consider applying our new approach by embedding αBB relaxations [9] of the original right-hand side $\mathbf{f}$ into (4.2.3) and (2.4.1). These approaches are then compared.

The αBB class of relaxation methods applies to twice-differentiable nonconvex functions, and constructs valid convex relaxations by adding a convex quadratic term with sufficiently large curvature to the original nonconvex function. Papamichail and Adjiman [1] describe a variant of αBB relaxations that relaxes ODE solutions $\mathbf{x}$ of (2.3.1) as follows. For each $i \in \{1,...,n_x\}$ and each $\mathbf{p} \in [\mathbf{p}^{\mathrm{L}}, \mathbf{p}^{\mathrm{U}}]$, define

$$x_i^{\mathrm{cv}}(t_f,\mathbf{p}) := x_i(t_f,\mathbf{p}) + \sum_{m=1}^{n_p} \alpha_m^i (p_m - p_m^{\mathrm{L}})(p_m - p_m^{\mathrm{U}}), \tag{4.3.21}$$

where the coefficients $\alpha_1^i,...,\alpha_{n_p}^i$ are ultimately determined so that for each $\mathbf{p}$, the Hessian matrix of $x_i^{\mathrm{cv}}(t_f,\mathbf{p})$ with respect to $\mathbf{p}$ is positive semidefinite. To compute appropriate coefficients $\boldsymbol{\alpha}^i$, Papamichail and Adjiman [1] first compute interval bounds of of the Hessian matrix of $x_i(t_f,\cdot)$ on P by applying Harrison's bounding method [69] to the second-order sensitivity system of $\mathbf{x}$ with respect to $\mathbf{p}$, as summarized in [13, Appendix]. With these bounds, appropriate $\boldsymbol{\alpha}^i$ values in (4.3.21) are then computed using the Scaled Gerschgorin Theorem (see [9, Theorem 3.13]). In this chapter, Papamichail and Adjiman's relaxation method will be referred to as the *PA method*.

Our new approach provides another way to use αBB relaxations to generate state relaxations, as follows. Instead of constructing state relaxations directly using αBB relaxations as in (4.3.21), we construct αBB relaxations $\mathbf{f}^{\mathrm{cv}}(t,\cdot,\cdot)$ and $\mathbf{f}^{\mathrm{cc}}(t,\cdot,\cdot)$ for the original right-hand side $\mathbf{f}(t,\cdot,\cdot)$ for each $t \in I$ as follows, according to the approach of [9]. Consider the αBB parameters $\boldsymbol{\alpha}_{\mathbf{p}}^{i,\mathrm{cv}}, \boldsymbol{\alpha}_{\mathbf{p}}^{i,\mathrm{cc}} : I \to \mathbb{R}^{n_p}$ and $\boldsymbol{\alpha}_{\mathbf{x}}^{i,\mathrm{cv}}, \boldsymbol{\alpha}_{\mathbf{x}}^{i,\mathrm{cc}} : I \to \mathbb{R}^{n_x}$ for each $i \in \{1,...,n_x\}$ as functions of t, to be determined subsequently. Then, for each $i \in \{1,...,n_x\}$, each $t \in I$, each $\mathbf{p} \in [\mathbf{p}^{\mathrm{L}}, \mathbf{p}^{\mathrm{U}}]$, and each $\boldsymbol{\xi} \in [\mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{U}}(t)]$, define

$$\begin{aligned}
f_i^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi}) &:= f_i(t,\mathbf{p},\boldsymbol{\xi}) + \sum_{m=1}^{n_p} \alpha_{m,\mathbf{p}}^{i,\mathrm{cv}}(t)(p_m - p_m^{\mathrm{L}})(p_m - p_m^{\mathrm{U}}) \\
&\quad + \Big(\sum_{m=1}^{n_x} \alpha_{m,\mathbf{x}}^{i,\mathrm{cv}}(t)(\xi_m - x_m^{\mathrm{L}}(t))(\xi_m - x_m^{\mathrm{U}}(t)) \Big), \\
f_i^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi}) &:= f_i(t,\mathbf{p},\boldsymbol{\xi}) - \sum_{m=1}^{n_p} \alpha_{m,\mathbf{p}}^{i,\mathrm{cc}}(t)(p_m - p_m^{\mathrm{L}})(p_m - p_m^{\mathrm{U}}) \\
&\quad - \Big(\sum_{m=1}^{n_x} \alpha_{m,\mathbf{x}}^{i,\mathrm{cc}}(t)(\xi_m - x_m^{\mathrm{L}}(t))(\xi_m - x_m^{\mathrm{U}}(t)) \Big).
\end{aligned} \tag{4.3.22}$$

If the αBB parameters $\boldsymbol{\alpha}^i(t) := (\boldsymbol{\alpha}_{\mathbf{p}}^{i,\mathrm{cv}}(t), \boldsymbol{\alpha}_{\mathbf{x}}^{i,\mathrm{cv}}(t), \boldsymbol{\alpha}_{\mathbf{p}}^{i,\mathrm{cc}}(t), \boldsymbol{\alpha}_{\mathbf{x}}^{i,\mathrm{cc}}(t))$ in (4.3.22) are computed for each t by the Scaled Gerschgorin Theorem as described in [9] based on the intervals $[\mathbf{p}^{\mathrm{L}}, \mathbf{p}^{\mathrm{U}}]$ and $[\mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{U}}(t)]$, then $f_i^{\mathrm{cv}}(t,\cdot,\cdot)$ and $f_i^{\mathrm{cc}}(t,\cdot,\cdot)$ are αBB relaxations of $f_i(t,\cdot,\cdot)$ on $P \times X(t)$, and we may then obtain state relaxations in our new approach by embedding (4.3.22) into (4.2.3), which is then embedded into (2.4.1). We may also compute t-invariant αBB parameters $\boldsymbol{\alpha}^i := (\boldsymbol{\alpha}_{\mathbf{p}}^{i,\mathrm{cv}}, \boldsymbol{\alpha}_{\mathbf{x}}^{i,\mathrm{cv}}, \boldsymbol{\alpha}_{\mathbf{p}}^{i,\mathrm{cc}}, \boldsymbol{\alpha}_{\mathbf{x}}^{i,\mathrm{cc}})$ by a simpler approach, when the original right-hand side function $\mathbf{f}$ in (2.3.1) is t-invariant. First, we compute state bounds $\mathbf{x}^{\mathrm{L}}$ and $\mathbf{x}^{\mathrm{U}}$ for (2.3.1) on I using Harrison's bounding method. Then, uniform state bounds $\underline{\mathbf{x}}, \overline{\mathbf{x}} \in \mathbb{R}^{n_x}$ are chosen so that $\underline{\mathbf{x}} \leq \mathbf{x}^{\mathrm{L}}(t) \leq \mathbf{x}^{\mathrm{U}}(t) \leq \overline{\mathbf{x}}$, for all $t \in I$. Finally, t-invariant $\boldsymbol{\alpha}^i$ in (4.3.22) are computed based on the intervals $[\mathbf{p}^{\mathrm{L}}, \mathbf{p}^{\mathrm{U}}]$ and $[\underline{\mathbf{x}}, \overline{\mathbf{x}}]$. These two approaches, based on αBB relaxations of $\mathbf{f}$ with either t-invariant $\boldsymbol{\alpha}^i$ or t-varying $\boldsymbol{\alpha}^i$, will be respectively referred to as the *OB-αTI method* and *OB-αTV method.*

The OB-αTI method uses our new approach, but is similar to the PA method in the sense that both of them require computing $\boldsymbol{\alpha}$ once for use in (4.3.21) and (4.3.22). For a fair comparison in this chapter, we consider both methods to employ natural interval extension [48] and Harrison's bounding method [69] for computing interval bounds whenever necessary, and to employ the Scaled Gerschgorin Theorem for computing parameters $\boldsymbol{\alpha}$. Since the two methods differ in structure, it is difficult to draw a general conclusion about which method is more efficient in general, and which method would yield tighter state relaxations. However, as will be shown in Example 4.2, in one instance of (2.3.1), our new OB-αTI method requires less computational time for evaluating state relaxations, and also yields tighter state relaxations than the PA method.

Intuitively, the OB-αTV method would yield tighter state relaxations than the OB-αTI method, since for each $t \in I$, the OB-αTV method computes less conservative $\boldsymbol{\alpha}(t)$ for use in (4.3.22) based on $[\mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{U}}(t)]$, while the OB-$\alpha$TI method computes a constant $\boldsymbol{\alpha}$ throughout the time horizon, based on the more conservative bounds $[\underline{\mathbf{x}}, \overline{\mathbf{x}}]$. On the other hand, the OB-αTI method would be less computationally expensive in general than the OB-αTV method, since the OB-αTV method must compute values of $\boldsymbol{\alpha}$ for each time step.

Comparison to discretize-then-relax approaches

Unlike the discretize-then-relax approaches outlined in Section 4.1, our new approach constructs state relaxations using an auxiliary ODE system. Thus, the new approach is able to exploit the adaptive time-stepping and error control of numerical ODE solvers when evaluating state relaxations. Moreover, compared to the Auxiliary Variable Method [50], our new approach avoids including auxiliary decision variables and constraints when discretizing the original ODE system (2.3.1). The new approach thereby does not enlarge the lower bounding problem of (1.1.1) in branch-and-bound, which may be advantageous for an overarching global optimization method.

4.4 Implementation and examples

4.4.1 Implementation

A proof-of-concept implementation was developed in MATLAB R2019a to construct and compute state relaxations for (2.3.1) by solving the auxiliary ODE system (2.4.1) with $(\mathbf{u},\mathbf{o})$ provided by (4.2.3) (available at https://github.com/kamilkhanlab/ob-ode-relaxations). In this implementation, Harrison's bounding method [69] is employed via operator overloading to compute state bounds $(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}})$ automatically. McCormick relaxations [5] were implemented via operator overloading, for use in automatically constructing $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$ that satisfy Assumption 4.2.1. For any inputs $(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \in I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ and any user-defined $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$, the optimal-value functions $\mathbf{u}$ and $\mathbf{o}$ in (2.4.1) are evaluated naively by solving the convex optimization problems in (4.2.3) using the local optimization solver `fmincon` with an optimality tolerance of 10^{-6}. Finally, the auxiliary ODE system (2.4.1) with $\mathbf{u}$ and $\mathbf{o}$ embedded is solved using the ODE solver `ode23` with an absolute tolerance of 10^{-4} and a relative tolerance of 10^{-4}. An analogous implementation for the SBM relaxation method [2] was also developed for comparison. All computation in this section was performed on a Dell desktop computer with two 3.00 GHz Intel Core i7-9700 CPUs and 16.0 GB of RAM.

4.4.2 Numerical examples

This subsection presents numerical examples that compare our new approach to the SBM method [2], the dynamic αBB relaxation method [1] discussed in Section 4.3.7, and a discretize-then-relax approach employing the state-of-the-art global optimization solver

BARON [60]. Table 4.1 lists the various relax-then-discretize approaches that will be compared in this subsection, along with abbreviations for each approach that will be employed from here on.

Table 4.1: A list of considered state relaxation methods and their abbreviations.

Method	Abbreviation
From Scott and Barton [2] with generalized McCormick relaxations	SBM
From Papamichail and Adjiman [1]	PA
New optimization-based method with embedded convex envelopes	OB-ENV
New method with embedded McCormick relaxations	OBM
New method with embedded αBB relaxations and time-varying α	OB-αTV
New method with embedded αBB relaxations and constant α	OB-αTI

Firstly, the following example illustrates the versatility of our new relaxation formulation, by showing that, unlike the SBM relaxations [2], we may use the αBB relaxation method [9] to construct right-hand side relaxations $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ according to (4.3.22). In this example, our new OB-αTI approach requires less computational time, and yields significantly tighter ODE convex relaxations than the PA method [1]. This example also illustrates the quadratic pointwise convergence of our new state relaxations established in Theorem 4.3.17.

Example 4.2. Let $P := [-2,2]$, and $I := [0,0.15]$, and consider the following instance of (2.3.1) with one state variable x and one parameter $p \in P$:

$$\begin{aligned} \dot{x}(t) &= p(x^2-1), \quad t \in I, \\ x(0) &= -2. \end{aligned} \tag{4.4.1}$$

The value $t_f := 0.15$ was chosen arbitrarily. As a function of p, $x(0.15,\cdot)$ is a nontrivial nonlinear function, as shown in Figure 4.1. Since $n_x = 1$ in this example, the right-hand side functions $\mathbf{u}$ and $\mathbf{o}$ from (4.2.3) were constructed in closed form using (4.2.5). The functions $(f^{\mathrm{cv}}, f^{\mathrm{cc}})$ in (4.2.5) were constructed using αBB relaxations both with t-varying $\boldsymbol{\alpha}$ and t-invariant $\boldsymbol{\alpha}$ following the procedure in Section 4.3.7. The Hessian system of $x(t,p)$ in (4.4.1) with respect to p was hard-coded in MATLAB, according to [13, Appendix]. Harrison's bounding method was applied to this system to obtain the Hessian bounds. Thus, state relaxations of the solution $x(0.15,\cdot)$ of (4.4.1) on P were generated numerically in MATLAB by the OB-αTI method, OB-αTV method, and PA method [1], as summarized in Section 4.3.7. When implementing the OB-αTV and OB-αTI methods, we integrated Harrison's state bounds and state relaxations for (4.4.1) at all mesh points $p_i \in P$ simultaneously, since doing so avoids repeatedly computing state bounds when evaluating each $(x^{\mathrm{cv}}(0.15,p_i), x^{\mathrm{cc}}(0.15,p_i))$. Table 4.2 presents the resulting CPU times for evaluating $(x^{\mathrm{cv}}(0.15,p_i), x^{\mathrm{cc}}(0.15,p_i))$ at one mesh point p_i by the three ODE relaxation methods. For the OB-αTV and OB-αTI methods, per-mesh-point CPU times were obtained by taking the total evaluation time for all mesh points p_i simultaneously, dividing this by the number of mesh points, and averaging this figure over ten runs. For both the PA and OB-αTI methods, the total CPU time includes the time for computing the constant $\boldsymbol{\alpha}$ in (4.3.21) and (4.3.22), respectively. Figure 4.1 depicts the resulting relaxations of $x(0.15,\cdot)$, along with the original ODE solution for comparison. Observe that, for both the OB-αTI and OB-αTV methods, the generated relaxations do indeed appear to be valid convex underestimators and concave overestimators, as is guaranteed by Theorems 4.3.7 and 4.3.9. Furthermore, the OB-αTV relaxations are evidently tighter than the OB-αTI relaxations, but require more computational effort to evaluate as shown in Table 4.2. This is to be expected, since the OB-αTV method updates values of $\boldsymbol{\alpha}$ for each right-hand side evaluation, but these time-varying $\boldsymbol{\alpha}$ are less conservative than the constant $\boldsymbol{\alpha}$ values of the OB-αTI method, which ultimately translate into tighter state relaxations through (4.3.22), (4.2.5), and (2.4.1). Observe that in this case, the new OB-αTI convex relaxations are significantly tighter than the PA convex relaxations [1], and also require less computational time than the PA relaxations as shown in Table 4.2. This could be because evaluating the Hessian bounds of $x(t,p)$ in (4.4.1) with respect to p is computationally expensive. We nevertheless note that since the PA method and the OB-αTI method differ so much in structure, it is difficult here to draw a general conclusion.

Table 4.2: Average computational times for evaluating state relaxations $(x^{\mathrm{cv}}(0.15,p), x^{\mathrm{cc}}(0.15,p))$ for (4.4.1) in Example 1.

State relaxation method	CPU time (seconds) *
OB-αTV	0.0006
OB-αTI	0.0002
PA	0.0006

* Each CPU time here was averaged over 10 runs.

To illustrate convergence properties of the new OB-αTV relaxations to (4.4.1) as the parametric subdomain shrinks, we compute state relaxations $\{x^{\mathrm{cv}}_{P(k)}, x^{\mathrm{cc}}_{P(k)}\}$ at $t := 0.15$ on $P(k) := [-1, -1+2^{k+1}]$ for each $k \in \{-10, -9, ..., 0\}$ by the OB-αTV method. Extending the "width" concept from Section 4.3.6, define the width of the enclosure X^{C}_k formed by the state relaxations on each parametric subinterval $P(k)$ as $\mathrm{wid}(X^{\mathrm{C}}_k) := \max_{\mathbf{p}\in P(k)} \left(x^{\mathrm{cc}}_{P(k)}(0.15,p) - x^{\mathrm{cv}}_{P(k)}(0.15,p)\right)$. Figure 4.2 plots $\mathrm{wid}(X^{\mathrm{C}}_k)$ against $\mathrm{wid}(P_k)$, and shows that the new αBB state relaxations do indeed exhibit second-order pointwise convergence in this case. This is guaranteed by Theorem 4.3.17, since αBB relaxations of closed-form smooth functions have second-order pointwise convergence [36].

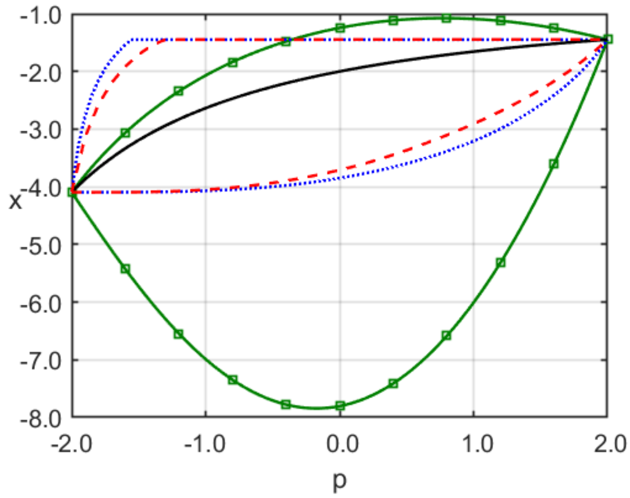

Figure 4.1: The solution $x(0.15, p)$ (*solid black*) of the parametric ODE (4.4.1) from Example 4.2, plotted against p along with corresponding state relaxations obtained by the OB-αTI method (*dotted blue*), the OB-αTV method (*dashed red*), and the PA method [1] (*squared green*).

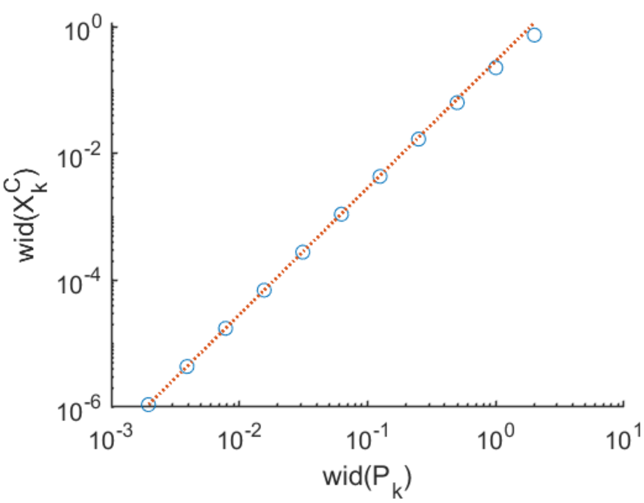

Figure 4.2: A plot on logarithmic axes of $\mathrm{wid}(X_k^{\mathrm{C}}) := \sup_{\mathbf{p}\in P(k)} \left(x_{P(k)}^{\mathrm{cc}}(0.15, p) - x_{P(k)}^{\mathrm{cv}}(0.15, p)\right)$ vs. $\mathrm{wid}(P(k)) := 2^{k+1}$ (*blue circles*) for Example 4.2, for $k := -10, -9, ..., 0$, with $(\mathbf{x}_{P(k)}^{\mathrm{cv}}, \mathbf{x}_{P(k)}^{\mathrm{cc}})$ generated by the OB-αTV method, along with a reference line (*dotted red*) corresponding to second-order pointwise convergence.

In the following example, we show that the discrepancy between our new ODE relaxations and the original ODE system may decrease over time, as discussed in Remark 4.3.18.

Example 4.3. Let $P := [-1.2, 1.2]$ and $I := [0, 0.7]$, and consider the following instance of (2.3.1) with one state variable x and one parameter $p \in P$:

$$\begin{aligned} \dot{x}(t) &= x^2 - e^x, \quad \forall t \in I, \\ x(0) &= p - \frac{p^3}{3}. \end{aligned} \tag{4.4.2}$$

The functions x_0^{cv} and x_0^{cc} in (2.4.1) were constructed as McCormick relaxations [5, 35] of the initial-value function $x_0(p) := p - \frac{p^3}{3}$ from (4.4.2), with the known convex and concave envelopes for $p \mapsto p^3$ [52] employed. For each $t \in [0.1, 0.7]$, state relaxations $x^{\mathrm{cv}}(t, \cdot)$ and $x^{\mathrm{cc}}(t, \cdot)$ for the solution $x(t, \cdot)$ of the ODE (4.4.2) were generated numerically on P by the new OB-αTV method in MATLAB. Thus, define the convex ODE relaxation discrepancy $\varepsilon^{\mathrm{cv}}$ as

$$\varepsilon^{\mathrm{cv}}(t) := \max_{p \in P} \left(x(t, p) - x^{\mathrm{cv}}(t, p) \right),$$

and define the concave ODE relaxation discrepancy $\varepsilon^{\mathrm{cc}}$ as

$$\varepsilon^{\mathrm{cc}}(t) := \max_{p \in P} \left(x^{\mathrm{cc}}(t, p) - x(t, p) \right).$$

Figure 4.3 depicts $\varepsilon^{\mathrm{cv}}$ and $\varepsilon^{\mathrm{cc}}$. We can see that as t increases, both $\varepsilon^{\mathrm{cv}}$ and $\varepsilon^{\mathrm{cc}}$ initially increase and then decrease with respect to t.

In the following example, we show that if McCormick relaxations [5] of the original right-hand side $\mathbf{f}$ are used to define $\mathbf{f}^{\mathrm{cv}}$ and $\mathbf{f}^{\mathrm{cc}}$ in (4.2.3), then the corresponding new OBM approach yields state relaxations that are at least as tight as the SBM relaxations [2], as

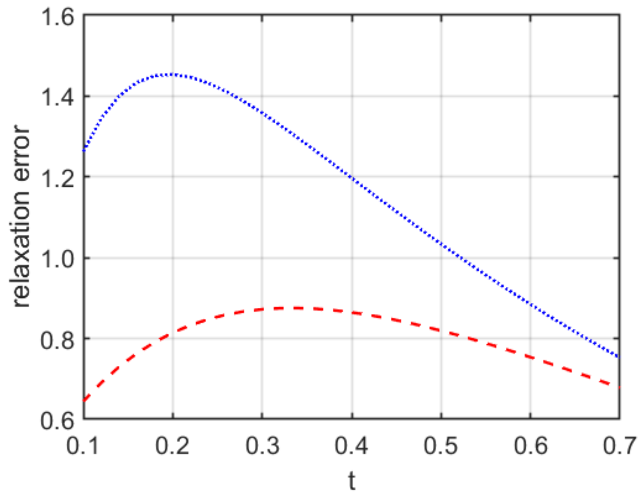

Figure 4.3: A plot of the convex state relaxation discrepancy $\varepsilon^{\mathrm{cv}}(t) := \max_{p\in P}\big(x(t,p) - x^{\mathrm{cv}}(t,p)\big)$ (*dotted blue*) and the concave state relaxation discrepancy $\varepsilon^{\mathrm{cc}}(t) := \max_{p\in P}\big(x^{\mathrm{cc}}(t,p) - x(t,p)\big)$ (*dashed red*) vs. time $t \in [0.1, 0.7]$, with $(x^{\mathrm{cv}}, x^{\mathrm{cc}})$ generated by the OB-αTV method, for Example 4.3.

guaranteed by Corollary 4.3.15.

Example 4.4. Consider the following variant of an established anaerobic digestion process model [81, 122]. Let $\mathbf{p} := (p_1, p_2, ..., p_8)$ denote parameters with known bounds listed in Table 4.3. Let $I := [0,2]$, and consider the following problem instance of (2.3.1):

$$\begin{aligned}
\dot{x}_1(t) &= (\mu_1(t) - 0.2)x_1(t), & x_1(0) &= 0.5, \\
\dot{x}_2(t) &= (\mu_2(t) - 0.2)x_2(t), & x_2(0) &= p_8, \\
\dot{x}_3(t) &= 0.4(5 - x_3(t)) - p_1\mu_1(t)x_1(t), & x_3(0) &= 1, \\
\dot{x}_4(t) &= 0.4(80 - x_4(t)) + p_2\mu_1(t)x_1(t) - p_3\mu_2(t)x_2(t), & x_4(0) &= 5, \\
\dot{x}_5(t) &= -0.4x_5(t) - q(t) + p_4\mu_1(t)x_1(t) + p_5\mu_2(t)x_2(t), & x_5(0) &= 40,
\end{aligned} \tag{4.4.3}$$

with

$$\mu_1(t) = \frac{1.2x_3(t)}{x_3(t)+7.1}, \qquad \mu_2(t) = \frac{0.74x_4(t)}{x_4(t)+p_7+(x_4(t))^2/256},$$
$$\phi(t) = x_5(t)+x_4(t)-34+\frac{p_6\mu_2(t)x_4(t)}{19.8}, \qquad q(t) = 19.8(x_5(t)+x_4(t)-50-0.5\phi(t)).$$

Table 4.3: The interval bounds of uncertain parameters $\mathbf{p}$ in Example 4.4.

i	Lower bounds of p_i	Upper bounds of p_i
1	22.14	62.14
2	80.0	146.5
3	238	298
4	30.6	70.6
5	313.6	373.6
6	423	483
7	4.28	14.28
8	0.84	1.16

For the solution $\mathbf{x}(2,\cdot)$ of (4.4.3) at $t_f = 2$, our new OBM relaxations were generated numerically by applying our MATLAB implementation, and the SBM relaxations were generated analogously for comparison. Figures 4.4 and 4.5 present two cross-sectional plots, comparing the new OBM relaxations, the SBM relaxations, and the original ODE solution. Table 4.4 summarizes the resulting CPU times for evaluating state relaxations $(\mathbf{x}^{\mathrm{cv}}(2,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(2,\mathbf{p}))$ for (4.4.3) at one $\mathbf{p}$, both using the SBM method and the OBM method. Observe that in both Figures 4.4 and 4.5, our new OBM relaxations are at least as tight as the SBM relaxations; this was shown to hold generally in Corollary 4.3.15. However, the new OBM relaxations took longer to evaluate than the SBM relaxations as shown in Table 4.4, largely because the proof-of-concept implementation of the OBM method described in Section 4.4.1 naively solves convex optimization problems with NLP solvers, at each evaluation. We expect that this implementation may be improved with the techniques outlined in Section 4.3.7.

Table 4.4: Average computational times for evaluating state relaxations $(\mathbf{x}^{cv}(2,\mathbf{p}),\mathbf{x}^{cc}(2,\mathbf{p}))$ for (4.4.3) in Example 4.4.

State relaxation method	CPU time (seconds) *
OBM	327.0
SBM	2.6

* Each CPU time here was averaged over 10 runs.

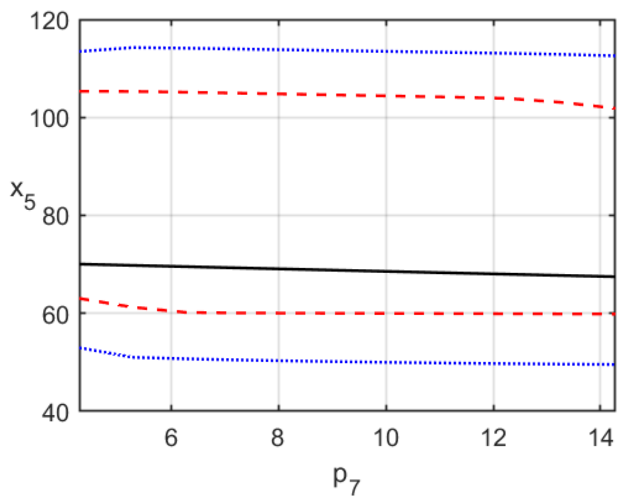

Figure 4.4: A cross-section at $(p_1, p_2, p_3, p_4, p_5, p_6, p_8) := (42.14, 116.5, 269, 50.6, 343.6, 450, 1)$ of the solution $x_5(2,\cdot)$ (*solid black*) of the ODE system (4.4.3) from Example 4.4, along with corresponding state relaxations obtained by the new OBM method (*dashed red*) and by the SBM method [2] (*dotted blue*).

The following example shows that our new approach can employ convex envelopes for $\mathbf{f}$ to yield state relaxations that are at least as tight as the SBM relaxations [2], as discussed in Remark 4.3.12. For comparison, we also applied the OB-αTI method, the OB-αTV method, and the PA method [1] as summarized in Section 4.3.7 to this example. In this

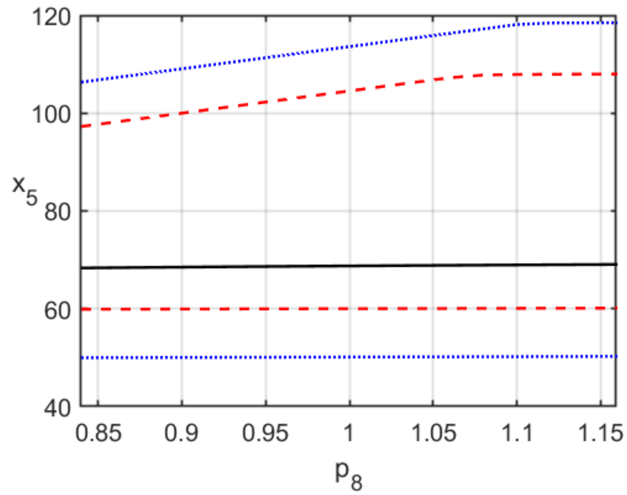

Figure 4.5: A cross-section at $(p_1, p_2, p_3, p_4, p_5, p_6, p_7) := (42.14, 116.5, 269, 50.6, 343.6, 450, 9.28)$ of the solution $x_5(2,\cdot)$ (*solid black*) of the ODE system (4.4.3) from Example 4.4, along with corresponding state relaxations obtained by the new OBM method (*dashed red*) and by the SBM method [2] (*dotted blue*).

case, the OB-αTI relaxations and OB-αTV relaxations ultimately reduce to the predefined Harrison state bounds, and the PA method ultimately fails, since Harrison interval bounds of the Hessian of the ODE solution x explode as t increases.

Example 4.5. For $P := [-1.2, -0.2]$ and $I := [0, 0.9]$, consider the following instance of (2.3.1) with one state variable x and one parameter $p \in P$:

$$\begin{aligned} \dot{x}(t) &= x^4 - 3x^2 - x + 0.4, \quad t \in I, \\ x(0) &= p - \frac{p^3}{3}. \end{aligned} \tag{4.4.4}$$

Define a set $S := \{(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) \in \mathbb{R}^3 : -1.5 \leq \xi^{\mathrm{L}} \leq \xi \leq \xi^{\mathrm{U}} \leq 0.5\}$. With $(x^{\mathrm{L}}, x^{\mathrm{U}})$ denoting Harrison state bounds for (4.4.4), it is empirically verified that for each $t \in I$, $-1.5 \leq x^{\mathrm{L}}(t) \leq x^{\mathrm{U}}(t) \leq 0.5$. Consider the first-order derivative function $f' : \xi \mapsto 4\xi^3 - 6\xi - 1$ of the right-hand side $f : \xi \mapsto \xi^4 - 3\xi^2 - \xi + 0.4$ of (4.4.4). Following the procedures presented in [5,51], we constructed functions $f^{\mathrm{cv,env}}, f^{\mathrm{cc,env}} : S \to \mathbb{R}$ so that for any $[\xi^{\mathrm{L}}, \xi^{\mathrm{U}}] \subseteq [-1.5, 0.5]$, the mappings $f^{\mathrm{cv,env}}(\cdot, \xi^{\mathrm{L}}, \xi^{\mathrm{U}})$ and $f^{\mathrm{cc,env}}(\cdot, \xi^{\mathrm{L}}, \xi^{\mathrm{U}})$ are respectively convex and concave envelopes of f on $[\xi^{\mathrm{L}}, \xi^{\mathrm{U}}]$. For each $(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) \in S$, $f^{\mathrm{cv,env}}$ and $f^{\mathrm{cc,env}}$ are thus evaluated as follows:

- if $\xi^{\mathrm{L}} \geq -\frac{\sqrt{2}}{2}$,

$$\begin{aligned} f^{\mathrm{cv,env}}(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) &:= \frac{f(\xi^{\mathrm{U}}) - f(\xi^{\mathrm{L}})}{\xi^{\mathrm{U}} - \xi^{\mathrm{L}}}(\xi - \xi^{\mathrm{L}}) + f(\xi^{\mathrm{L}}), \\ f^{\mathrm{cc,env}}(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) &:= f(\xi), \end{aligned}$$

- if $\xi^{\mathrm{U}} \leq -\frac{\sqrt{2}}{2}$,

$$\begin{aligned} f^{\mathrm{cv,env}}(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) &:= f(\xi), \\ f^{\mathrm{cc,env}}(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) &:= \frac{f(\xi^{\mathrm{U}}) - f(\xi^{\mathrm{L}})}{\xi^{\mathrm{U}} - \xi^{\mathrm{L}}}(\xi - \xi^{\mathrm{L}}) + f(\xi^{\mathrm{L}}), \end{aligned}$$

- if $\xi^{\mathrm{L}} < -\frac{\sqrt{2}}{2} < \xi^{\mathrm{U}}$, with $\xi^{\mathrm{A}} := \frac{-\xi^{\mathrm{U}} - \sqrt{9 - 2(\xi^{\mathrm{U}})^2}}{3}$ and $\xi^{\mathrm{B}} := \frac{-\xi^{\mathrm{L}} - \sqrt{9 - 2(\xi^{\mathrm{L}})^2}}{3}$,

$$f^{\mathrm{cv,env}}(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) := \begin{cases} \frac{f(\xi^{\mathrm{U}}) - f(\xi^{\mathrm{L}})}{\xi^{\mathrm{U}} - \xi^{\mathrm{L}}}(\xi - \xi^{\mathrm{L}}) + f(\xi^{\mathrm{L}}), & \text{if } \xi^{\mathrm{A}} \leq \xi^{\mathrm{L}}, \\ f(\xi), & \text{if } \xi^{\mathrm{A}} > \xi^{\mathrm{L}} \text{ and } \xi \leq \xi^{\mathrm{A}}, \\ f'(\xi^{\mathrm{A}})(\xi - \xi^{\mathrm{A}}) + f(\xi^{\mathrm{A}}), & \text{if } \xi^{\mathrm{A}} > \xi^{\mathrm{L}} \text{ and } \xi > \xi^{\mathrm{A}}, \end{cases}$$

$$f^{\mathrm{cc,env}}(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) := \begin{cases} \frac{f(\xi^{\mathrm{U}}) - f(\xi^{\mathrm{L}})}{\xi^{\mathrm{U}} - \xi^{\mathrm{L}}}(\xi - \xi^{\mathrm{L}}) + f(\xi^{\mathrm{L}}), & \text{if } \xi^{\mathrm{B}} \geq \xi^{\mathrm{U}}, \\ f(\xi) & \text{if } \xi^{\mathrm{B}} < \xi^{\mathrm{U}} \text{ and } \xi > \xi^{\mathrm{B}}, \\ f'(\xi^{\mathrm{B}})(\xi - \xi^{\mathrm{B}}) + f(\xi^{\mathrm{B}}), & \text{if } \xi^{\mathrm{B}} < \xi^{\mathrm{U}} \text{ and } \xi \leq \xi^{\mathrm{B}}. \end{cases}$$

Then, appropriate functions $(f^{\mathrm{cv}}, f^{\mathrm{cc}})$ for use in (4.2.5) were constructed by setting

$$f^{\mathrm{cv}}(t,p,\xi) \equiv f^{\mathrm{cv,env}}(\xi, x^{\mathrm{L}}(t), x^{\mathrm{U}}(t)) \text{ and } f^{\mathrm{cc}}(t,p,\xi) \equiv f^{\mathrm{cc,env}}(\xi, x^{\mathrm{L}}(t), x^{\mathrm{U}}(t)).$$

The functions x_0^{cv} and x_0^{cc} in (2.4.1) were constructed as McCormick relaxations [5, 35] of the initial-value function $x_0(p) \equiv p - \frac{p^3}{3}$ from (4.4.4), with the known convex and concave envelopes for $p \mapsto p^3$ [52] employed. State relaxations for the solution $x(0.9,\cdot)$ of the ODE (4.4.4) were generated numerically on P by the *OB-ENV method* (in which our new approach has $(f^{\mathrm{cv,env}}, f^{\mathrm{cc,env}})$ embedded), the SBM method [2], the new OB-αTI method, and the new OB-αTV method in MATLAB. Note that since $n_x = 1$, the OBM method here reduces to the SBM method as discussed in Remark 4.2.5. In this case, the PA method [1] failed to give valid state relaxations, since the required Harrison interval bounds of the Hessian of x in (4.4.4) exploded as t increased. Table 4.5 summarizes the average CPU times for evaluating state relaxations $(x^{\mathrm{cv}}(0.9,p), x^{\mathrm{cc}}(0.9,p))$ for (4.4.4) at one p, using each of these methods. Figure 4.6 presents the corresponding state relaxations, along with the original ODE solution $x(0.9,\cdot)$. Observe that the OB-ENV relaxations are tighter than the the SBM relaxations, which is guaranteed in general as discussed in Remark 4.3.12.The OB-ENV relaxations also required significantly less computational time to evaluate than the SBM relaxations, as shown in Table 4.5. One reason for this may be that the envelopes $(f^{\mathrm{cv,env}}, f^{\mathrm{cc,env}})$ could be entered in closed-form, while constructing McCormick relaxations of the original right-hand side of (4.4.4) required access to operator overloading. We also note that both the OB-αTV relaxations and the OB-αTI relaxations reduce to Harrison state bounds in this example. The reason is that for any $p \in P$, our new αBB-based ODE relaxation $x^{\mathrm{cv}}(\tau,p)$ coincided with $x^{\mathrm{L}}(\tau)$ at some $\tau < t_f$, and then $x^{\mathrm{cv}}(t,p)$ remained identical with $x^{\mathrm{L}}(t)$ for $t > \tau$ due to the if-statements in (2.4.1). A similar situation arose with x^{cc} and x^{U}.

Table 4.5: Average computational times for evaluating state relaxations $(x^{\mathrm{cv}}(0.9,p),x^{\mathrm{cc}}(0.9,p))$ for (4.4.4) in Example 4.5.

State relaxation method	CPU time (seconds) *
OB-αTI	0.008
OB-αTV	0.010
OB-ENV	0.004
SBM	0.012
PA	failed

* Each CPU time here was averaged over 10 runs.

In the following example, we describe how a dynamic optimization problem instance was solved to global optimality in Julia v1.4.2 [95], both with the OB-ENV relaxations and the SBM relaxations [2] for an embedded ODE system. This implementation only addresses the $n_x = 1$ case, in which Remark 4.2.4 shows that (4.2.3) reduces to the simpler (4.2.5). The results show that, to reach global optimality in this case, the new OB-ENV relaxations require significantly fewer branch-and-bound iterations than the SBM relaxations, for this example. The same problem instance was also supplied to the state-of-the-art global optimization solver BARON v19.12.7 [60], under the Auxiliary Variable Method [50].

Example 4.6. Consider the following instance of the global dynamic optimization problem (1.1.1) with (4.4.4) embedded:

$$\min_{p\in[-1.2,-0.2]} \; -3(x(0.9,p))^3+(1+p)x(0.9,p) \tag{4.4.5}$$

A proof-of-concept implementation for solving (4.4.5) to global optimality was developed in Julia v1.4.2 [95]. EAGO v0.4.1 [119] was used to apply a branch-and-bound framework [27], without any range reduction. As described in [27], a branch-and-bound method

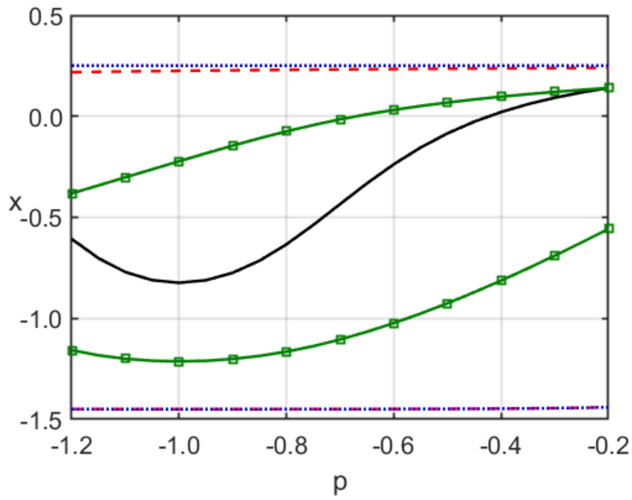

Figure 4.6: The solution $x(0.9,\cdot)$ (*solid black*) of the ODE (4.4.4) on P from Example 4.5, along with corresponding state relaxations obtained by the OB-ENV method (*squared green*), the SBM method [2] (*dashed red*), and the OB-αTV method (*dotted blue*). (In this example, the OB-αTI relaxations overlap with the OB-αTV relaxations, and the OB-αTV convex relaxations overlap with the SBM convex relaxations).

computes upper and lower bounds of the globally optimal objective value and progressively refines these bounds as the decision space is subdivided. EAGO is currently the only open-source branch-and-bound framework that admits user-defined upper and lower bounding procedures. In our implementation, on any subinterval $\tilde{P} \subseteq [-1.2, -0.2]$, the upper bounding procedure solves the problem (4.4.5) locally using IPOPT [120], and the lower bounding procedure is as follows:

1. Compute state relaxations for (4.4.4) on $\hat{P}$ using the OB-ENV method or the SBM method [2].

2. Based on these state relaxations, compute a convex relaxation of the mapping $p \mapsto -3(x(0.9,p))^3 + (1+p)x(0.9,p)$ on $\hat{P}$ using the generalized McCormick relaxation method [6].

3. Minimize this convex relaxation on $\hat{P}$, yielding a valid lower bound of globally optimal objective values of (4.4.5).

In the upper and lower bounding procedures above, we employ EAGO to compute natural interval extensions [48] and generalized McCormick relaxations [6] when necessary, we employ the ODE solver `BS3()` from the package DifferentialEquations v6.15.0 [123] to solve related ODE systems, and we employ JuMP v0.21.3 [121] as an interface with the local NLP solver IPOPT v3.13.2 [120], which is used for all local minimization.

While IPOPT was developed to solve smooth NLPs, it was applied here even for the nonsmooth lower bounding NLPs, as it is the only NLP solver implemented in JuMP. We observed that, when attempting to minimize nonsmooth convex relaxations, IPOPT usually iterated from initial points to nonsmooth optimal points within several steps, and then often remained at or near the optimal points without terminating, perhaps unsuccessfully reducing the dual infeasibility due to the nonsmoothness. In this proof-of-concept implementation, we sidestepped this issue by permitting at most 10 iterations, and considering the objective value at this point to be acceptable. All non-default IPOPT settings in our implementation are listed in Table 4.6, and were adapted from [124] wherein Watson et al. applied IPOPT to a nondifferentiable model of a gas liquefaction process. Any necessary gradients were approximated by the centered finite difference method with a step length of 10^{-6}. The absolute and relative convergence tolerances of EAGO's branch-and-bound were set to be 10^{-4} and 10^{-3}, respectively. Both absolute and relative tolerances of the ODE solver `BS3()` were set to be 10^{-8}.

Using this implementation, we solved the problem (4.4.5) to numerical global optimality both with the OB-ENV and SBM methods [2] used to relax (4.4.4). Table 4.7

Table 4.6: Non-default IPOPT options specified for nonsmooth convex minimization, for Example 4.6.

IPOPT option	Value
`tol`	10^{-5}
`max_iter`	10
`mu_strategy`	`adaptive`
`hessian_approximation`	`limited-memory`

summarizes the corresponding CPU times and numbers of iterations required for branch-and-bound, averaged over ten runs. Observe that the new OB-ENV method required less CPU time and about 30% fewer iterations than the SBM method [2] to obtain the globally optimal solution.

Table 4.7: Global optimization results for the problem (4.4.5) in Example 4.6.

State relaxation method	Glob. optim. obj. *	CPU time (seconds) *	# iterations in B&B *
OB-ENV	-0.06068	7.3	23
SBM	-0.06068	9.7	33

* Each number here is the average of 10 runs. The abbreviation "glob. optim. obj." stands for "global optimal objective value", and "B&B" stands for "branch-and-bound".

For comparison, we also attempted to solve (4.4.5) with BARON v19.12.7 [60] in GAMS v30.3.0, using a discretize-then-relax approach. We discretized the embedded parametric ODE system (4.4.4) by the forward Euler method with predefined time steps $t_0, t_1, ..., t_k$ and an even step size $\Delta t := \frac{0.9}{k}$, and included auxiliary variables $\xi_0, \xi_1, ..., \xi_k$ with bounds $[-1.5, 0.5]$ at each t_i to represent the state x, according to [35]. Thus, the dynamic optimization problem (4.4.5) was approximated by the NLP:

$$\begin{aligned}
\min_{p,\xi_0,\xi_1,\ldots,\xi_k} \quad & -3\xi_k^3+(1+p)\xi_k \\
\text{s.t.} \quad & \xi_j=\xi_{j-1}+\Delta t(\xi_{j-1}^4-3\xi_{j-1}^2-\xi_{j-1}+0.4), \quad \forall j\in\{1,\ldots,k\}, \\
& \xi_0=p-\frac{p^3}{3}, \\
& -1.2\le p\le -0.2, \\
& -1.5\le \xi_j\le 0.5, \quad \forall j\in\{1,\ldots,k\}.
\end{aligned} \tag{4.4.6}$$

This NLP was formulated in GAMS and solved with BARON both with $k=10$ and $k=100$, and both with and without range reduction. Recall that our global dynamic optimization implementation in Julia did not incorporate range reduction, for a fair comparison between different ODE relaxation methods. The corresponding computational results for these solves are summarized in Table 4.8. Observe that with range reduction, for both $k=10$ and $k=100$, BARON converged to a globally optimal solution for the approximation (4.4.6) of (4.4.5). Moreover, the globally optimal objective value -0.06090 of $k=100$ is closer to the value -0.06068 obtained by our Julia implementation with a sequential approach, as was reported in Table 4.7. This is expected; the discretized formulation (4.4.6) inevitably introduces discretization error, which typically becomes smaller as the step size Δt shrinks. Without range reduction, BARON failed to solve the problem to global optimality. For $k=10$, BARON reported that the "problem is numerically sensitive"; we were unable to determine why. BARON's documentation suggests that in this case, the reported "best possible" value is likely a correct globally optimal objective value, but this was not true for this example. For $k=100$, the obtained upper and lower bounds for the globally optimal objective value failed to converge within 500 seconds.

Table 4.8: Computational results for solving (4.4.6) with BARON in Example 4.6.

k	Range reduction?	Successfully solved?	Glob. optim. obj. *	CPU time (seconds) *	# iterations in B&B *
10	Yes	Yes	−0.06292	0.21	5
100	Yes	Yes	−0.06090	0.75	25
10	No	No	"problem is numerically sensitive", "best possible" = −0.47320		
100	No	No	LB = −0.4695, UB = −0.06090 after 500 seconds		

* Each number here is the average of 10 runs. The abbreviation "glob. optim. obj." stands for "globally optimal objective value", and "B&B" stands for "branch-and-bound.

4.5 Conclusions and future work

Based on Scott and Barton's general ODE relaxation framework [2], we have proposed a new approach for generating convex and concave relaxations for the solutions of nonconvex parametric ODE systems (2.3.1), for use in deterministic global dynamic optimization. This approach furnishes new right-hand sides $(\mathbf{u}, \mathbf{o})$ in the Scott–Barton relaxation framework (2.4.1) as optimal-value functions defined by (4.2.3). Prior to this work, the SBM relaxations summarized in Section 2.4 were the only established way to actually generate valid relaxations within the framework (2.4.1). While the SBM relaxations require the generalized McCormick relaxations of $\mathbf{f}$, our new approach admits any relaxations of $\mathbf{f}$ that satisfy Assumption 4.2.1, such as affine relaxations, αBB relaxations [9], McCormick relaxations, envelopes, and even the pointwise tightest among multiple relaxations. Moreover, Theorem 4.3.11 shows that tighter relaxations for $\mathbf{x}_0$ and $\mathbf{f}$ in (2.3.1) translate into tighter state relaxations for $\mathbf{x}$; this in turn incentivizes seeking tighter relaxations for closed-form functions such as $\mathbf{f}$. Corollary 4.3.15 moreover shows that if McCormick relaxations are applied in our approach, then our new state relaxations are at least as tight as the SBM relaxations [2] that are based on generalized McCormick relaxations. These properties are beneficial in the context of global dynamic optimization, since tighter relaxations lead to fewer iterations required by branch-and-bound-based deterministic optimization algorithms. Theorem 4.3.17 shows that our new ODE relaxations also inherit second-order pointwise convergence from the relaxations for $\mathbf{f}$, thus mitigating the cluster effect in an overarching global dynamic optimization method. Proof-of-concept implementations of these relaxations were developed in MATLAB, and a proof-of-concept global optimization solver was formulated in Julia for the $n_x = 1$ case, by combining our new relaxation approach with EAGO, JuMP, and DifferentialEquations.

We remark that our proof-of-concept implementation in MATLAB may not yet be appropriate for general usage in deterministic global dynamic optimization. Firstly, this implementation solves convex optimization problems at each right-hand side evaluation, which invokes significant computational expense. Potential methods for improving computational efficiency, including warm-started optimization and KKT reformulation, were outlined in Section 4.3.7. Secondly, since nonsmooth relaxations $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ such as McCormick relaxations may be embedded, and since MATLAB's NLP solvers assume smoothness, these solvers may perform poorly when solving the NLPs embedded in (4.2.3). Thus, nonsmooth convex optimization solvers such as Nesterov's level method [40] would be preferred for handling nonsmooth $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$, but are not yet implemented here. Thirdly, in order to handle the right-hand sides' discontinuities in $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ introduced by the if-statements in (2.4.1), an event detection scheme in [2] would be preferred. In the current implementation, such discontinuities are represented by if-statements in MATLAB, which may introduce numerical difficulties during integration [125].

Future work will involve developing this implementation further to compute the proposed ODE relaxations more efficiently. Developing a method for computing subgradients for solutions of (2.4.1) would also be particularly useful, since subgradients are important when obtaining lower bounds for the globally optimal objective value of (1.1.1) using off-the-shelf convex optimization solvers.

Chapter 5

Extending Optimization-Based Convex Relaxations for Dynamic Process Models

This chapter, reproduced from the manuscript in preparation [93], describes two extensions of the optimization-based ODE relaxations proposed in Chapter 4. When simulating the optimization-based relaxation system, the MATLAB implementation in Chapter 4 numerically solves convex optimization problems for each right-hand side evaluation, which may require expensive computational efforts. In the first extension, it is shown that if the employed relaxations of the original right-hand side functions have pre-known monotonicity, then closed-form extrema can be identified directly, and thus the optimization-based ODE relaxation system's right-hand side can be efficiently evaluated in closed form. A numerical example then suggests that by using this method, the resulting optimization-based ODE relaxations may be significantly tighter, yet as efficient as the SBM relaxations. The second extension is a new ODE relaxation approach based on the optimization-based relaxations, which constructs different convex optimization problems at the relaxation system's right-hand side. These convex optimization problems are constructed from relaxing the original right-hand side using the effective Auxiliary Variable Method (AVM) [8, 50]. It is shown that the new AVM-based ODE relaxations are guaranteed to be at least as tight as the optimization-based relaxations and in some cases significantly tighter, as illustrated by several numerical examples based on a proof-of-concept implementation in Julia. The right-hand side of the AVM-based system may be evaluated in closed form in certain cases as well.

5.1 Introduction

This chapter considers the original ODE system (2.3.1) with a *factorable* right-hand side function $\mathbf{f}$. As in [6, Definition 8], a factorable function can be represented as a finite composition of predefined intrinsic functions. The Auxiliary Variable Method (AVM) [8, 50] and the McCormick relaxation method [5] are two methods for automatically constructing convex and concave relaxations for general factorable functions. The AVM was originally proposed for relaxing nonconvex optimization problems to yield convex programs, by first introducing auxiliary variables to capture any nonlinearities, and then bounding these variables by appropriate convex and concave relaxations. The AVM is employed in the state-of-the-art deterministic global optimization solver BARON [28], and has been shown to be empirically efficient for computing lower bounds in global optimization. As will be shown in Definition 5.4.6 below, the AVM may also be extended to construct convex relaxations for factorable functions as optimal-value functions (in the sense of e.g. [114]). The McCormick relaxation method [5] and its variants [6, 7, 33, 34] typically construct closed-form convex relaxations by recursively applying relaxation rules for intrinsic functions, without introducing auxiliary variables. It has been shown [35] that in certain cases, using McCormick relaxations in global optimization may lead to significant computational savings over the AVM. On the other hand, the AVM may in general yield tighter relaxations than the McCormick relaxations, by effectively handling repeated terms in a factorable function [7, 45].

For an original ODE system (2.3.1) with a factorable $\mathbf{f}$, this chapter proposes two extensions of the *optimization-based (OB) state relaxations* proposed in Chapter 4, for efficiently computing tight state relaxations for use in deterministic algorithms of global dynamic optimization. In the first extension, we show that if the employed relaxations of $\mathbf{f}$ in the OB relaxation formulation have known monotonicity on the considered box domain, then an optimal solution may be directly identified on the box. Thus, the right-hand side of the OB relaxation system can be efficiently evaluated in closed form; no need to use numerical NLP solvers. This method would be much more efficient than numerically solving convex NLPs at each right-hand side evaluation as in [3]. In Example 5.1 below, convex envelopes with known monotonicity of an original ODE right-hand side $\mathbf{f}$ are applied for constructing the OB relaxations. The results show that the OB relaxations are tighter and also as efficient as the SBM relaxations, which may ultimately improve computational efficiency in an overarching global optimization method.

In the second extension, we propose a new relax-then-discretize state relaxation approach in the Scott–Barton framework, which constructs a third new class of right-hand side functions after the SBM relaxations and OB relaxations. Similarly to the OB relaxation system's right-hand side, these new right-hand side functions are optimal-value functions with convex optimization problems embedded, but are constructed very differently. As mentioned above, the OB relaxation convex NLPs employ any relaxations of $\mathbf{f}$ as objective functions and employ box constraints. On the other hand, the new convex NLPs here are motivated by handling a factorable $\mathbf{f}$ using the AVM [8, 50], and employ linear objective functions and nonlinear convex constraints. In the new formulation, $\mathbf{f}$ is factorized, and new convex NLPs are constructed from bounding each factor with convex and concave relaxations of the corresponding intrinsic function, and also employing the box constraints in the OB relaxation formulation. Thus, this new formulation extends the efficient AVM from computing lower bounds in global optimization to effectively handling factorable ODE right-hand sides for constructing state relaxations. The new state relaxations (referred as *AVM-based relaxations*) have desirable tightness properties over the established relaxations [2, 3] in the Scott–Barton framework. As a necessity for establishing these results, we prove rigorously that for any given factorable function, the AVM relaxations are guaranteed to be at least as tight as the multivariate McCormick relaxations [7], and the multivariate McCormick relaxations are guaranteed to be at least as tight as the classical (univariate) McCormick relaxations [5]. These were briefly discussed in [7, Section 4], but not yet rigorously proved. By leveraging these results, it is shown that the new AVM-based state relaxations are at least as tight as both the SBM relaxations and the OB relaxations derived from the McCormick relaxations of $\mathbf{f}$ (denoted as the *optimization–based–McCormick (OBM) relaxations*). This is promising since McCormick relaxations are commonly used when tight relaxations of $\mathbf{f}$ are not directly available, and the new AVM-based state relaxations are superior to these state relaxations constructed from McCormick relaxations of $\mathbf{f}$. Furthermore, numerical examples in this chapter suggest that if in practice $\mathbf{f}$ has *repeated factors* as in Definition 5.4.26 or if convex envelopes of multivariate intrinsic functions are available, then the AVM-based relaxations may be significantly tighter than the SBM and OBM relaxations. We will also outline a proof-of-concept implementation of the new AVM-based relaxations in Julia [95], which numerically solves convex NLPs at each time step. This implementation may require expensive computational effort, similarly to the implementation for OB relaxations in [3]. We note that the techniques for efficiently solving ODE right-hand side convex NLPs in [3, Section 5.7.1] may also be useful here. Besides, if the factors of **f** are bounded by affine relaxations, then the right-hand side convex NLPs reduce to linear programming (LP) problems. Example 5.5 below shows that by solving LPs, the AVM-based relaxations can be evaluated much more efficiently than solving convex NLPs at the relaxation system's right-hand side, yet without compromising much tightness of the resulting relaxations. Lastly, the right-hand side of the AVM-based relaxation system may also be efficiently evaluated in closed form by leveraging the monotonicity of the employed relaxations for each factor, as will be seen in Example 5.5.

The remainder of this chapter is organized as follows. Section 5.2 summarizes the OB relaxation formulation proposed in Chapter 4. Section 5.3 proposes the first contribution of this chapter, which constructs closed-form OB relaxation formulation's right-hand side based on known monotonicity of relaxations of **f**. Section 5.4 then proposes the new AVM-based state relaxation approach. Established relaxation methods for factorable functions are firstly summarized, and the new state relaxation formulation is then presented, along with a comparison to the OB relaxation formulation. Next, theoretical properties of the new approach are established, including solutions' uniqueness, valid bounding and convexity properties, and desirable tightness properties. Lastly, a proof-of-concept implementation in Julia is outlined, and several numerical examples are presented to illustrate the tightness properties of the new AVM-based state relaxations.

5.2 Background: optimization-based relaxation formulation

This section summarizes the optimization-based state relaxation formulation proposed in Chapter 4. Consider the Scott–Barton framework (2.4.1) in Section 2.4. The optimization-based relaxation formulation furnishes Scott–Barton right-hand side functions $(\mathbf{u},\mathbf{o})$ as optimal-value functions (in the sense of e.g. [114]), constructed as follows. Define a function $\mathbf{v}:\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$ so that for all $i\in\{1,\dots,n_x\}$ and $\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}\in\mathbb{R}^{n_x}$,

$$v_i(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \tfrac{1}{2}[(\alpha_i+1)\xi_i^{\mathrm{cc}}-(\alpha_i-1)\xi_i^{\mathrm{cv}}]. \tag{5.2.1}$$

Intuitively, $v_i(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})$ is a linear combination of $\boldsymbol{\xi}^{\mathrm{cv}}$ and $\boldsymbol{\xi}^{\mathrm{cc}}$, weighted in a particular way based on the value of $\boldsymbol{\alpha}$. Consider functions $\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}}:I\times P\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$ that satisfy the following assumption.

Assumption 5.2.1 (from [3])**.** Suppose that functions $\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}}:I\times P\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$ satisfy the following conditions:

1. $\mathbf{f}^{\mathrm{cv}}$ and $\mathbf{f}^{\mathrm{cc}}$ are continuous,
2. $\mathbf{f}^{\mathrm{cv}}(t,\mathbf{p},\cdot)$ and $\mathbf{f}^{\mathrm{cc}}(t,\mathbf{p},\cdot)$ are Lipschitz continuous on $\mathbb{R}^{n_x}$, uniformly over $(t,\mathbf{p})\in I\times P$, and
3. for a.e. $t\in I$, the functions $\mathbf{f}^{\mathrm{cv}}(t,\cdot,\cdot)$ and $\mathbf{f}^{\mathrm{cc}}(t,\cdot,\cdot)$ are, respectively, convex and concave relaxations of $\mathbf{f}(t,\cdot,\cdot)$ in (2.3.1) on $P\times X(t)$.

Then, for each $i\in\{1,\dots,n_x\}$ and $(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})\in I\times P\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$, [3] constructs the following (u_i, o_i):

$$\begin{aligned} & u_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \min_{\boldsymbol{\alpha}\in[-1,1]^{n_x}} f_i^{\mathrm{cv}}(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})) \text{ subject to } \alpha_i = -1, \\ \text{and} \quad & o_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \max_{\boldsymbol{\alpha}\in[-1,1]^{n_x}} f_i^{\mathrm{cc}}(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})) \text{ subject to } \alpha_i = +1. \end{aligned} \tag{5.2.2}$$

Thus, this state relaxation approach is to solve (2.4.1) with $(\mathbf{u},\mathbf{o})$ defined in (5.2.2). Observe that on a set $S := \{(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \in I\times P\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x} : \boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}} \in X(t) \text{ and } \boldsymbol{\xi}^{\mathrm{cv}} \leq \boldsymbol{\xi}^{\mathrm{cc}}\}$, the (u_i, o_i) in (5.2.2) reduce to

$$\begin{aligned} & u_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \min_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cv}}, \\ \text{and} \quad & o_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \max_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cc}}. \end{aligned} \tag{5.2.3}$$

As discussed in [3, Remark 1], for each $t \in I$, $\mathbf{f}^{\mathrm{cv}}(t,\cdot,\cdot)$ and $\mathbf{f}^{\mathrm{cc}}(t,\cdot,\cdot)$ above may be constructed as any Lipschitz continuous convex and concave relaxations of $\mathbf{f}(t,\cdot,\cdot)$ on $P \times X(t)$. Lipschitz continuity on the full space may be enforced by passing to an appropriate Lipschitz extension [126]. Note that [3, Theorem 1] guarantees that any solution $(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}))$ of (2.4.1) with (5.2.2) embedded is always within the box $X(t)$. Defining $\mathbf{f}^{\mathrm{cv}}(t,\mathbf{p},\cdot)$ and $\mathbf{f}^{\mathrm{cc}}(t,\mathbf{p},\cdot)$ outside $X(t)$ is for the convenience of validating solutions' theoretical properties.

This state relaxation approach can utilize established convex and concave relaxations for $\mathbf{f}$ including McCormick relaxations [5,6], αBB relaxations [9,49], affine relaxations [4, 35, 44], and convex envelopes. Moreover, tighter such relaxations will necessarily translate into at least as tight state relaxations, as guaranteed in [3, Theorem 5]. Notably, [3, Corollary 2] shows that using McCormick relaxations [5] of $\mathbf{f}$ in this approach necessarily yields state relaxations that are at least as tight as the SBM relaxations [2]. In this chapter, we refer to the state relaxations of this approach with any $(\mathbf{f}^{\text{cv}}, \mathbf{f}^{\text{cc}})$ embedded as the *optimization-based (OB) relaxations*. Specifically, if $(\mathbf{f}^{\text{cv}}, \mathbf{f}^{\text{cc}})$ are McCormick relaxations, then the resulting state relaxations will be referred as the *optimization-based-McCormick (OBM) relaxations*.

5.3 Using closed-form minima of convex envelopes

When computing the OB relaxations [3], the proof-of-concept-implementation proposed in [3] numerically solves the optimization problems in (5.2.2) using NLP solvers (e.g. `fmincon` in MATLAB or IPOPT [120]). As shown in [3, Example 4], even though the OB relaxations can be tighter than the SBM relaxations [2], this naive implementation may require more expensive computational effort, and thus ultimately lead to longer computational time for an overarching global optimization method. As the first contribution of this chapter, Example 5.1 below shows that if the relaxations $(\mathbf{f}^{\text{cv}}, \mathbf{f}^{\text{cc}})$ (including envelopes) employed in (5.2.2) have pre-known monotonicity on the interval X, then $(\mathbf{u}, \mathbf{o})$ in (5.2.2) can be efficiently evaluated in closed forms, by directly identifying optimal solutions of the right-hand side optimization problems at corners of the box $[-1,1]^{n_x}$. With this method, the resulting OB relaxations may be equally efficiently evaluated, yet tighter than the SBM relaxations.

Example 5.1. Let $P := [0.1, 1.5]$, and $I := [0, 2.5]$, and consider the following instance of (2.3.1) on I:

$$\begin{aligned} \dot{x}_1(t) &= \frac{\sqrt{x_2}}{x_1^2}, \qquad \dot{x}_2(t) = x_2 e^{-x_1}, \\ x_1(0) &= 1 + p^2, \quad x_2(0) = p - \frac{p^3}{3} + 5. \end{aligned} \tag{5.3.1}$$

Harrison's state bounding method was applied to the ODEs above, and it was empirically verified that the resulting state bounds $\mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{U}}(t) > 0$ for each $t \in I$. Define the right-hand side functions of (5.3.1) as $g : \boldsymbol{\xi} \mapsto \sqrt{\xi_2}/\xi_1^2$ and $h : \boldsymbol{\xi} \mapsto \xi_2 e^{-\xi_1}$. Let $\Xi = \{(\boldsymbol{\xi}, \boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}}) \in \mathbb{R}^6 : 0 < \boldsymbol{\xi}^{\mathrm{L}} \leq \boldsymbol{\xi} \leq \boldsymbol{\xi}^{\mathrm{U}}\}$. Following [54, Corollaries 1 and 2], we constructed functions $g^{\mathrm{cv,env}}, h^{\mathrm{cv,env}} : \Xi \to \mathbb{R}$ so that for any $\boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}} \in \mathbb{R}^2$ for which $0 < \boldsymbol{\xi}^{\mathrm{L}} \leq \boldsymbol{\xi}^{\mathrm{U}}$, $g^{\mathrm{cv,env}}(\cdot, \boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}})$ and $h^{\mathrm{cv,env}}(\cdot, \boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}})$ are respectively convex envelopes of g and h on $[\boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}}]$, as follows. For each $(\boldsymbol{\xi}, \boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}}) \in \Xi$, let

$$\lambda_1 = \frac{\xi_2^{\mathrm{U}} - \xi_2}{\xi_2^{\mathrm{U}} - \xi_2^{\mathrm{L}}} \quad \text{and} \quad \lambda_2 = \frac{\xi_2 - \xi_2^{\mathrm{L}}}{\xi_2^{\mathrm{U}} - \xi_2^{\mathrm{L}}},$$

let $\alpha = (\xi_2^{\mathrm{L}}/\xi_2^{\mathrm{U}})^{1/6}$ and then [54, Corollary 1] indicates that

$$g^{\mathrm{cv,env}}(\boldsymbol{\xi}, \boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}}) := \begin{cases} \lambda_1 (\xi_1^{\mathrm{L}})^{-2}\sqrt{\xi_2^{\mathrm{L}}} + \lambda_2^3(\xi_1 - \lambda_1\xi_1^{\mathrm{L}})^{-2}\sqrt{\xi_2^{\mathrm{U}}}, & \text{if } \xi_1^{\mathrm{L}} \leq \xi_1 \leq \lambda_1\xi_1^{\mathrm{L}} + \lambda_2 \min\{\xi_1^{\mathrm{L}}/\alpha, \xi_1^{\mathrm{U}}\}, \\ \xi_1^{-2}(\lambda_1(\xi_2^{\mathrm{L}})^{1/6} + \lambda_2(\xi_2^{\mathrm{U}})^{1/6})^3, & \text{if } (\lambda_1 + \lambda_2/\alpha)\xi_1^{\mathrm{L}} \leq \xi_1 \leq (\lambda_1\alpha + \lambda_2)\xi_1^{\mathrm{U}}, \\ \lambda_1^3(\xi_1 - \lambda_2\xi_1^{\mathrm{U}})^{-2}\sqrt{\xi_2^{\mathrm{L}}} + \lambda_2(\xi_1^{\mathrm{U}})^{-2}\sqrt{\xi_2^{\mathrm{U}}}, & \text{if } \lambda_1 \max\{\xi_1^{\mathrm{L}}, \alpha\xi_1^{\mathrm{U}}\} + \lambda_2\xi_1^{\mathrm{U}} \leq \xi_1 \leq \xi_1^{\mathrm{U}}, \end{cases}$$

let $\alpha' = -\ln(\xi_2^{\mathrm{U}}/\xi_2^{\mathrm{L}})$ and then [54, Corollary 2] indicates that

$$h^{\mathrm{cv,env}}(\boldsymbol{\xi}, \boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}}) := \begin{cases} \lambda_1\xi_2^{\mathrm{L}}e^{-\xi_1^{\mathrm{L}}} + \lambda_2 e^{(-\xi_1 + \lambda_1\xi_1^{\mathrm{L}})/\lambda_2}\xi_2^{\mathrm{U}}, & \text{if } \xi_1^{\mathrm{L}} \leq \xi_1 \leq \lambda_1\xi_1^{\mathrm{L}} + \lambda_2\min\{\xi_1^{\mathrm{L}} - \alpha', \xi_1^{\mathrm{U}}\}, \\ e^{-\xi_1}(\xi_2^{\mathrm{L}})^{\lambda_1}(\xi_2^{\mathrm{U}})^{\lambda_2}, & \text{if } \xi_1^{\mathrm{L}} - \lambda_2\alpha' \leq \xi_1 \leq \xi_1^{\mathrm{U}} + \lambda_1\alpha', \\ \lambda_1 e^{(-\xi_1 + \lambda_2\xi_1^{\mathrm{U}})/\lambda_1}\xi_2^{\mathrm{L}} + \lambda_2 e^{-\xi_1^{\mathrm{U}}}\xi_2^{\mathrm{U}}, & \text{if } \lambda_1\max\{\xi_1^{\mathrm{L}}, \xi_1^{\mathrm{U}} + \alpha'\} + \lambda_2\xi_1^{\mathrm{U}} \leq \xi_1 \leq \xi_1^{\mathrm{U}}. \end{cases} \tag{5.3.2}$$

Next, appropriate functions $(f_1^{\mathrm{cv}}, f_2^{\mathrm{cv}})$ for use in (5.2.2) were constructed by setting

$$f_1^{\mathrm{cv}}(t, p, \boldsymbol{\xi}) := g^{\mathrm{cv,env}}(\boldsymbol{\xi}, \mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{U}}(t)) \quad \text{and} \quad f_2^{\mathrm{cv}}(t, p, \boldsymbol{\xi}) := h^{\mathrm{cv,env}}(\boldsymbol{\xi}, \mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{U}}(t)). \tag{5.3.3}$$

Instead of numerically solving the optimization problems for defining (u_1,u_2) in (5.2.2) with (5.3.3) embedded, we now describe how these functions can be evaluated in closed form in this case, and thus lead to a more efficient evaluation method. We observed that for several $t \in I$, with slight abuse of notation, for each fixed $\xi_1 \in [x_1^{\mathrm{L}}(t), x_1^{\mathrm{U}}(t)]$, the function $g^{\mathrm{cv,env}}(\xi_1,\cdot,\mathbf{x}^{\mathrm{L}}(t),\mathbf{x}^{\mathrm{U}}(t))$ is monotonic increasing on $[x_2^{\mathrm{L}}(t), x_2^{\mathrm{U}}(t)]$, and for each fixed $\xi_2 \in [x_2^{\mathrm{L}}(t), x_2^{\mathrm{U}}(t)]$, the function $h^{\mathrm{cv,env}}(\cdot,\xi_2,\mathbf{x}^{\mathrm{L}}(t),\mathbf{x}^{\mathrm{U}}(t))$ is monotonic decreasing on $[x_2^{\mathrm{L}}(t), x_2^{\mathrm{U}}(t)]$. However, due to the complexity of these functions, it is difficult to verify the mentioned monotonicity rigorously for each $t \in I$; we directly assume such monotonicity of $(g^{\mathrm{cv,env}}, h^{\mathrm{cv,env}})$ for each $t \in I$. Thus, the optimization problems for defining u_1 and u_2 in (5.2.2) always have optimal solutions $\boldsymbol{\alpha}^{*,u_1} := (-1,-1)$ and $\boldsymbol{\alpha}^{*,u_2} := (1,-1)$, respectively. Therefore, the functions (u_1,u_2) defined using (5.2.2) with (5.3.3) embedded have the following closed forms:

$$\begin{aligned} & u_1(t,p,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv f_1^{\mathrm{cv}}(t,p,\mathbf{v}(\boldsymbol{\alpha}^{*,u_1},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})) \\ \text{and} \quad & u_2(t,p,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv f_2^{\mathrm{cv}}(t,p,\mathbf{v}(\boldsymbol{\alpha}^{*,u_2},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})). \end{aligned} \tag{5.3.4}$$

Since there are no available concave envelopes for g and h that have the desirable monotonicity as $(g^{\mathrm{cv,env}}, h^{\mathrm{cv,env}})$ above, we constructed closed-form McCormick concave relaxations $h^{\mathrm{cc,MC}}, g^{\mathrm{cc,MC}} : \Xi \to \mathbb{R}$ as follows. It will be shown that these concave relaxations have the required monotonicity for identifying closed-form maxima. For each $(\boldsymbol{\xi},\boldsymbol{\xi}^{\mathrm{L}},\boldsymbol{\xi}^{\mathrm{U}}) \in \Xi$, let

$$\beta^g = \frac{-(\xi_1^{\mathrm{L}}+\xi_1^{\mathrm{U}})}{(\xi_1^{\mathrm{L}}\xi_1^{\mathrm{U}})^2}(\xi_1-\xi^{\mathrm{L}}) + \frac{1}{(\xi_1^{\mathrm{L}})^2} \quad \text{and} \quad \beta^h = \frac{e^{-\xi_1^{\mathrm{U}}} - e^{-\xi_1^{\mathrm{L}}}}{\xi_1^{\mathrm{U}} - \xi_1^{\mathrm{L}}}(\xi_1-\xi_1^{\mathrm{L}}) + e^{-\xi_1^{\mathrm{L}}},$$

and then,

$$\begin{aligned} g^{\mathrm{cc,MC}}(\boldsymbol{\xi},\boldsymbol{\xi}^{\mathrm{L}},\boldsymbol{\xi}^{\mathrm{U}}) &:= \min\{\frac{\sqrt{\xi_2}}{(\xi_1^{\mathrm{U}})^2}+\sqrt{\xi_2^{\mathrm{U}}}\beta^g-\frac{\sqrt{\xi_2^{\mathrm{U}}}}{(\xi_1^{\mathrm{U}})^2},\frac{\sqrt{\xi_2}}{(\xi_1^{\mathrm{L}})^2}+\sqrt{\xi_2^{\mathrm{L}}}\beta^g-\frac{\sqrt{\xi_2^{\mathrm{L}}}}{(\xi_1^{\mathrm{L}})^2}\} \\ \text{and} \quad h^{\mathrm{cc,MC}}(\boldsymbol{\xi},\boldsymbol{\xi}^{\mathrm{L}},\boldsymbol{\xi}^{\mathrm{U}}) &:= \min\{e^{-\xi_1^{\mathrm{U}}}\xi_2+\xi_2^{\mathrm{U}}\beta^h-\xi_2^{\mathrm{U}}e^{-\xi_1^{\mathrm{U}}},e^{-\xi_1^{\mathrm{L}}}\xi_2+\xi_2^{\mathrm{L}}\beta^h-\xi_2^{\mathrm{L}}e^{-\xi_1^{\mathrm{L}}}\}. \end{aligned} \tag{5.3.5}$$

Note that if $\xi_1^{\mathrm{U}} \equiv \xi_1^{\mathrm{L}}$, then β^h above is defined as $e^{-\xi_1^{\mathrm{L}}}$. Since $\boldsymbol{\xi}^{\mathrm{L}},\boldsymbol{\xi}^{\mathrm{U}} > 0$, on any $[\boldsymbol{\xi}^{\mathrm{L}},\boldsymbol{\xi}^{\mathrm{U}}]$, for each fixed ξ_1, $g^{\mathrm{cc,MC}}$ is monotonic increasing with respect to ξ_2, and for each fixed ξ_2, $h^{\mathrm{cc,MC}}$ is monotonic decreasing with respect to ξ_1. Thus, with the definition:

$$f_1^{\mathrm{cc}}(t,p,\boldsymbol{\xi}) := g^{\mathrm{cc,MC}}(\boldsymbol{\xi},\mathbf{x}^{\mathrm{L}}(t),\mathbf{x}^{\mathrm{U}}(t)) \quad \text{and} \quad f_2^{\mathrm{cc}}(t,p,\boldsymbol{\xi}) := h^{\mathrm{cc,MC}}(\boldsymbol{\xi},\mathbf{x}^{\mathrm{L}}(t),\mathbf{x}^{\mathrm{U}}(t)),$$

the optimization problems for defining o_1 and o_2 in (5.2.2) always have optimal solutions $\boldsymbol{\alpha}^{*,o_1} := (1,1)$ and $\boldsymbol{\alpha}^{*,o_2} := (-1,1)$, respectively. Therefore, we may construct closed-form (o_1,o_2) defined in (5.2.2) as:

$$\begin{aligned} o_1(t,p,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &\equiv f_1^{\mathrm{cc}}(t,p,\mathbf{v}(\boldsymbol{\alpha}^{*,o_1},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})) \\ \text{and} \quad o_2(t,p,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &\equiv f_2^{\mathrm{cc}}(t,p,\mathbf{v}(\boldsymbol{\alpha}^{*,o_2},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})). \end{aligned} \tag{5.3.6}$$

The initial-value functions in (5.3.1) were relaxed using the McCormick relaxation method, and then the OB relaxation system were constructed by embedding closed-form $(\mathbf{u},\mathbf{o})$ in (5.3.4) and (5.3.6) into (2.4.1). We computed both the SBM relaxations and the OB relaxations based on implementations in Julia v1.4.2 [95]. Harrison's state bounds [69] $(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}})$ were integrated simultaneously with the relaxation systems. The closed-form functions $(\mathbf{u},\mathbf{o})$ in (5.3.4) and (5.3.6) were hard-coded. Any necessary computations of natural interval extension [48] and generalized McCormick relaxations [6] were performed via operator overloading using EAGO v0.4.1 [94]. All ODE systems were solved using the ODE solver `BS3()` with an absolute tolerance of 10^{-4} and a relative tolerance of 10^{-4} from the package DifferentialEquations v6.15.0 [123]. These computations were performed on a Dell desktop computer with two 3.00GHz Intel Core i7-9700 CPUs and 16.0 GB of RAM.

Figure 5.1 depicts the resulting state relaxations, along with the original ODE solution for comparison. Observe that the OB relaxations are mostly visually tighter than the SBM relaxations [2]. Table 5.1 presents the CPU times for evaluating $(\mathbf{x}^{\mathrm{cv}}(2.5, p_i), \mathbf{x}^{\mathrm{cc}}(2.5, p_i))$ at one mesh point p_i by the considered two state relaxation methods. These per-mesh-point CPU times were obtained by taking the total evaluation time for all mesh points p_i simultaneously, dividing this by the number of mesh points, and averaging this figure over ten runs. Observe that both methods roughly took the same CPU time, since both relaxation systems' right-hand sides were evaluated in closed form. Thus, the OB relaxations obtained by this method are tighter, but at the same time as efficient as the SBM relaxations, which may ultimately improve efficiency for an overarching global optimization method.

Table 5.1: Average computational times for evaluating state relaxations $(\mathbf{x}^{\mathrm{cv}}(2.5, p), \mathbf{x}^{\mathrm{cc}}(2.5, p))$ for (5.3.1) in Example 5.1.

State relaxation method	CPU time (seconds)*
SBM relaxations	0.06
OB relaxations	0.05

* Each CPU time here was averaged over 10 runs, with a sample standard deviation that is much smaller than the reported average.

For this example, if concave envelopes with desirable monotonicity for g and h are available, then it is reasonable to expect that using these to construct closed-form (o_1, o_2) defined in (5.2.2) would yield further tighter OB relaxations than in Figure 5.1, yet as

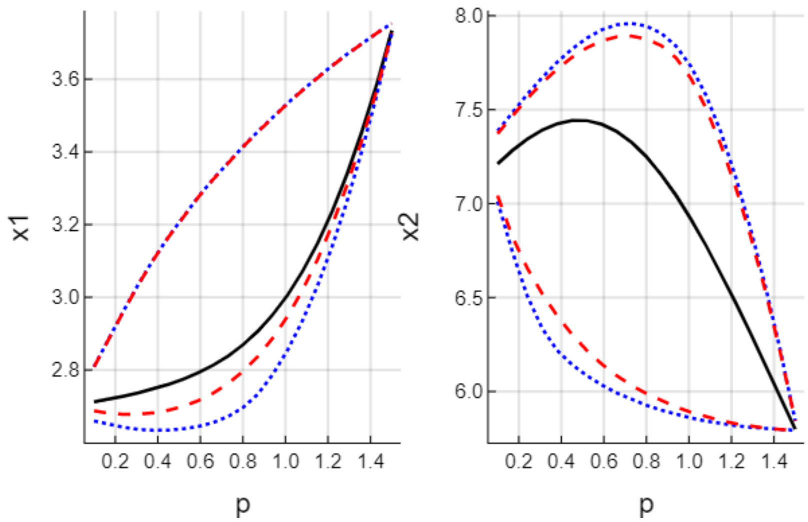

Figure 5.1: The solution $x_1(2.5, p)$ (*left, solid black*) and $x_2(2.5, p)$ (*right, solid black*) of the parametric ODEs (5.3.1) from Example 5.1, plotted against p along with corresponding SBM relaxations [2] (*dotted blue*) and the OB relaxations (*dashed red*) obtained by solving (2.4.1) with closed-form (5.3.4) and (5.3.6) embedded. (In the left subfigure, the concave SBM relaxation overlaps with the new concave relaxation for x_1.)

efficient as the SBM relaxations. This efficient evaluation method for OB relaxations is applicable to an original parametric ODE system with multi-parameters, more than two state variables, and explicit parameter-dependence at right-hand side, as long as $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ have pre-known monotonicity on X analogous to this example.

5.4 New state relaxation approach

As the second contribution of this chapter, we now present a new state relaxation approach for the original parametric ODE system (2.3.1). This approach is extended from the OB relaxation approach [3], is applicable to a factorable original right-hand side function $\mathbf{f}$, and yields state relaxations that are at least as tight as the OBM relaxations (OB relaxations derived from McCormick relaxations of $\mathbf{f}$). This section is organized as follows. Section 5.4.1

gives a definition of factorable functions considered in this chapter, and summarizes several existing relaxation approaches for factorable functions. Section 5.4.2 presents the new relaxation formulation, and Section 5.4.3 describes how this formulation compares to the OB relaxation formulation. Section 5.4.4 then validates theoretical properties of the new formulation, including solutions' uniqueness, validity as state relaxations, and tightness. Section 5.4.5 outlines a proof-of-concept implementation in Julia, and Section 5.4.6 presents several numerical examples to illustrate the desirable tightness properties of the new state relaxations.

5.4.1 Background: convex relaxations of factorable functions

In this subsection, we give a definition of factorable functions employed in this chapter, and introduce three established convex relaxation methods for factorable functions: the McCormick relaxation method [5, 6], a formalized Tsoukalas–Mitsos–McCormick relaxation method based on multivariate McCormick relaxations [6, 7], and the Auxiliary Variable Method [8, 50]. These methods are useful for developing and validating a new ODE relaxation approach later.

Definition 5.4.1. Given $V \subseteq \mathbb{R}^n$, a function $w : V \to \mathbb{R}$ is called a *multivariate intrinsic function* if,

1. functions $w^{\mathrm{L}}, w^{\mathrm{U}} : \mathbb{I}V \to \mathbb{R}$ are available so that for each $\tilde{V} \in \mathbb{I}V$ and each $\mathbf{v} \in \tilde{V}$, $w^{\mathrm{L}}(\tilde{V}) \leq w(\mathbf{v}) \leq w^{\mathrm{U}}(\tilde{V})$,

2. functions $w^{\mathrm{cv}}, w^{\mathrm{cc}} : V \times \mathbb{I}V \to \mathbb{R}$ are available so that for each $\tilde{V} \in \mathbb{I}V$, the mappings $w^{\mathrm{cv}}(\cdot, \tilde{V})$ and $w^{\mathrm{cc}}(\cdot, \tilde{V})$ are respectively convex and concave relaxations of w on $\tilde{V}$, and

3. the following holds:

$$w^{\mathrm{L}}(\tilde{V}) \le w^{\mathrm{cv}}(\mathbf{v},\tilde{V}) \le w^{\mathrm{cc}}(\mathbf{v},\tilde{V}) \le w^{\mathrm{U}}(\mathbf{v},\tilde{V}), \quad \forall \tilde{V} \in \mathbb{I}V, \, \forall \mathbf{v} \in \tilde{V}. \tag{5.4.1}$$

For any given $\tilde{V} \in \mathbb{I}V$, we are only interested in evaluating $w^{\mathrm{cv}}(\mathbf{v},\tilde{V})$ and $w^{\mathrm{cc}}(\mathbf{v},\tilde{V})$ for each $\mathbf{v} \in \tilde{V}$. Defining $(w^{\mathrm{cv}}, w^{\mathrm{cc}})$ on $V \times \mathbb{I}V$ is for convenience of notation. For any relaxations $(w^{\mathrm{cv,A}}, w^{\mathrm{cc,A}})$ that do not satisfy (5.4.1), new relaxations $(w^{\mathrm{cv,B}}, w^{\mathrm{cc,B}})$ that satisfy (5.4.1) can be constructed by setting for any $\tilde{V} \in \mathbb{I}V$ and any $\mathbf{v} \in \tilde{V}$,

$$\begin{aligned} & w^{\mathrm{cv,B}}(\mathbf{v},\tilde{V}) := \max(w^{\mathrm{L}}(\tilde{V}), w^{\mathrm{cv,A}}(\mathbf{v},\tilde{V})), \\ \text{and} \quad & w^{\mathrm{cc,B}}(\mathbf{v},\tilde{V}) := \min(w^{\mathrm{U}}(\tilde{V}), w^{\mathrm{cc,A}}(\mathbf{v},\tilde{V})). \end{aligned}$$

Now, a factorable function considered in this chapter is defined below.

Definition 5.4.2. Given $Y \subseteq \mathbb{R}^n$, a function $h: Y \to \mathbb{R}$ is *factorable* if it can be expressed in terms of a finite number of factors $v_1, ..., v_m$ such that for any given $\mathbf{y} \in Y$,

1. for each $j \in \{1,...,m\}$, let $V_j \subseteq \mathbb{R}^{j-1}$, there exists a multivariate intrinsic function $w_j : Y \times V_j \to \mathbb{R}$ such that $v_j := w_j(\mathbf{y}, \mathbf{v}_{1:j-1})$,
2. $h(\mathbf{y}) \equiv v_m$.

The following assumption is made for any factorable functions considered in this chapter, and is standard in interval analysis [48].

Assumption 5.4.3. Given $Y \subseteq \mathbb{R}^n$, consider a factorable function $h: Y \to \mathbb{R}$ with related quantities and functions in Definitions 5.4.2 and 5.4.1. Let $\hat{V}_1$ be a set of null vector, and for each $\tilde{Y} \in \mathbb{I}Y$, for each $j \in \{2,...,m\}$, recursively define

$$\hat{V}_j := [w_1^{\mathrm{L}}(\tilde{Y},\hat{V}_1), w_1^{\mathrm{U}}(\tilde{Y},\hat{V}_1)] \times ... \times [w_{j-1}^{\mathrm{L}}(\tilde{Y},\hat{V}_{j-1}), w_{j-1}^{\mathrm{U}}(\tilde{Y},\hat{V}_{j-1})],$$

and assume that $\hat{V}_j \subseteq V_j$.

Definition 5.4.2 above generalizes the conventional definition of factorable functions in [5, 6, 35]. [6, Definition 8] considers a factorable function as a finite composition of binary addition, binary multiplication, and known univariate intrinsic functions. Definition 5.4.2 considers a factorable function as a finite composition of known multivariate intrinsic functions. These multivariate intrinsic functions may include, but are not limited to, these operations in [6, Definition 8]. In fact, nearly every function that can be represented finitely on a computer is factorable in the sense of [6, Definition 8]. However, Definition 5.4.2 has the advantage that it allows using known tight convex relaxations of multivariate intrinsic functions for constructing convex relaxations for a factorable function, as illustrated in the following example.

Example 5.2 (adapted from [53])**.** Consider a function $\phi : [0,1] \times [0.1,2] \times [0,1] \to \mathbb{R}$:

$$\phi : (y_1, y_2, y_3) \mapsto (\sqrt{y_1} - y_2)\exp(-y_3) + y_1 y_2.$$

The conventional definition [6, Definition 8] of a factorable function may factorize ϕ as follows: for each $\mathbf{y}$,

$$\begin{aligned} &v_1 = \sqrt{y_1}, \quad v_2 = -y_2, \qquad v_3 = v_1 + v_2, \\ &v_4 = -y_3, \quad v_5 = \exp(v_4), \qquad v_6 = v_3 v_5, \\ &v_7 = y_1 y_2, \quad v_8 = v_6 + v_7, \quad \phi(\mathbf{y}) = v_8. \end{aligned}$$

Observe that the factor representation above involves the following nonlinear univariate intrinsic functions:

$$w_1 : y_1 \mapsto \sqrt{y_1} \quad \text{and} \quad w_5 : v_4 \mapsto \exp(v_4).$$

With known convex and concave relaxations of the nonlinear (w_1, w_5) above, the McCormick relaxation method [5, 6] computes convex relaxations for ϕ by recursively applying relaxation rules for univariate composition, binary addition, and binary multiplication. On the other hand, Definition 5.4.2 in this chapter may factorize ϕ using another different factor representation: for each $\mathbf{y}$,

$$v_1 = (\sqrt{y_1} - y_2)\exp(-y_3), \quad v_2 = y_1 y_2, \quad v_3 = v_1 + v_2, \quad \phi(\mathbf{y}) = v_3. \tag{5.4.2}$$

Observe that this factor representation involves the following nonlinear multivariate intrinsic functions:

$$w_1 : (y_1, y_2, y_3) \mapsto (\sqrt{y_1} - y_2)\exp(-y_3) \quad \text{and} \quad w_2 : (y_1, y_2) \mapsto y_1 y_2. \tag{5.4.3}$$

Convex and concave envelopes of w_1 are available in [53, Example 3], and w_2 may be relaxed using the well-known McCormick envelope [5]. However, the standard McCormick relaxation method as in [6, Definition 9] cannot utilize the tight relaxations of such multivariate (w_1, w_2) for computing ϕ's relaxations. On the other hand, in concert with Definition 5.4.2, the following *Tsoukalas–Mitsos–McCormick relaxation method* [6, 7] can relax ϕ with the factor representation (5.4.2). This method can utilize any convex and concave relaxations, even convex envelopes, for known multivariate intrinsic functions such as (w_1, w_2) in (5.4.3), and recursively compose these relaxations via the *multivariate McCormick relaxation rule* [7, Theorem 2], as defined below.

Definition 5.4.4 (adapted from [6] and [7])**.** Given $Y \subseteq \mathbb{R}^n$, consider a factorable function $h : Y \to \mathbb{R}$ and the related quantities and functions in Definitions 5.4.2 and 5.4.1. Denote the *Tsoukalas–Mitsos–McCormick (TMC) relaxations* of h as functions $h^{\mathrm{cv,TMC}}, h^{\mathrm{cc,TMC}} : Y \times \mathbb{I}Y \to \mathbb{R}$, where for each $\tilde{Y} \in \mathbb{I}Y$ and $\mathbf{y} \in \tilde{Y}$, $h^{\mathrm{cv,TMC}}(\mathbf{y}, \tilde{Y})$ and $h^{\mathrm{cc,TMC}}(\mathbf{y}, \tilde{Y})$ are defined by the following procedure:

1. Initialize $\mathbf{v}^{\mathrm{cv}}$ and $\mathbf{v}^{\mathrm{cc}}$ as null vectors.

2. Set $j := 1$.

3. Compute the interval $\hat{V}_j$ defined in Assumption 5.4.3.

4. Compute v_j^{cv} as

$$v_j^{\mathrm{cv}} := \min\{w_j^{\mathrm{cv}}(\mathbf{y}, \mathbf{v}_{1:j-1}, \tilde{Y}, \hat{V}_j) : \mathbf{v}_{1:j-1}^{\mathrm{cv}} \leq \mathbf{v}_{1:j-1} \leq \mathbf{v}_{1:j-1}^{\mathrm{cc}}\} \tag{5.4.4}$$

   and compute v_j^{cc} as

$$v_j^{\mathrm{cc}} := \max\{w_j^{\mathrm{cc}}(\mathbf{y}, \mathbf{v}_{1:j-1}, \tilde{Y}, \hat{V}_j) : \mathbf{v}_{1:j-1}^{\mathrm{cv}} \leq \mathbf{v}_{1:j-1} \leq \mathbf{v}_{1:j-1}^{\mathrm{cc}}\}. \tag{5.4.5}$$

5. If $j = m$, go to 6. Otherwise, assign $j := j+1$ and go to 3.

6. Set $h^{\mathrm{cv,TMC}}(\mathbf{y}) := v_m^{\mathrm{cv}}$ and $h^{\mathrm{cc,TMC}}(\mathbf{y}) := v_m^{\mathrm{cc}}$.

Then, $h^{\mathrm{cv,TMC}}(\cdot, \tilde{Y})$ and $h^{\mathrm{cc,TMC}}(\cdot, \tilde{Y})$ are guaranteed to be respectively convex and concave relaxations for h on $\tilde{Y}$.

The formulas (5.4.4) and (5.4.5) are derived from the multivariate McCormick relaxations [7, Theorem 2]. By applying this to h represented by a finite composition of factors,

$h^{\mathrm{cv,TMC}}$ and $h^{\mathrm{cc,TMC}}$ are indeed valid convex and concave relaxations of h. The TMC relaxation method was conceptually discussed in [7], and there is thus far no developed implementation for automatically executing this procedure.

For convenience, we reframe the standard McCormick relaxation method as in [6, Definition 9] so that it is applicable to a factorable function h as in Definition 5.4.2, with factorable multivariate intrinsic functions w_j in the sense of [6, Definition 8]. Note that such h is also overall factorable in the sense of [6, Definition 8]. For such h, the following definition of McCormick relaxations is equivalent to [6, Definition 9].

Definition 5.4.5 (adapted from [6])**.** Given $Y \subseteq \mathbb{R}^n$, consider a factorable function $h: Y \to \mathbb{R}$ and the related quantities and functions in Definitions 5.4.2 and 5.4.1. Suppose that each multivariate intrinsic function w_j is factorable in the sense of [6, Definition 8]. Denote the *McCormick (MC) relaxations* of h as functions $h^{\mathrm{cv,MC}}, h^{\mathrm{cc,MC}}: Y \times \mathbb{I}Y \to \mathbb{R}$, where for each $\tilde{Y} \in \mathbb{I}Y$ and $\mathbf{y} \in \tilde{Y}$, $h^{\mathrm{cv,MC}}(\mathbf{y}, \tilde{Y})$ and $h^{\mathrm{cc,MC}}(\mathbf{y}, \tilde{Y})$ are defined by the following procedure:

1. Initialize $\bar{\mathbf{v}}^{\mathrm{cv}}$ and $\bar{\mathbf{v}}^{\mathrm{cc}}$ as null vectors.

2. Set $j := 1$.

3. Compute the interval $\hat{V}_j$ defined in Assumption 5.4.3.

4. Denote the *generalized McCormick (gMC) relaxations* as in [6, Definition 15] of w_j as $w_j^{\mathrm{cv,gMC}}, w_j^{\mathrm{cc,gMC}}: \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{j-1} \times \mathbb{R}^{j-1} \times \mathbb{I}Y \times \mathbb{I}V_j \to \mathbb{R}$. Compute $\bar{v}_j^{\mathrm{cv}}$ and $\bar{v}_j^{\mathrm{cc}}$ as
$$\begin{aligned} \bar{v}_j^{\mathrm{cv}} &:= w_j^{\mathrm{cv,gMC}}(\mathbf{y}, \mathbf{y}, \bar{\mathbf{v}}_{1:j-1}^{\mathrm{cv}}, \bar{\mathbf{v}}_{1:j-1}^{\mathrm{cc}}, \tilde{Y}, \hat{V}_j), \\ \bar{v}_j^{\mathrm{cc}} &:= w_j^{\mathrm{cc,gMC}}(\mathbf{y}, \mathbf{y}, \bar{\mathbf{v}}_{1:j-1}^{\mathrm{cv}}, \bar{\mathbf{v}}_{1:j-1}^{\mathrm{cc}}, \tilde{Y}, \hat{V}_j). \end{aligned} \tag{5.4.6}$$

5. If $j = m$, go to 6. Otherwise, assign $j := j + 1$ and go to 3.

6. Set $h^{\mathrm{cv,MC}}(\mathbf{y}) := \bar{v}_m^{\mathrm{cv}}$ and $h^{\mathrm{cc,MC}}(\mathbf{y}) := \bar{v}_m^{\mathrm{cc}}$.

Then, $h^{\mathrm{cv,MC}}(\cdot, \tilde{Y})$ and $h^{\mathrm{cc,MC}}(\cdot, \tilde{Y})$ are guaranteed to be respectively convex and concave relaxations for h on $\tilde{Y}$.

It will be shown in Theorem 5.4.18 that the TMC relaxations of a factorable function in Definition 5.4.4 are at least as tight as the MC relaxations in Definition 5.4.5.

The Auxiliary Variable Method [8, 50] was originally proposed for relaxing nonconvex optimization problems to yield convex programs, by first introducing auxiliary variables to capture any nonconvexities, and then bounding these variables by appropriate convex and concave relaxations. This method is used in the deterministic global optimization solver BARON [60] to compute the required lower bounds of the globally optimal objective values. In this chapter, as a straightforward extension, we define the Auxiliary Variable Method for computing convex relaxations of a factorable function in the sense of Definition 5.4.2.

Definition 5.4.6 (adapted from [50]). Given $Y \subseteq \mathbb{R}^n$, consider a factorable function $h : Y \to \mathbb{R}$ with related quantities and functions in Definitions 5.4.2 and 5.4.1. Consider the sets

$\hat{V}_1, ..., \hat{V}_m$ defined in Assumption 5.4.3. The *Auxiliary Variable Method* constructs optimal-value functions $h^{\mathrm{cv,AVM}}, h^{\mathrm{cc,AVM}} : Y \times \mathbb{I}Y \to \mathbb{R}$ so that for each $\tilde{Y} \in \mathbb{I}Y$ and $\mathbf{y} \in \tilde{Y}$,

$$\begin{aligned}
h^{\mathrm{cv,AVM}}(\mathbf{y}, \tilde{Y}) := \min_{\mathbf{v}} \quad & v_m \\
\text{s.t.} \quad & \forall j \in \{1, ..., m\}, \\
& w_j^{\mathrm{cv}}(\mathbf{y}, \mathbf{v}_{1:j-1}, \tilde{Y}, \hat{V}_j) \leq v_j \leq w_j^{\mathrm{cc}}(\mathbf{y}, \mathbf{v}_{1:j-1}, \tilde{Y}, \hat{V}_j), \\
h^{\mathrm{cc,AVM}}(\mathbf{y}, \tilde{Y}) := \max_{\mathbf{v}} \quad & v_m \\
\text{s.t.} \quad & \forall j \in \{1, ..., m\}, \\
& w_j^{\mathrm{cv}}(\mathbf{y}, \mathbf{v}_{1:j-1}, \tilde{Y}, \hat{V}_j) \leq v_j \leq w_j^{\mathrm{cc}}(\mathbf{y}, \mathbf{v}_{1:j-1}, \tilde{Y}, \hat{V}_j).
\end{aligned} \tag{5.4.7}$$

Then, the mappings $h^{\mathrm{cv,AVM}}(\cdot, \tilde{Y})$ and $h^{\mathrm{cc,AVM}}(\cdot, \tilde{Y})$ are relatively convex and concave relaxations of h on $\tilde{Y}$. These mappings will be referred as the *AVM relaxations* for h on $\tilde{Y}$.

Observe that the optimization problems in (5.4.7) are convex. It will be shown in Theorem 5.4.20 that for any given factorable function and any bounds and relaxations of the factors, the AVM relaxations defined above are at least as tight as the TMC relaxations in Definition 5.4.4.

5.4.2 New state relaxation formulation

Consider the original parametric ODE system (2.3.1) in Section 2.3, and further assume that the right-hand side function $\mathbf{f}$ is factorable as in Definition 5.4.2. This new state relaxation formulation is based on the Scott–Barton ODE relaxation framework (2.4.1), and employs a novel construction of the right-hand side functions $(\mathbf{u}, \mathbf{o})$ as follows.

For each $i \in \{1, ..., n_x\}$, let $m(i)$ denote the number of factors $v_1, ..., v_{m(i)}$ for f_i, and let

$J_i := \{1,...,m(i)\}$. Thus, based on Definitions 5.4.1 and 5.4.2 and Assumption 5.4.3, for each $i \in \{1,...,n_x\}$ and $j \in J_i$,

1. let $V_{j,i} \subseteq \mathbb{R}^{j-1}$, and define multivariate intrinsic functions $w_{j,i} : I \times P \times D \times V_{j,i} \to \mathbb{R}$ so that $v_j := w_{j,i}(t, \mathbf{p}, \boldsymbol{\xi}, \mathbf{v}_{1:j-1})$ for each $(t, \mathbf{p}, \boldsymbol{\xi}, \mathbf{v}_{1:j-1}) \in I \times P \times D \times V_{j,i}$,

2. define functions $w^{\mathrm{L}}_{j,i}, w^{\mathrm{U}}_{j,i} : I \times \mathbb{I}P \times \mathbb{I}D \times \mathbb{I}V_{j,i} \to \mathbb{R}$ so that for each $(t, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}) \in I \times \mathbb{I}P \times \mathbb{I}D \times \mathbb{I}V_{j,i}$,

$$w^{\mathrm{L}}_{j,i}(t, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}) \le w_{j,i}(t, \mathbf{p}, \boldsymbol{\xi}, \mathbf{v}_{1:j-1}) \le w^{\mathrm{U}}_{j,i}(t, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}),\ \forall (\mathbf{p}, \boldsymbol{\xi}, \mathbf{v}_{1:j-1}) \in \tilde{P} \times \tilde{D} \times \tilde{V}_{j,i},$$

3. define functions $w^{\mathrm{cv}}_{j,i}, w^{\mathrm{cc}}_{j,i} : I \times P \times D \times V_{j,i} \times \mathbb{I}P \times \mathbb{I}D \times \mathbb{I}V_{j,i} \to \mathbb{R}$ so that for each $(t, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}) \in I \times \mathbb{I}P \times \mathbb{I}D \times \mathbb{I}V_{j,i}$, the mappings $w^{\mathrm{cv}}_{j,i}(t, \cdot, \cdot, \cdot, \tilde{P}, \tilde{D}, \tilde{V}_{j,i})$ and $w^{\mathrm{cc}}_{j,i}(t, \cdot, \cdot, \cdot, \tilde{P}, \tilde{D}, \tilde{V}_{j,i})$ are respectively convex and concave relaxations of $w_{j,i}(t, \cdot, \cdot, \cdot)$ on $\tilde{P} \times \tilde{D} \times \tilde{V}_{j,i}$, and for each $(\mathbf{p}, \boldsymbol{\xi}, \mathbf{v}_{1:j-1}) \in \tilde{P} \times \tilde{D} \times \tilde{V}_{j,i}$,

$$\begin{aligned} w^{\mathrm{L}}_{j,i}(t, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}) &\le w^{\mathrm{cv}}_{j,i}(t, \mathbf{p}, \boldsymbol{\xi}, \mathbf{v}_{1:j-1}, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}) \\ &\le w^{\mathrm{cc}}_{j,i}(t, \mathbf{p}, \boldsymbol{\xi}, \mathbf{v}_{1:j-1}, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}) \le w^{\mathrm{U}}_{j,i}(t, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}), \end{aligned} \tag{5.4.8}$$

4. for each $t \in I$, let $\hat{V}_{1,i}(t)$ be a set of null vector, and for each $j \in \{2,...,m(i)\}$, recursively define $\hat{V}_{j,i}(t)$:

$$\begin{aligned} \hat{V}_{j,i}(t) := {} & [w^{\mathrm{L}}_{1,i}(t, P, X(t), \hat{V}_{1,i}(t)), w^{\mathrm{U}}_{1,i}(t, P, X(t), \hat{V}_{1,i}(t))] \times ... \\ & \times [w^{\mathrm{L}}_{j-1,i}(t, P, X(t), \hat{V}_{j-1,i}(t)), w^{\mathrm{U}}_{j-1,i}(t, P, X(t), \hat{V}_{j-1,i}(t))], \end{aligned} \tag{5.4.9}$$

   and assume that $\hat{V}_{j,i}(t) \in \mathbb{I}V_{j,i}$.

Let $S := \{(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \in I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} : \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}} \in X(t) \text{ and } \boldsymbol{\xi}^{\mathrm{cv}} \le \boldsymbol{\xi}^{\mathrm{cc}}\}$, and define functions $\bar{\mathbf{u}}, \bar{\mathbf{o}} : S \to \mathbb{R}^{n_x}$ so that for each $i \in \{1,...,n_x\}$ and $(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \in S$,

$$\begin{aligned}
\bar{u}_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \min_{\boldsymbol{\xi},\mathbf{v}} \quad & v_{m(i)} \\
\text{s.t.} \quad & \xi_i = \xi_i^{\mathrm{cv}}, \\
& \boldsymbol{\xi}^{\mathrm{cv}} \le \boldsymbol{\xi} \le \boldsymbol{\xi}^{\mathrm{cc}}, \\
& \forall j \in J_i, \\
& -v_j \le -w_{j,i}^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi},\mathbf{v}_{1:j-1},P,X(t),\hat{V}_{j,i}(t)), && \text{(5.4.10a)} \\
& \quad v_j \le w_{j,i}^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi},\mathbf{v}_{1:j-1},P,X(t),\hat{V}_{j,i}(t)), && \text{(5.4.10b)} \\
\bar{o}_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \max_{\boldsymbol{\xi},\mathbf{v}} \quad & v_{m(i)} \\
\text{s.t.} \quad & \xi_i = \xi_i^{\mathrm{cc}}, \\
& \boldsymbol{\xi}^{cv} \le \boldsymbol{\xi} \le \boldsymbol{\xi}^{\mathrm{cc}}, \\
& \forall j \in J_i, \\
& -v_j \le -w_{j,i}^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi},\mathbf{v}_{1:j-1},P,X(t),\hat{V}_{j,i}(t)), \\
& \quad v_j \le w_{j,i}^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi},\mathbf{v}_{1:j-1},P,X(t),\hat{V}_{j,i}(t)).
\end{aligned}$$

Observe that the right-hand sides' optimization problems above are convex.

Remark 5.4.7. The functions $(w_j^{\mathrm{cv}}, w_j^{\mathrm{cc}})$ can be any convex and concave relaxations for w_j, including McCormick relaxations [5, 6], αBB relaxations [9, 49], affine relaxations [4, 35, 44], and convex envelopes.

Remark 5.4.8. For any $i \in \{1,...,n_x\}$, $j \in J_i$, and $t \in I$, if the multivariate intrinsic function $w_{j,i}(t,\cdot,\cdot,\cdot)$ is affine on $P \times X(t) \times \hat{V}_{j,i}(t)$, then we can choose the relaxations $w_{j,i}^{\mathrm{cv}}(t,\cdot,\cdot,\cdot,P,X(t),\hat{V}_{j,i}(t))$ and $w_{j,i}^{\mathrm{cc}}(t,\cdot,\cdot,\cdot,P,X(t),\hat{V}_{j,i}(t))$ to be identical to $w_{j,i}(t,\cdot,\cdot,\cdot)$. Thus, for these (i,j,t), the constraints (5.4.10a) and (5.4.10b) are reduced to a linear equality constraint $v_j = w_{j,i}(t,\mathbf{p},\boldsymbol{\xi},\mathbf{v}_{1:j-1})$.

Remark 5.4.9. If all $(w^{\mathrm{cv}}_{j,i}, w^{\mathrm{cc}}_{j,i})$ are chosen to be affine relaxations, then the convex optimization problems in (5.4.10) reduce to linear optimization problems.

We make the following blanket assumptions concerning the uniform Lipschitz continuity of the constructed $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ in (5.4.10).

Assumption 5.4.10. There exists a scalar $l > 0$ so that for each $t \in I$, $\mathbf{p} \in P$, and $\boldsymbol{\xi}^{\mathrm{cv,A}}, \boldsymbol{\xi}^{\mathrm{cc,A}}, \boldsymbol{\xi}^{\mathrm{cv,B}}, \boldsymbol{\xi}^{\mathrm{cc,B}} \in X(t)$ for which $\boldsymbol{\xi}^{\mathrm{cv,A}} \leq \boldsymbol{\xi}^{\mathrm{cc,A}}$ and $\boldsymbol{\xi}^{\mathrm{cv,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,B}}$,

$$\begin{aligned}\|\bar{\mathbf{u}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}) - \bar{\mathbf{u}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}})\|_\infty + \|\bar{\mathbf{o}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}) - \bar{\mathbf{o}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}})\|_\infty \\ \leq l\Big(\|\boldsymbol{\xi}^{\mathrm{cv,A}} - \boldsymbol{\xi}^{\mathrm{cv,B}}\|_\infty + \|\boldsymbol{\xi}^{\mathrm{cc,A}} - \boldsymbol{\xi}^{\mathrm{cc,B}}\|_\infty\Big).\end{aligned}$$

It will be shown in Section 5.4.4 that under stronger assumptions on the functions for defining $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ in (5.4.10), $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ are guaranteed to satisfy the assumption above. Now, consider functions $\mathbf{u}, \mathbf{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ that satisfy the following assumption.

Assumption 5.4.11. Consider functions $\mathbf{u}, \mathbf{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ defined in (5.4.10), for which suppose that

1. for each $(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \in S$,

$$\mathbf{u}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \bar{\mathbf{u}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \text{ and } \mathbf{o}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \bar{\mathbf{o}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}),$$

2. $\mathbf{u}(t, \mathbf{p}, \cdot, \cdot)$ and $\mathbf{o}(t, \mathbf{p}, \cdot, \cdot)$ are Lipschitz continuous on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ uniformly over $(t, \mathbf{p}) \in I \times P$.

Since Assumption 5.4.10 holds for $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$, [126, Theorem 5] implies that the $(\mathbf{u}, \mathbf{o})$ in the assumption above exist and may be constructed as appropriate Lipschitz extensions of $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ from S to $I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$.

Then, our new approach for constructing state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ for (2.3.1) is to solve the auxiliary ODE system (2.4.1) with $(\mathbf{u}, \mathbf{o})$ satisfying Assumption 5.4.11. Unfortunately, we encounter difficulties for showing certain continuity and measurability of these $(\mathbf{u}, \mathbf{o})$. Such properties can help validate solutions' existence of the new ODE relaxation system (c.f. [102]). We instead make the following assumption about solutions' existence.

Assumption 5.4.12. Assume that there exists at least one solution for the ODE system (2.4.1) with $(\mathbf{u}, \mathbf{o})$ satisfying Assumption 5.4.11.

It will be shown in Section 5.4.4 that the new ODE relaxation system has right-uniqueness, whose unique solution is guaranteed to be valid state relaxations for (2.3.1).

Remark 5.4.13. It will be shown in Section 5.4.4 that any solution $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ of our new ODE relaxation system satisfies that for all $(t, \mathbf{p}) \in I \times P$, $\mathbf{x}^{\mathrm{L}}(t) \leq \mathbf{x}^{\mathrm{cv}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cc}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{U}}(t)$. Thus, the solutions do not visit the region outside of S. Therefore, we may practically construct functions $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ in (5.4.10), and then solve (2.4.1) with $(\mathbf{u}, \mathbf{o}) \leftarrow (\bar{\mathbf{u}}, \bar{\mathbf{o}})$. This would yield identical state relaxations. The Lipschitz extensions $(\mathbf{u}, \mathbf{o})$ have the advantage that they are defined on an open set $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ of state variables, which is convenient for validating theoretical properties of the new relaxation approach. As discussed in [3, 72], using $(\mathbf{u}, \mathbf{o})$ for numerically solving the ODEs may be preferable over using $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$, since ODE solvers typically require the right-hand side functions defined on an open set. However, from the authors' numerical experiments, using $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ did not cause numerical issues for ODE solvers with explicit integration methods.

5.4.3 Comparison to optimization-based relaxations

The new ODE relaxation system (2.4.1) with $(\mathbf{u}, \mathbf{o})$ satisfying Assumption 5.4.11 is motivated by embedding the AVM relaxations as in Definition 5.4.6 of the original right-hand side f_i into the optimization-based ODE relaxation formulation (5.2.3) and (2.4.1). For illustrating this, consider the quantities and functions in (5.4.10), and apply the Auxiliary Variable Method as in Definition 5.4.6 to the factorable function $f_i(t,\cdot,\cdot)$ for each $i \in \{1,...,n_x\}$ and $t \in I$. Thus, define the AVM relaxations $\mathbf{f}^{\mathrm{cv,AVM}},\mathbf{f}^{\mathrm{cc,AVM}} : I \times P \times D \times \mathbb{I}P \times \mathbb{I}D \to \mathbb{R}^{n_x}$ so that for each $i \in \{1,...,n_x\}$, each $\mathbf{p} \in P$, and each $\boldsymbol{\xi} \in X(t)$,

$$
\begin{aligned}
f_i^{\mathrm{cv,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X(t)) := \min_{\mathbf{v}} \quad & v_{m(i)} \\
\text{s.t.} \quad & \forall j \in J_i, \\
& -v_j \leq -w_{j,i}^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi},\mathbf{v}_{1:j-1},P,X(t),\hat{V}_{j,i}(t)), \\
& v_j \leq w_{j,i}^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi},\mathbf{v}_{1:j-1},P,X(t),\hat{V}_{j,i}(t)), \\
f_i^{\mathrm{cc,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X(t)) := \max_{\mathbf{v}} \quad & v_{m(i)} \\
\text{s.t.} \quad & \forall j \in J_i, \\
& -v_j \leq -w_{j,i}^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi},\mathbf{v}_{1:j-1},P,X(t),\hat{V}_{j,i}(t)), \\
& v_j \leq w_{j,i}^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi},\mathbf{v}_{1:j-1},P,X(t),\hat{V}_{j,i}(t)).
\end{aligned}
\tag{5.4.11}
$$

According to Definition 5.4.6, for each $i \in \{1,...,n_x\}$ and each $t \in I$, $f_i^{\mathrm{cv,AVM}}(t,\cdot,\cdot,P,X(t))$ and $f_i^{\mathrm{cc,AVM}}(t,\cdot,\cdot,P,X(t))$ are respectively convex and concave relaxations of $f_i(t,\cdot,\cdot)$ on $P \times X(t)$. Moreover, observe that for each $(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \in S$, $\bar{u}_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})$ and $\bar{o}_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})$ in (5.4.10) reduce to

$$
\begin{aligned}
\bar{u}_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &\equiv \min_{\boldsymbol{\xi} \in [\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cv,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X(t)) \text{ subject to } \xi_i = \xi_i^{\mathrm{cv}}, \\
\bar{o}_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &\equiv \max_{\boldsymbol{\xi} \in [\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cc,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X(t)) \text{ subject to } \xi_i = \xi_i^{\mathrm{cc}}
\end{aligned}
\tag{5.4.12}
$$

which resemble (5.2.3) with

$$f_i^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi}) := f_i^{\mathrm{cv,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X(t)) \text{ and } f_i^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi}) := f_i^{\mathrm{cc,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X(t)).$$

Thus, throughout the remainder of this chapter, the new ODE relaxation method that solves (2.4.1) with $(\mathbf{u},\mathbf{o})$ satisfying Assumption 5.4.11 will be referred as the *AVM-based ODE relaxation method*.

The new AVM-based relaxation method is similar to the OB relaxation method [3] in the sense that both solve convex optimization problems at relaxation systems' right-hand side. However, they are different in the following respects:

1. Though the new $(\bar{u}_i,\bar{o}_i)$ resemble (5.4.12), these $(\bar{u}_i,\bar{o}_i)$ can be evaluated by solving the convex optimization problems in (5.4.10), instead of solving the nested optimization problems in (5.4.12) with (5.4.11) embedded. The reformulation (5.4.12) is useful for validating theoretical properties of the new state relaxation approach.

2. The OB relaxation approach does not assume a factorable original right-hand side f_i in general, while the AVM-based approach assumes a factorable f_i. Nevertheless, almost all functions that can be presented in a scientific calculator are factorable [2, 6].

3. If the relaxations $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$ used in (5.2.2) are nonsmooth (e.g. McCormick relaxations [5]), then the right-hand side NLPs in (5.2.2) are nonsmooth and in principle require dedicated nonsmooth convex optimization solvers such as *Nesterov's Level Method* [40] or general nonsmooth solvers such as *bundle methods* [42, 43].

On the other hand, if the relaxations $(w^{\mathrm{cv}}_{j,i}, w^{\mathrm{cc}}_{j,i})$ for use in (5.4.10) are piecewise-differentiable in the sense of Scholtes [127], then such relaxations can be easily decomposed into a set of smooth constraints in (5.4.10). For example, the McCormick envelope [5] for bilinear terms can be represented by two lower bounding constraints and two upper bounding constraints [8]. Therefore, the resulting right-hand side NLPs in (5.4.10) are smooth and can be solved by off-the-shelf smooth NLP solvers such as IPOPT [120].

4. It will be shown in Theorem 5.4.23 that the new AVM-based relaxations are guaranteed to be at least as tight as the OBM relaxations (using McCormick relaxations of $\mathbf{f}$). Note that when the relaxations $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ are not directly available, McCormick relaxation method is a primary relaxation method for such factorable $\mathbf{f}$. Moreover, numerical examples in Section 5.4.6 will show that the new relaxation approach can effectively handle repeated factors as in Definition 5.4.26 of $\mathbf{f}$ and use convex envelopes $(w^{\mathrm{cv}}_{j,i}, w^{\mathrm{cc}}_{j,i})$ in (5.4.10), to yield significantly tighter relaxations than the SBM and OBM relaxations.

Lastly, both relaxation methods may require expensive computational effort, if the NLPs at relaxation systems' right-hand sides are naively solved using numerical NLP solvers. Section 5.3 has proposed to construct closed-form OB relaxation system's right-hand side, and several methods for improving computational efficiency of the OB relaxations are summarized in [3, Section 5.7.1]. These methods are also useful for efficiently evaluating the AVM-based ODE relaxations. For example, the right-hand side NLPs in (5.4.10) can be reformulated using Karush-Kuhn-Tucker complementarity conditions. Thus, the AVM-based ODE relaxation system is reformulated as an equivalent complementarity system (NCS), which could be solved by efficient NCS solvers such as

SICONOS [117]. In addition, the optimal solutions in (5.4.10) at a time t may be nearly optimal in the near future, and thus warm-starting an NLP solver may be particularly useful. As will be seen in Example 5.5, the AVM-based relaxation system's right-hand side may also be expressed in closed form, by directly identifying an optimal solution of (5.4.10). If all employed $(w_{j,i}^{\mathrm{cv}}, w_{j,i}^{\mathrm{cc}})$ are affine relaxations, i.e. (5.4.10) solves linear optimization problems, then solving these LPs is in principle much more efficient than solving convex NLPs during ODE solving. Moreover, in this setting, the feasible-basis tracking approach for solving ODEs with LP embedded described in [72, Section 5.3] may also be useful. Overall, we expect that the AVM-based relaxation method would be in general at most as efficient as the OB relaxation method, since the former explores detailed factor structure of $\mathbf{f}$ and employs more constraints in the right-hand side convex NLPs, to tighten ODE relaxations.

5.4.4 Properties of new state relaxations

This subsection establishes the following useful properties of the new AVM-based ODE relaxation system:

- Under mild additional assumptions on $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}, w_{j,i}^{\mathrm{L}}, w_{j,i}^{\mathrm{U}}, w_{j,i}^{\mathrm{cv}}, w_{j,i}^{\mathrm{cc}})$, $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ in (5.4.10) are guaranteed to exhibit the Lipschitz properties in Assumption 5.4.10, which in turn guarantees that the functions $(\mathbf{u}, \mathbf{o})$ used in (2.4.1) satisfy Assumption 5.4.11.
- The AVM-based ODE relaxation system has right-uniqueness.
- The unique solution of the AVM-based ODE relaxation system is guaranteed to be valid state relaxations for (2.3.1).

- For a given factorable function in Definition 5.4.2, the TMC relaxations in Definition 5.4.4 are at least as tight as the MC relaxations in Definition 5.4.5, and the AVM relaxations in Definition 5.4.6 are at least as tight as the TMC relaxations. For a given original parametric ODE system (2.3.1), the new AVM-based state relaxations are at least as tight as both the SBM and OBM relaxations [3] in the Scott–Barton framework.

Lipschitz continuity

Due to theoretical difficulties, we have to generally assume the Lipschitz continuity of $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ in (5.4.10), as in Assumption 5.4.10. Such property is crucial for validating theoretical properties of solutions of the AVM-based ODE relaxation system later. In this subsection, we show that if some mild additional assumptions are applied to the functions for defining $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$, $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ are guaranteed to satisfy Assumption 5.4.10.

Assumption 5.4.14. Consider the functions and quantities for defining $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ in (5.4.10). Further assume that the following conditions hold:

1. The state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ for (2.3.1) are Lipschitz continuous.
2. For each $i \in \{1, ..., n_x\}$ and each $j \in J_i$, the functions $(w^{\mathrm{cv}}_{j,i}, w^{\mathrm{cc}}_{j,i}, w^{\mathrm{L}}_{j,i}, w^{\mathrm{U}}_{j,i})$ are Lipschitz continuous, and for each $(t, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}) \in I \times \mathbb{I}P \times \mathbb{I}D \times \mathbb{I}V_{j,i}$ and each $(\mathbf{p}, \boldsymbol{\xi}, \mathbf{v}_{1:j-1}) \in \tilde{P} \times \tilde{D} \times \tilde{V}_{j,i}$,

$$w^{\mathrm{cv}}_{j,i}(t, \mathbf{p}, \boldsymbol{\xi}, \mathbf{v}_{1:j-1}, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}) < w^{\mathrm{cc}}_{j,i}(t, \mathbf{p}, \boldsymbol{\xi}, \mathbf{v}_{1:j-1}, \tilde{P}, \tilde{D}, \tilde{V}_{j,i}). \tag{5.4.13}$$

The assumption above can often be satisfied in practice. For example, Harrison's state bounds [69] $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ are Lipschitz continuous. The relaxations $(w^{\mathrm{cv}}_{j,i}, w^{\mathrm{cc}}_{j,i})$ constructed by

McCormick relaxations [5, 6], αBB relaxations [9, 49], and certain convex envelopes (e.g. envelopes developed in [52–54]) are Lipschitz continuous. The bounds $(w^{\mathrm{L}}_{j,i}, w^{\mathrm{U}}_{j,i})$ constructed by natural interval extension [48] are Lipschitz continuous. For any $(w^{\mathrm{cv,A}}_{j,i}, w^{\mathrm{cc,A}}_{j,i})$ that only satisfy a weak inequality as in (5.4.8), new $(w^{\mathrm{cv,B}}_{j,i}, w^{\mathrm{cc,B}}_{j,i})$ that satisfy (5.4.13) can be constructed by adding a small perturbation $\varepsilon > 0$ to $(w^{\mathrm{cv,A}}_{j,i}, w^{\mathrm{cc,A}}_{j,i})$, as shown below:

$$w^{\mathrm{cv,B}}_{j,i} := w^{\mathrm{cv,A}}_{j,i} - \varepsilon \quad \text{and} \quad w^{\mathrm{cc,B}}_{j,i} := w^{\mathrm{cc,A}}_{j,i} + \varepsilon.$$

Proposition 5.4.15. Under Assumption 5.4.14, the functions $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ defined in (5.4.10) satisfy Assumption 5.4.10.

Proof. Consider any fixed $i \in \{1, ..., n_x\}$. Under Assumption 5.4.14, for each $j \in J_i$, the mappings $w^{\mathrm{cv}}_{j,i}(\cdot,\cdot,\cdot,\cdot,P,X(\cdot),\hat{V}_{j,i}(\cdot))$ and $w^{\mathrm{cc}}_{j,i}(\cdot,\cdot,\cdot,\cdot,P,X(\cdot),\hat{V}_{j,i}(\cdot))$ are Lipschitz continuous. Moreover, since (5.4.13) holds, for each $(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \in S$, the domains of the optimization problems in (5.4.10) are compact convex sets with interior points, and thus satisfy the Slater's Conditions (c.f. [128, Theorem 1 (ii)]). Thus, [128, Theorem 1] implies that $(\bar{u}_i, \bar{o}_i)$ is locally Lipschitz continuous at each $(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \in S$. Note that [128, Theorem 1] requires the mappings $w^{\mathrm{cv}}_{j,i}(t, \mathbf{p}, \cdot, \cdot, P, X(t), \hat{V}_{j,i}(t))$ and $w^{\mathrm{cc}}_{j,i}(t, \mathbf{p}, \cdot, \cdot, P, X(t), \hat{V}_{j,i}(t))$ to be respectively convex and concave on $\mathbb{R}^{n_x} \times \mathbb{R}^{j-1}$for each $t \in I$; this can be easily satisfied by applying convex extensions (c.f. [129, Proposition 3.1.4]) to these mappings from $X(t) \times \hat{V}_{j,i}(t)$ to $\mathbb{R}^{n_x} \times \mathbb{R}^{j-1}$. Furthermore, since S is compact, it follows that $(\bar{u}_i, \bar{o}_i)$ are Lipschitz continuous on S. Then, it is readily verified that this guarantees $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ to satisfy Assumption 5.4.10. $\square$

Right-uniqueness

The following theorem establishes the right-uniqueness of solutions of the new AVM-based ODE relaxation system.

Theorem 5.4.16. Suppose that Assumption 5.4.10 holds. For each $\mathbf{p} \in P$, the ODE system (2.4.1) with $(\mathbf{u}, \mathbf{o})$ satisfying Assumption 5.4.11 has right-uniqueness on I.

Proof. It is readily verified that under Assumption 5.4.11, for each $\mathbf{p} \in P$, the overall right-hand side functions of (2.4.1) satisfy the sufficient condition for right-uniqueness established in [102, §10, Theorem 1]. □

Valid state relaxations

The following theorem shows that our new AVM-based ODE relaxation system yields valid state relaxations for the original ODE system (2.3.1).

Theorem 5.4.17. Suppose that Assumption 5.4.10 holds. Let $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ be a solution of (2.4.1) with $(\mathbf{u}, \mathbf{o})$ satisfying Assumption 5.4.11. Then, for each $(t, \mathbf{p}) \in I \times P$,

$$\mathbf{x}^{\mathrm{L}}(t) \leq \mathbf{x}^{\mathrm{cv}}(t, \mathbf{p}) \quad \text{and} \quad \mathbf{x}^{\mathrm{cc}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{U}}(t), \tag{5.4.14}$$

and $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ are valid state relaxations for (2.3.1) as in Definition 2.4.1.

Proof. [2, Lemma 1] implies that any solution $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ (in the Carathéodory sense) of the Scott–Barton ODE relaxation framework (2.4.1) satisfies (5.4.14). As discussed in Section 5.4.2, $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ in (5.4.10) are equivalent to (5.4.12), where $f_i^{\mathrm{cv,AVM}}/f_i^{\mathrm{cc,AVM}}(t, \cdot, \cdot, P, X(t))$ are valid convex and concave relaxations for $f_i(t, \cdot, \cdot)$ on $P \times X(t)$ for each $t \in I$. Then, since Assumption 5.4.11 holds, according to [3, Lemma 1 and Lemma 2], $(\mathbf{u}, \mathbf{o})$ describe both bound preserving dynamics and convexity preserving dynamics (c.f. [2, Definitions 6 and 7]) for the original ODE system (2.3.1). Finally, [2, Corollary 1 and Theorem 3] imply that any solution $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ of the new AVM-based ODE relaxation system are valid state relaxations for (2.3.1). □

The theorem above implies that the functions $(\mathbf{u}, \mathbf{o})$ in Assumption 5.4.11 satisfy Conditions 2 and 3 of the Scott–Barton right-hand side functions as in Definition 2.4.9. However, these $(\mathbf{u}, \mathbf{o})$ do not necessarily satisfy Condition 1 in Definition 2.4.9. Nevertheless, this is not a problem, since Condition 1 is only concerned with Scott–Barton relaxation framework solutions' existence, which has been addressed in Assumption 5.4.12 for the new AVM-based relaxation system.

Tightness

Tightness of relaxations for factorable functions Now, we establish tightness results for comparing the various relaxations of factorable functions introduced in Section 5.4.1. These results are essential for comparing our new ODE relaxations in this chapter to the established ODE relaxations [2,3]. The following theorem shows that for a given factorable function h in the sense of Definition 5.4.2, if McCormick relaxations for each factor w_j of h are used in Definition 5.4.4, then the TMC relaxations in Definition 5.4.4 are at least as tight as the MC relaxations in Definition 5.4.5.

Theorem 5.4.18. Given $Y \subseteq \mathbb{R}^n$, consider a factorable function $h : Y \to \mathbb{R}$ with related quantities and functions in Definitions 5.4.2 and 5.4.1. Suppose that each w_j is factorable in the sense of [6, Definition 8]. Consider the MC relaxations $(h^{\mathrm{cv,MC}}, h^{\mathrm{cc,MC}})$ for h and the gMC relaxations $(w_j^{\mathrm{cv,gMC}}, w_j^{\mathrm{cc,gMC}})$ for each w_j in Definition 5.4.5. For each $j \in \{1, ..., m\}$, define $w_j^{\mathrm{cv,MC}}, w_j^{\mathrm{cc,MC}} : Y \times V_j \times \mathbb{I}Y \times \mathbb{I}V_j \to \mathbb{R}$ by setting for each $\tilde{Y} \in \mathbb{I}Y$, $\tilde{V}_j \in \mathbb{I}V_j$, $\mathbf{y} \in \tilde{Y}$, and $\mathbf{v}_{1:j-1} \in \tilde{V}_j$,

$$\begin{aligned} w_j^{\mathrm{cv,MC}}(\mathbf{y}, \mathbf{v}_{1:j-1}, \tilde{Y}, \tilde{V}_j) &:= w_j^{\mathrm{cv,gMC}}(\mathbf{y}, \mathbf{y}, \mathbf{v}_{1:j-1}, \mathbf{v}_{1:j-1}, \tilde{Y}, \tilde{V}_j), \\ w_j^{\mathrm{cc,MC}}(\mathbf{y}, \mathbf{v}_{1:j-1}, \tilde{Y}, \tilde{V}_j) &:= w_j^{\mathrm{cc,gMC}}(\mathbf{y}, \mathbf{y}, \mathbf{v}_{1:j-1}, \mathbf{v}_{1:j-1}, \tilde{Y}, \tilde{V}_j). \end{aligned} \tag{5.4.15}$$

Then, $(w_j^{\mathrm{cv,MC}}, w_j^{\mathrm{cc,MC}})$ are valid McCormick relaxations for w_j. Now, consider the TMC relaxations $(h^{\mathrm{cv,TMC}}, h^{\mathrm{cc,TMC}})$ for h in Definition 5.4.4 with $(w_j^{\mathrm{cv}}, w_j^{\mathrm{cc}}) \leftarrow (w_j^{\mathrm{cv,MC}}, w_j^{\mathrm{cc,MC}})$. Then, for each $\tilde{Y} \in \mathbb{I}Y$ and each $\mathbf{y} \in \tilde{Y}$,

$$h^{\mathrm{cv,MC}}(\mathbf{y}, \tilde{Y}) \leq h^{\mathrm{cv,TMC}}(\mathbf{y}, \tilde{Y}) \leq h^{\mathrm{cc,TMC}}(\mathbf{y}, \tilde{Y}) \leq h^{\mathrm{cc,MC}}(\mathbf{y}, \tilde{Y}). \tag{5.4.16}$$

Proof. According to [6, 79], $(w_j^{\mathrm{cv,MC}}, w_j^{\mathrm{cc,MC}})$ defined in (5.4.15) are standard McCormick relaxations of w_j.

Consider any fixed $\tilde{Y} \in \mathbb{I}Y$ and $\mathbf{y} \in \tilde{Y}$. Consider the set $\hat{V}_j$ for each $j \in \{1, ..., m\}$ defined in Assumption 5.4.3. Consider the $(\mathbf{v}^{\mathrm{cv}}, \mathbf{v}^{\mathrm{cc}})$ computed using (5.4.4) and (5.4.5) in Definition 5.4.4, and consider the $(\bar{\mathbf{v}}^{\mathrm{cv}}, \bar{\mathbf{v}}^{\mathrm{cc}})$ computed using (5.4.6) in Definition 5.4.5. We prove (5.4.16) by showing that $\bar{v}_m^{\mathrm{cv}} \leq v_m^{\mathrm{cv}} \leq v_m^{\mathrm{cc}} \leq \bar{v}_m^{\mathrm{cc}}$. This will be proved using strong induction.

Firstly, we have

$$\begin{aligned} \bar{v}_2^{\mathrm{cv}} &:= w_2^{\mathrm{cv,gMC}}(\mathbf{y}, \mathbf{y}, \bar{v}_1^{\mathrm{cv}}, \bar{v}_1^{\mathrm{cc}}, \tilde{Y}, \hat{V}_2), \\ \bar{v}_2^{\mathrm{cc}} &:= w_2^{\mathrm{cc,gMC}}(\mathbf{y}, \mathbf{y}, \bar{v}_1^{\mathrm{cv}}, \bar{v}_1^{\mathrm{cc}}, \tilde{Y}, \hat{V}_2), \\ v_2^{\mathrm{cv}} &:= \min\{w_2^{\mathrm{cv,gMC}}(\mathbf{y}, \mathbf{y}, v_1, v_1, \tilde{Y}, \hat{V}_2) : v_1^{\mathrm{cv}} \leq v_1 \leq v_1^{\mathrm{cc}}\}, \\ \text{and} \quad v_2^{cc} &:= \max\{w_2^{\mathrm{cc,gMC}}(\mathbf{y}, \mathbf{y}, v_1, v_1, \tilde{Y}, \hat{V}_2) : v_1^{\mathrm{cv}} \leq v_1 \leq v_1^{\mathrm{cc}}\}, \end{aligned}$$

where

$$
\begin{aligned}
v_1^{\mathrm{cv}} \equiv \bar{v}_1^{\mathrm{cv}} &:= w_1^{\mathrm{cv,gMC}}(\mathbf{y},\mathbf{y},\bar{\mathbf{v}}_{1:0}^{\mathrm{cv}},\bar{\mathbf{v}}_{1:0}^{\mathrm{cc}},\tilde{Y},\hat{V}_1)\\
\text{and}\quad v_1^{\mathrm{cc}} \equiv \bar{v}_1^{\mathrm{cc}} &:= w_1^{\mathrm{cc,gMC}}(\mathbf{y},\mathbf{y},\bar{\mathbf{v}}_{1:0}^{\mathrm{cv}},\bar{\mathbf{v}}_{1:0}^{\mathrm{cc}},\tilde{Y},\hat{V}_1).
\end{aligned}
$$

Since generalized McCormick relaxations are *inclusion monotonic* as shown in [39, Theorem 2.4.32], for any $v_1 \in [\bar{v}_1^{\mathrm{cv}}, \bar{v}_1^{\mathrm{cc}}]$,

$$
\begin{aligned}
w_2^{\mathrm{cv,gMC}}(\mathbf{y},\mathbf{y},\bar{v}_1^{\mathrm{cv}},\bar{v}_1^{\mathrm{cc}},\tilde{Y},\hat{V}_2) &\leq w_2^{\mathrm{cv,gMC}}(\mathbf{y},\mathbf{y},v_1,v_1,\tilde{Y},\hat{V}_2)\\
&\leq w_2^{\mathrm{cc,gMC}}(\mathbf{y},\mathbf{y},v_1,v_1,\tilde{Y},\hat{V}_2) \leq w_2^{\mathrm{cc,gMC}}(\mathbf{y},\mathbf{y},\bar{v}_1^{\mathrm{cv}},\bar{v}_1^{\mathrm{cc}},\tilde{Y},\hat{V}_2),
\end{aligned}
$$

and thus $\bar{v}_2^{\mathrm{cv}} \leq v_2^{\mathrm{cv}} \leq v_2^{\mathrm{cc}} \leq \bar{v}_2^{\mathrm{cc}}$. Now, consider any $k \in \{3,\ldots,m\}$ for which suppose that $\bar{v}_r^{\mathrm{cv}} \leq v_r^{\mathrm{cv}} \leq v_r^{\mathrm{cc}} \leq \bar{v}_r^{\mathrm{cc}}$, for all $r \in \{1,\ldots,k-1\}$. We have

$$
\begin{aligned}
\bar{v}_k^{\mathrm{cv}} &:= w_k^{\mathrm{cv,gMC}}(\mathbf{y},\mathbf{y},\bar{\mathbf{v}}_{1:k-1}^{\mathrm{cv}},\bar{\mathbf{v}}_{1:k-1}^{\mathrm{cc}},\tilde{Y},\hat{V}_k),\\
\bar{v}_k^{\mathrm{cc}} &:= w_k^{\mathrm{cc,gMC}}(\mathbf{y},\mathbf{y},\bar{\mathbf{v}}_{1:k-1}^{\mathrm{cv}},\bar{\mathbf{v}}_{1:k-1}^{\mathrm{cc}},\tilde{Y},\hat{V}_k),\\
v_k^{\mathrm{cv}} &:= \min\{w_k^{\mathrm{cv,gMC}}(\mathbf{y},\mathbf{y},\mathbf{v}_{1:k-1},\mathbf{v}_{1:k-1},\tilde{Y},\hat{V}_k) : \mathbf{v}_{1:k-1}^{\mathrm{cv}} \leq \mathbf{v}_{1:k-1} \leq \mathbf{v}_{1:k-1}^{\mathrm{cc}}\},\\
\text{and}\quad v_k^{\mathrm{cc}} &:= \max\{w_k^{\mathrm{cc,gMC}}(\mathbf{y},\mathbf{y},\mathbf{v}_{1:k-1},\mathbf{v}_{1:k-1},\tilde{Y},\hat{V}_k) : \mathbf{v}_{1:k-1}^{\mathrm{cv}} \leq \mathbf{v}_{1:k-1} \leq \mathbf{v}_{1:k-1}^{\mathrm{cc}}\}.
\end{aligned}
$$

By a similar argument of inclusion monotonicity, it can be shown that $\bar{v}_k^{\mathrm{cv}} \leq v_k^{\mathrm{cv}} \leq v_k^{\mathrm{cc}} \leq \bar{v}_k^{\mathrm{cc}}$. Finally, (5.4.16) is proved by letting k increase inductively from 3 to m. □

Remark 5.4.19. The proof above implies that if non-McCormick relaxations $(w_j^{\mathrm{cv}}, w_j^{\mathrm{cc}})$ in Definition 5.4.4 are at least as tight as $(w_j^{\mathrm{cv,MC}}, w_j^{\mathrm{cc,MC}})$, then the TMC relaxations are also at least as tight as the MC relaxations for h.

The following theorem shows that for a given factorable function and any given bounds and relaxations of the factors, the AVM relaxations in Definition 5.4.6 are guaranteed to be at least as tight as the TMC relaxations in Definition 5.4.4.

Theorem 5.4.20. Given $Y \subseteq \mathbb{R}^n$, consider a factorable function $h : Y \to \mathbb{R}$ in the sense of Definition 5.4.2. Consider the TMC relaxations $(h^{\mathrm{cv,TMC}}, h^{\mathrm{cc,TMC}})$ for h in Definition 5.4.4 and the AVM relaxations $(h^{\mathrm{cv,AVM}}, h^{\mathrm{cc,AVM}})$ for h in Definition 5.4.6. Assume that same functions $(w_j^{\mathrm{L}}, w_j^{\mathrm{U}}, w_j^{\mathrm{cv}}, w_j^{\mathrm{cc}})$ in Definition 5.4.1 are applied to construct both $(h^{\mathrm{cv,TMC}}, h^{\mathrm{cc,TMC}})$ and $(h^{\mathrm{cv,AVM}}, h^{\mathrm{cc,AVM}})$. Then, for each $\tilde{Y} \in \mathbb{I}Y$ and each $\mathbf{y} \in \tilde{Y}$,

$$h^{\mathrm{cv,TMC}}(\mathbf{y}, \tilde{Y}) \leq h^{\mathrm{cv,AVM}}(\mathbf{y}, \tilde{Y}) \leq h^{\mathrm{cc,AVM}}(\mathbf{y}, \tilde{Y}) \leq h^{\mathrm{cc,TMC}}(\mathbf{y}, \tilde{Y}). \tag{5.4.17}$$

Proof. Consider any fixed $\tilde{Y} \in \mathbb{I}Y$ and $\mathbf{y} \in \tilde{Y}$. Consider the set $\hat{V}_j$ for each $j \in \{1, ..., m\}$ in Assumption 5.4.3. Consider the $(\mathbf{v}^{\mathrm{cv}}, \mathbf{v}^{\mathrm{cc}})$ computed in Definition 5.4.4 and the decision variables $\mathbf{v}$ in (5.4.7) in Definition 5.4.6. By observing the optimization problems in (5.4.7), to prove the claimed result, we need to show that for any $\mathbf{v}_{1:m-1} \in \hat{V}_m$ for which

$$w_r^{\mathrm{cv}}(\mathbf{y}, \mathbf{v}_{1:r-1}, \tilde{Y}, \hat{V}_r) \leq v_r \leq w_r^{\mathrm{cc}}(\mathbf{y}, \mathbf{v}_{1:r-1}, \tilde{Y}, \hat{V}_r), \quad \forall r \in \{1, ..., m-1\},$$

the following holds:

$$v_m^{\mathrm{cv}} \leq w_m^{\mathrm{cv}}(\mathbf{y}, \mathbf{v}_{1:m-1}, \tilde{Y}, \hat{V}_m) \leq w_m^{\mathrm{cc}}(\mathbf{y}, \mathbf{v}_{1:m-1}, \tilde{Y}, \hat{V}_m) \leq v_m^{\mathrm{cc}}.$$

We prove this using strong induction. Firstly, according to the formulations (5.4.4) and (5.4.5), we have

$$\begin{aligned} v_2^{\mathrm{cv}} &:= \min \{w_2^{\mathrm{cv}}(\mathbf{y}, v_1, \tilde{Y}, \hat{V}_2) : v_1^{\mathrm{cv}} \leq v_1 \leq v_1^{\mathrm{cc}}\} \\ \text{and} \quad v_2^{\mathrm{cc}} &:= \max \{w_2^{\mathrm{cc}}(\mathbf{y}, v_1, \tilde{Y}, \hat{V}_2) : v_1^{\mathrm{cv}} \leq v_1 \leq v_1^{\mathrm{cc}}\}, \end{aligned} \tag{5.4.18}$$

where $v_1^{\mathrm{cv}} := w_1^{\mathrm{cv}}(\mathbf{y}, \mathbf{v}_{1:0}, \tilde{Y}, \hat{V}_1)$ and $v_1^{\mathrm{cc}} := w_1^{\mathrm{cc}}(\mathbf{y}, \mathbf{v}_{1:0}, \tilde{Y}, \hat{V}_1)$. Observe that for any $v_1 \in \mathbb{R}$

such that $w_1^{\mathrm{cv}}(\mathbf{y},\mathbf{v}_{1:0},\tilde{Y},\hat{V}_1) \leq v_1 \leq w_1^{\mathrm{cc}}(\mathbf{y},\mathbf{v}_{1:0},\tilde{Y},\hat{V}_1)$,

$$v_2^{\mathrm{cv}} \leq w_2^{\mathrm{cv}}(\mathbf{y},v_1,\tilde{Y},\hat{V}_2) \leq w_2^{\mathrm{cc}}(\mathbf{y},v_1,\tilde{Y},\hat{V}_2) \leq v_2^{\mathrm{cc}}.$$

Now we show that for any $k := \{3,...,m\}$, if for any $\mathbf{v}_{1:k-1} \in \hat{V}_k$ for which

$$v_r^{\mathrm{cv}} \leq w_r^{\mathrm{cv}}(\mathbf{y},\mathbf{v}_{1:r-1},\tilde{Y},\hat{V}_r) \leq v_r \leq w_r^{\mathrm{cc}}(\mathbf{y},\mathbf{v}_{1:r-1},\tilde{Y},\hat{V}_r) \leq v_r^{\mathrm{cc}}, \quad \forall r \in \{1,...,k-1\}, \tag{5.4.19}$$

then the following holds:

$$v_k^{\mathrm{cv}} \leq w_k^{\mathrm{cv}}(\mathbf{y},\mathbf{v}_{1:k-1},\tilde{Y},\hat{V}_k) \leq w_k^{\mathrm{cc}}(\mathbf{y},\mathbf{v}_{1:k-1},\tilde{Y},\hat{V}_k) \leq v_k^{\mathrm{cc}}.$$

Recall that

$$\begin{aligned} & v_k^{\mathrm{cv}} := \min\,\{w_k^{\mathrm{cv}}(\mathbf{y},\mathbf{v}_{1:k-1},\tilde{Y},\hat{V}_k) : \mathbf{v}_{1:k-1}^{\mathrm{cv}} \leq \mathbf{v}_{1:k-1} \leq \mathbf{v}_{1:k-1}^{\mathrm{cc}}\} \\ \text{and} \quad & v_k^{\mathrm{cc}} := \max\,\{w_k^{\mathrm{cc}}(\mathbf{y},\mathbf{v}_{1:k-1},\tilde{Y},\hat{V}_k) : \mathbf{v}_{1:k-1}^{\mathrm{cv}} \leq \mathbf{v}_{1:k-1} \leq \mathbf{v}_{1:k-1}^{\mathrm{cc}}\}. \end{aligned}$$

Define the domain of the optimization problems above as

$$\Gamma_k := \{\mathbf{v}_{1:k-1} \in \hat{V}_k : \mathbf{v}_{1:k-1}^{\mathrm{cv}} \leq \mathbf{v}_{1:k-1} \leq \mathbf{v}_{1:k-1}^{\mathrm{cc}}\},$$

and define a set

$$\tilde{\Gamma}_k := \{\mathbf{v}_{1:k-1} \in \hat{V}_k : w_r^{\mathrm{cv}}(\mathbf{y},\mathbf{v}_{1:r-1},\tilde{Y},\hat{V}_r) \leq v_r \leq w_r^{\mathrm{cc}}(\mathbf{y},\mathbf{v}_{1:r-1},\tilde{Y},\hat{V}_r),\ \forall r \in \{1,...,k-1\}\}.$$

Since (5.4.19) holds, $\tilde{\Gamma}_k \subseteq \Gamma_k$. Moreover, since v_k^{cv} minimizes $w_k^{\mathrm{cv}}(\mathbf{y},\cdot,\tilde{Y},\hat{V}_k)$ on Γ_k and v_k^{cc}

maximizes $w_k^{\mathrm{cc}}(\mathbf{y},\cdot,\tilde{Y},\hat{V}_k)$ on Γ_k, it follows that

$$v_k^{\mathrm{cv}} \leq w_k^{\mathrm{cv}}(\mathbf{y},\mathbf{v}_{1:k-1},\tilde{Y},\hat{V}_k) \leq w_k^{\mathrm{cc}}(\mathbf{y},\mathbf{v}_{1:k-1},\tilde{Y},\hat{V}_k) \leq v_k^{\mathrm{cc}}, \quad \forall \mathbf{v}_{1:k-1} \in \tilde{\Gamma}_k.$$

Finally, (5.4.17) is proved by letting k increase inductively from 3 to m. □

Remark 5.4.21. Theorems 5.4.18 and 5.4.20 together imply the following tightness result. Consider a factorable function h as in Definition 5.4.2, and suppose that $(w_j^{\mathrm{cv}}, w_j^{\mathrm{cc}})$ are at least as tight as $(w_j^{\mathrm{cv,MC}}, w_j^{\mathrm{cc,MC}})$. Then, the AVM relaxations $(h^{\mathrm{cv,AVM}}, h^{\mathrm{cc,AVM}})$ derived from $(w_j^{\mathrm{cv}}, w_j^{\mathrm{cc}})$ as in Definition 5.4.6 are at least as tight as the McCormick relaxations $(h^{\mathrm{cv,MC}}, h^{\mathrm{cc,MC}})$ as in Definition 5.4.5. This result is based on a fundamental assumption that both the AVM and McCormick relaxation methods use the same factor representation in the sense of Definition 5.4.2. However, a factorable function can be factorized in multiple ways, which may in turn result in different AVM relaxations and McCormick relaxations. In this situation, the tightness result in this remark implies that for a given factorable function, the tightest possible McCormick relaxations are guaranteed to be no tighter than the tightest AVM relaxations.

Tightness results for new ODE relaxations Theorem 5.4.23 below shows that for a given original ODE system (2.3.1) with a factorable right-hand side function $\mathbf{f}$, if McCormick relaxations of each multivariate intrinsic function $w_{j,i}$ of f_i are employed in (5.4.10), then our new AVM-based state relaxations are guaranteed to be at least as tight as the OBM relaxations [3] (using McCormick relaxations of $\mathbf{f}$), as introduced in Section 5.2. Note that when tight relaxations $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ of a factorable $\mathbf{f}$ are not directly available, the McCormick relaxation method is a primary method for relaxing such general factorable functions. Theorem 5.4.23 requires the following basic tightness result reframed from [3, Theorem 5] concerning the state relaxation formulation (5.2.2) and (2.4.1) with any appropriate $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$. This result shows that if tighter state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ for (2.3.1), tighter relaxations $(\mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}})$ for $\mathbf{x}_0$, and tighter relaxations $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ are available, then these tighter relaxations will necessarily translate into at least as tight state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ for (2.3.1) through (5.2.2) and (2.4.1).

Proposition 5.4.22 (adapted from [3]). Consider state lower bounds $\mathbf{x}^{\mathrm{L,A}}, \mathbf{x}^{\mathrm{L,B}} : I \to \mathbb{R}^{n_x}$ and state upper bounds $\mathbf{x}^{\mathrm{U,A}}, \mathbf{x}^{\mathrm{U,B}} : I \to \mathbb{R}^{n_x}$ for (2.3.1) that are absolutely continuous, and suppose for all $t \in I$ that $\mathbf{x}^{\mathrm{L,A}}(t) \leq \mathbf{x}^{\mathrm{L,B}}(t) \leq \mathbf{x}^{\mathrm{U,B}}(t) \leq \mathbf{x}^{\mathrm{U,A}}(t)$. For each $t \in I$, denote the intervals $[\mathbf{x}^{\mathrm{L,A}}(t), \mathbf{x}^{\mathrm{U,A}}(t)]$ and $[\mathbf{x}^{\mathrm{L,B}}(t), \mathbf{x}^{\mathrm{U,B}}(t)]$ as $X^{\mathrm{A}}(t)$ and $X^{\mathrm{B}}(t)$, respectively. Consider convex relaxations $\mathbf{x}_0^{\mathrm{cv,A}}, \mathbf{x}_0^{\mathrm{cv,B}} : P \to \mathbb{R}^{n_x}$ and concave relaxations $\mathbf{x}_0^{\mathrm{cc,A}}, \mathbf{x}_0^{\mathrm{cc,B}} : P \to \mathbb{R}^{n_x}$ for the initial-value function $\mathbf{x}_0$ in (2.3.1), and suppose for all $\mathbf{p} \in P$ that $\mathbf{x}_0^{\mathrm{cv,A}}(\mathbf{p}) \leq \mathbf{x}_0^{\mathrm{cv,B}}(\mathbf{p}) \leq \mathbf{x}_0^{\mathrm{cc,B}}(\mathbf{p}) \leq \mathbf{x}_0^{\mathrm{cc,A}}(\mathbf{p})$. Consider functions $\mathbf{f}^{\mathrm{cv,A}}, \mathbf{f}^{\mathrm{cc,A}}, \mathbf{f}^{\mathrm{cv,B}}, \mathbf{f}^{\mathrm{cc,B}} : I \times P \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ for which suppose that for a.e. $t \in I$, $\mathbf{f}^{\mathrm{cv,A(B)}}(t, \cdot, \cdot)$ and $\mathbf{f}^{\mathrm{cc,A(B)}}(t, \cdot, \cdot)$ are respectively convex and concave relaxations for $\mathbf{f}(t, \cdot, \cdot)$ in (2.3.1) on $P \times X^{\mathrm{A(B)}}(t)$, and for each $(\mathbf{p}, \boldsymbol{\xi}) \in P \times X^{\mathrm{B}}(t)$,

$$\boldsymbol{f}^{\mathrm{cv,A}}(t, \mathbf{p}, \boldsymbol{\xi}) \leq \boldsymbol{f}^{\mathrm{cv,B}}(t, \mathbf{p}, \boldsymbol{\xi}) \leq \boldsymbol{f}^{\mathrm{cc,B}}(t, \mathbf{p}, \boldsymbol{\xi}) \leq \boldsymbol{f}^{\mathrm{cc,A}}(t, \mathbf{p}, \boldsymbol{\xi}). \tag{5.4.20}$$

Define a set $S^{\mathrm{A}} := \{(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \in I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} : \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}} \in X^{\mathrm{A}}(t) \text{ and } \boldsymbol{\xi}^{\mathrm{cv}} \leq \boldsymbol{\xi}^{\mathrm{cc}}\}$ and a set $S^{\mathrm{B}} := \{(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \in I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} : \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}} \in X^{\mathrm{B}}(t) \text{ and } \boldsymbol{\xi}^{\mathrm{cv}} \leq \boldsymbol{\xi}^{\mathrm{cc}}\}$. Consider functions $\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ so that $\mathbf{u}^{\mathrm{A}}(t, \mathbf{p}, \cdot, \cdot)$, $\mathbf{o}^{\mathrm{A}}(t, \mathbf{p}, \cdot, \cdot)$, $\mathbf{u}^{\mathrm{B}}(t, \mathbf{p}, \cdot, \cdot)$, and $\mathbf{o}^{\mathrm{B}}(t, \mathbf{p}, \cdot, \cdot)$ are Lipschitz continuous on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly over $(t, \mathbf{p}) \in$

$I \times P$, and also suppose that for each $i \in \{1,...,n_x\}$, for each $(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \in S^{\mathrm{A}}$,

$$u_i^{\mathrm{A}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \equiv \min_{\boldsymbol{\xi} \in [\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cv,A}}(t, \mathbf{p}, \boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cv}},$$

$$o_i^{\mathrm{A}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \equiv \max_{\boldsymbol{\xi} \in [\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cc,A}}(t, \mathbf{p}, \boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cc}},$$

and for each $(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \in S^{\mathrm{B}}$,

$$u_i^{\mathrm{B}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \equiv \min_{\boldsymbol{\xi} \in [\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cv,B}}(t, \mathbf{p}, \boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cv}},$$

$$o_i^{\mathrm{B}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \equiv \max_{\boldsymbol{\xi} \in [\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cc,B}}(t, \mathbf{p}, \boldsymbol{\xi}) \text{ subject to } \xi_i = \xi_i^{\mathrm{cc}}.$$

Let $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}, \mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}}, \mathbf{u}, \mathbf{o}) \leftarrow (\mathbf{x}^{\mathrm{L,A}}, \mathbf{x}^{\mathrm{U,A}}, \mathbf{x}_0^{\mathrm{cv,A}}, \mathbf{x}_0^{\mathrm{cc,A}}, \mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}}).$$

Let $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}, \mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}}, \mathbf{u}, \mathbf{o}) \leftarrow (\mathbf{x}^{\mathrm{L,B}}, \mathbf{x}^{\mathrm{U,B}}, \mathbf{x}_0^{\mathrm{cv,B}}, \mathbf{x}_0^{\mathrm{cc,B}}, \mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}}).$$

Then, $(\mathbf{x}^{\mathrm{cv,A}}, \mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}}, \mathbf{x}^{\mathrm{cc,B}})$ are both valid state relaxations for (2.3.1). Moreover, for any $(t, \mathbf{p}) \in I \times P$,

$$\mathbf{x}^{\mathrm{cv,A}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cv,B}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,B}}(t, \mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,A}}(t, \mathbf{p}).$$

Theorem 5.4.23. Consider functions $(\mathbf{x}^{\mathrm{L,A}}, \mathbf{x}^{\mathrm{L,B}}, \mathbf{x}^{\mathrm{U,A}}, \mathbf{x}^{\mathrm{U,B}})$ and $(\mathbf{x}_0^{\mathrm{cv,A}}, \mathbf{x}_0^{\mathrm{cv,B}}, \mathbf{x}_0^{\mathrm{cc,A}}, \mathbf{x}_0^{\mathrm{cc,B}})$, intervals $X^{\mathrm{A}}(t)$ and $X^{\mathrm{B}}(t)$, and sets S^{A} and S^{B} as in Proposition 5.4.22. Suppose that each multivariate intrinsic function $w_{j,i}$ for representing f_i in Section 5.4.2 is factorable in the sense of [6, Definition 8]. Following Definition 5.4.5, define McCormick relaxations $\mathbf{f}^{\mathrm{cv,MC}}, \mathbf{f}^{\mathrm{cc,MC}} : I \times P \times D \times \mathbb{I}P \times \mathbb{I}D \to \mathbb{R}^{n_x}$ so that for each $t \in I$, $\mathbf{f}^{\mathrm{cv,MC}}(t,\cdot,\cdot,P,X^{\mathrm{A}}(t))$ and $\mathbf{f}^{\mathrm{cc,MC}}(t,\cdot,\cdot,P,X^{\mathrm{A}}(t))$ are respectively convex and concave relaxations of $\mathbf{f}(t,\cdot,\cdot)$ on $P \times X^{\mathrm{A}}(t)$. Define functions $\bar{\mathbf{u}}^{\mathrm{A}}, \bar{\mathbf{o}}^{\mathrm{A}} : S^{\mathrm{A}} \to \mathbb{R}^{n_x}$ so that for each $(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \in S^{\mathrm{A}}$,

$$\begin{aligned} \bar{u}_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &:= \min_{\boldsymbol{\xi} \in [\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cv,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{A}}(t)) \text{ subject to } \xi_i = \xi_i^{\mathrm{cv}}, \\ \bar{o}_i^{\mathrm{A}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &:= \max_{\boldsymbol{\xi} \in [\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cc,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{A}}(t)) \text{ subject to } \xi_i = \xi_i^{\mathrm{cc}}. \end{aligned} \tag{5.4.21}$$

For each $i \in \{1,...,n_x\}$ and $j \in J_i$, consider McCormick relaxations $(w_{j,i}^{\mathrm{cv,MC}}, w_{j,i}^{\mathrm{cv,MC}})$ for the intrinsic functions $w_{j,i}$ of f_i, and for each $t \in I$, consider sets $\hat{V}_{j,i}^{\mathrm{B}}(t)$ defined using (5.4.9) with $X(t) \leftarrow X^{\mathrm{B}}(t)$ and $\hat{V}_{j,i}(t) \leftarrow \hat{V}_{j,i}^{\mathrm{B}}(t)$. Define functions $\bar{\mathbf{u}}^{\mathrm{B}}, \bar{\mathbf{o}}^{\mathrm{B}} : S^{\mathrm{B}} \to \mathbb{R}^{n_x}$ using (5.4.10) with $(\bar{\mathbf{u}},\bar{\mathbf{o}}) \leftarrow (\bar{\mathbf{u}}^{\mathrm{B}},\bar{\mathbf{o}}^{\mathrm{B}})$, $(w_{j,i}^{\mathrm{cv}}, w_{j,i}^{\mathrm{cc}}) \leftarrow (w_{j,i}^{\mathrm{cv,MC}}, w_{j,i}^{\mathrm{cc,MC}})$, and $(X(t),\hat{V}_{j,i}(t)) \leftarrow (X^{\mathrm{B}}(t),\hat{V}_{j,i}^{\mathrm{B}}(t))$. Suppose that Assumption 5.4.10 is satisfied with $(\bar{\mathbf{u}},\bar{\mathbf{o}}) \leftarrow (\bar{\mathbf{u}}^{\mathrm{A}},\bar{\mathbf{o}}^{\mathrm{A}})$ and with $(\bar{\mathbf{u}},\bar{\mathbf{o}}) \leftarrow (\bar{\mathbf{u}}^{\mathrm{B}},\bar{\mathbf{o}}^{\mathrm{B}})$. Consider Lipschitz extensions $\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}},\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ of $(\bar{\mathbf{u}}^{\mathrm{A}},\bar{\mathbf{o}}^{\mathrm{A}},\bar{\mathbf{u}}^{\mathrm{B}},\bar{\mathbf{o}}^{\mathrm{B}})$ so that Assumption 5.4.11 is satisfied with $(\mathbf{u},\mathbf{o}) \leftarrow (\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}})$ and with $(\mathbf{u},\mathbf{o}) \leftarrow (\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}})$. Let $(\mathbf{x}^{\mathrm{cv,A}},\mathbf{x}^{\mathrm{cc,A}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}},\mathbf{x}_0^{\mathrm{cv}},\mathbf{x}_0^{\mathrm{cc}},\mathbf{u},\mathbf{o}) \leftarrow (\mathbf{x}^{\mathrm{L,A}},\mathbf{x}^{\mathrm{U,A}},\mathbf{x}_0^{\mathrm{cv,A}},\mathbf{x}_0^{\mathrm{cc,A}},\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}}).$$

Let $(\mathbf{x}^{\mathrm{cv,B}},\mathbf{x}^{\mathrm{cc,B}})$ be a solution of (2.4.1) with

$$(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}},\mathbf{x}_0^{\mathrm{cv}},\mathbf{x}_0^{\mathrm{cc}},\mathbf{u},\mathbf{o}) \leftarrow (\mathbf{x}^{\mathrm{L,B}},\mathbf{x}^{\mathrm{U,B}},\mathbf{x}_0^{\mathrm{cv,B}},\mathbf{x}_0^{\mathrm{cc,B}},\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}}).$$

Then, $(\mathbf{x}^{\mathrm{cv,A}},\mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}},\mathbf{x}^{\mathrm{cc,B}})$ are both valid state relaxations for (2.3.1). Moreover, for any $(t,\mathbf{p}) \in I \times P$,

$$\mathbf{x}^{\mathrm{cv,A}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cv,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,B}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc,A}}(t,\mathbf{p}). \tag{5.4.22}$$

Proof. By construction, on the set S^{A}, $(\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}}) \equiv (\bar{\mathbf{u}}^{\mathrm{A}},\bar{\mathbf{o}}^{\mathrm{A}})$, and on S^{B}, $(\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}}) \equiv (\bar{\mathbf{u}}^{\mathrm{B}},\bar{\mathbf{o}}^{\mathrm{B}})$. Moreover, by a similar argument to (5.4.12), $(\bar{\mathbf{u}}^{\mathrm{B}},\bar{\mathbf{o}}^{\mathrm{B}})$ are equivalent to the following: for each $i \in \{1,...,n_x\}$ and each $(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \in S^{\mathrm{B}}$,

$$\begin{aligned}
\bar{u}_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &\equiv \min_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cv,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t)) \text{ subject to } \xi_i = \xi_i^{\mathrm{cv}},\\
\bar{o}_i^{\mathrm{B}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &\equiv \max_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cc,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t)) \text{ subject to } \xi_i = \xi_i^{\mathrm{cc}}
\end{aligned} \tag{5.4.23}$$

where the AVM relaxations $f_i^{\mathrm{cv,AVM}}, f_i^{\mathrm{cc,AVM}} : I \times P \times D \times \mathbb{I}P \times \mathbb{I}D \to \mathbb{R}$ are constructed following Definition 5.4.6 so that for each $t \in I$, $\mathbf{f}^{\mathrm{cv,AVM}}(t,\cdot,\cdot,P,X^{\mathrm{B}}(t))$ and $\mathbf{f}^{\mathrm{cc,AVM}}(t,\cdot,\cdot,P,X^{\mathrm{B}}(t))$ are respectively convex and concave relaxations of $\mathbf{f}(t,\cdot,\cdot)$ on $P \times X^{\mathrm{B}}(t)$. Since Assumption 5.4.11 is satisfied both with $(\mathbf{u},\mathbf{o}) \leftarrow (\mathbf{u}^{\mathrm{A}},\mathbf{o}^{\mathrm{A}})$ and with $(\mathbf{u},\mathbf{o}) \leftarrow (\mathbf{u}^{\mathrm{B}},\mathbf{o}^{\mathrm{B}})$, and since (5.4.21) and (5.4.23) hold, $(\mathbf{x}^{\mathrm{cv,A}},\mathbf{x}^{\mathrm{cc,A}})$ and $(\mathbf{x}^{\mathrm{cv,B}},\mathbf{x}^{\mathrm{cc,B}})$ are both valid state relaxations for (2.3.1), according to Proposition 5.4.22.

Now, we prove (5.4.22) by verifying the tightness hypothesis (5.4.20) in Proposition 5.4.22 with the following substitution: for each $t \in I$, $\mathbf{p} \in P$, and $\boldsymbol{\xi} \in X^{\mathrm{B}}(t)$,

$$\begin{aligned}
\mathbf{f}^{\mathrm{cv,A}}(t,\mathbf{p},\boldsymbol{\xi}) &\leftarrow \mathbf{f}^{\mathrm{cv,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{A}}(t)),\\
\mathbf{f}^{\mathrm{cc,A}}(t,\mathbf{p},\boldsymbol{\xi}) &\leftarrow \mathbf{f}^{\mathrm{cc,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{A}}(t)),\\
\mathbf{f}^{\mathrm{cv,B}}(t,\mathbf{p},\boldsymbol{\xi}) &\leftarrow \mathbf{f}^{\mathrm{cv,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t)),\\
\mathbf{f}^{\mathrm{cc,B}}(t,\mathbf{p},\boldsymbol{\xi}) &\leftarrow \mathbf{f}^{\mathrm{cc,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t)).
\end{aligned}$$

Consider any $t \in I$, $\mathbf{p} \in P$, and $\boldsymbol{\xi} \in X^{\mathrm{B}}(t)$. Since McCormick relaxations $(w_{j,i}^{\mathrm{cv,MC}}, w_{j,i}^{\mathrm{cc,MC}})$ of $w_{j,i}$ are used for defining $(\mathbf{f}^{\mathrm{cv,AVM}}, \mathbf{f}^{\mathrm{cc,AVM}})$, Theorems 5.4.18 and 5.4.20 together imply that

$$\begin{aligned}\mathbf{f}^{\mathrm{cv,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t)) &\leq \mathbf{f}^{\mathrm{cv,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t))\\ &\leq \mathbf{f}^{\mathrm{cc,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t)) \leq \mathbf{f}^{\mathrm{cc,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t)).\end{aligned}$$

Since McCormick relaxations are *partition monotonic* as in [39, Theorem 2.6.5], and since $X^{\mathrm{B}}(t) \subseteq X^{\mathrm{A}}(t)$, it follows that

$$\begin{aligned}\mathbf{f}^{\mathrm{cv,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{A}}(t)) &\leq \mathbf{f}^{\mathrm{cv,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t))\\ &\leq \mathbf{f}^{\mathrm{cc,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t)) \leq \mathbf{f}^{\mathrm{cc,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{A}}(t)).\end{aligned}$$

Combining the inequalities above yields

$$\begin{aligned}\mathbf{f}^{\mathrm{cv,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{A}}(t)) &\leq \mathbf{f}^{\mathrm{cv,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t))\\ &\leq \mathbf{f}^{\mathrm{cc,AVM}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{B}}(t)) \leq \mathbf{f}^{\mathrm{cc,MC}}(t,\mathbf{p},\boldsymbol{\xi},P,X^{\mathrm{A}}(t)),\end{aligned}$$

which satisfies (5.4.20) in Proposition 5.4.22. □

In the theorem above, the ODE system (2.4.1) with $(\mathbf{u}^{\mathrm{A}}, \mathbf{o}^{\mathrm{A}})$ embedded comes from the OBM relaxation approach in [3], and (2.4.1) with $(\mathbf{u}^{\mathrm{B}}, \mathbf{o}^{\mathrm{B}})$ embedded comes from the new AVM-based ODE relaxation method in this chapter.

Remark 5.4.24. Proposition 5.4.22 and Theorem 5.4.23 also imply the following tightness results:

1. By using $(w_{j,i}^{\mathrm{cv,MC}}, w_{j,i}^{\mathrm{cc,MC}})$, the new state relaxations in this chapter are also at least as tight as the SBM relaxations [2], since it is shown in [3] that the OBM relaxations

are at least as tight as the SBM relaxations.

2. If non-McCormick relaxations $(w_{j,i}^{\mathrm{cv}}, w_{j,i}^{\mathrm{cc}})$ that are at least as tight as $(w_{j,i}^{\mathrm{cv,MC}}, w_{j,i}^{\mathrm{cc,MC}})$ are available, then using these for defining the new $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ in (5.4.10) also leads to state relaxations that are at least as tight as the OBM relaxations.

3. Suppose that there are two considered choices of relaxations for $w_{j,i}$; call these $(w_{j,i}^{\mathrm{cv,B}}, w_{j,i}^{\mathrm{cc,B}})$ and $(w_{j,i}^{\mathrm{cv,A}}, w_{j,i}^{\mathrm{cc,A}})$. If $(w_{j,i}^{\mathrm{cv,B}}, w_{j,i}^{\mathrm{cc,B}})$ are at least as tight as $(w_{j,i}^{\mathrm{cv,A}}, w_{j,i}^{\mathrm{cc,A}})$, then using $(w_{j,i}^{\mathrm{cv,B}}, w_{j,i}^{\mathrm{cc,B}})$ in the new AVM-based ODE relaxation system necessarily leads to state relaxations that are at least as tight as these obtained by using $(w_{j,i}^{\mathrm{cv,A}}, w_{j,i}^{\mathrm{cc,A}})$.

4. The TMC relaxations $(\mathbf{f}^{\mathrm{cv,TMC}}, \mathbf{f}^{\mathrm{cc,TMC}})$ in Definition 5.4.4 of $\mathbf{f}$ may also be used for generating state relaxations via (5.2.2) and (2.4.1) (analogously denoted as the *optimization–based–TMC (OBT) relaxations*). Theorems 5.4.18 and 5.4.20 imply that for a given factorable function, the TMC relaxations are no looser than the MC relaxations, and are no tighter than the AVM relaxations. Combining this with Proposition 5.4.22, it can be verified that the OBT state relaxations are no looser than the OBM relaxations, and are no tighter than the new AVM-based relaxations. In this chapter, we focus on the AVM-based relaxations rather than the OBT relaxations, since the AVM-based relaxations are the tightest among these potential relaxations.

Remark 5.4.25. Since the AVM-based relaxations are at least as tight as the OBM relaxations and since the OBM relaxations have *second-order pointwise convergence* [36, 79] as proved in [3, Theorem 6], the new AVM-based relaxations in this chapter also have such desirable convergence property, which can help mitigate the *cluster effect* [37, 38] in branch-and-bound methods for deterministic nonconvex optimization. Moreover, due to the same tightness result, similarly to the OB relaxations as discussed in [3, Remark 10 and Example 3], the new AVM-based relaxations $(\mathbf{x}^{\mathrm{cv}}(t,\cdot),\mathbf{x}^{\mathrm{cc}}(t,\cdot))$ may get tighter over time t under appropriate conditions.

5.4.5 Implementation

Proof-of-concept implementations were developed in Julia v1.4.2 [95] to compute state relaxations for (2.3.1) by constructing and solving the auxiliary ODE system (2.4.1) with $(\mathbf{u},\mathbf{o})$ satisfying Assumption 5.4.11. According to Remark 5.4.13, we practically construct $(\bar{\mathbf{u}},\bar{\mathbf{o}})$ defined in (5.4.10) in place of the Lipschitz extensions $(\mathbf{u},\mathbf{o})$ in (2.4.1), which would yield identical solutions. Harrison's bounding method [69] is employed to compute state bounds $(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}})$ automatically. For any given instance of the original ODE system (2.3.1), the initial-value function $\mathbf{x}_0$ is relaxed using the McCormick relaxation method [5], and the right-hand side function $\mathbf{f}$ is factorized manually. Natural interval extension [48] is employed for constructing all factor bounds $(w^{\mathrm{L}}_{j,i},w^{\mathrm{U}}_{j,i})$, and the optimal-value functions $(\bar{\mathbf{u}},\bar{\mathbf{o}})$ defined in (5.4.10) are hard-coded with predefined $(w^{\mathrm{cv}}_{j,i},w^{\mathrm{cc}}_{j,i})$ for each numerical example in the next subsection. For each $(t,\mathbf{p})\in I\times P$, unless otherwise specified, the right-hand side optimization problems in (5.4.10) are solved numerically using local optimizer IPOPT v3.13.2 [120] with an overall tolerance of 10^{-6}. Any necessary user-provided local sensitivities are approximated using central finite difference with a step length of 10^{-10}. JuMP v0.21.3 [121] is employed as an interface with IPOPT. The auxiliary ODE system (2.4.1) with $(\bar{\mathbf{u}},\bar{\mathbf{o}})$ in place of $(\mathbf{u},\mathbf{o})$ is solved using the ODE solver `BS3()` with an absolute tolerance of 10^{-4} and a relative tolerance of 10^{-4} from the package DifferentialEquations v6.15.0 [123]. For comparison, we also developed an analogous implementation in Julia for the OBM relaxations [3], which automatically constructs and solves optimization problems in (5.2.2) with McCormick relaxations [5] ($\mathbf{f}^{\mathrm{cv,MC}}$,$\mathbf{f}^{\mathrm{cc,MC}}$) embedded. This implementation is similar to the previous MATLAB implementation proposed in [3, Section 6.1]. An implementation of the SBM relaxations [2] was also developed in Julia. Any necessary computations of natural interval extension and McCormick relaxations are executed via operator overloading using EAGO v0.4.1 [119]. All computations in the next subsection were performed on the same computer as in Example 5.1.

5.4.6 Numerical examples

As established in Theorem 5.4.23 and Remark 5.4.24, the new AVM-based state relaxations in this chapter are guaranteed to be at least as tight as both the SBM relaxations [2] and the OBM relaxations [3]. In this section, we present several numerical examples to show that, if $\mathbf{f}$ has certain features, then the new AVM-based state relaxations may be significantly tighter than the two other types of relaxations. Example 5.3 shows that the AVM-based relaxation method can yield tighter state relaxations by recognizing and eliminating *repeated factors* (in Definition 5.4.26 below) of $\mathbf{f}$. Example 5.4 shows that if $\mathbf{f}$ contains multivariate intrinsic functions with known convex envelopes, then the new relaxation method can utilize these envelopes to yield tighter state relaxations. In Example 5.5, we apply the AVM-based relaxations to a chemical reaction network model from [74, Example 1], whose right-hand side has both repeated factors and multivariate intrinsic functions with known convex envelopes. The results show that the new relaxations are significantly tighter than the SBM relaxations and the OBM relaxations.

Definition 5.4.26. Given $Y \subset \mathbb{R}^n$, consider a factorable function $h : Y \to \mathbb{R}$ with related factors $v_1,...,v_m$ in Definition 5.4.2. For each $\mathbf{y} \in Y$, for any $i, j \in \{1,...,m\}$, v_i and v_j are *repeated factors* if $v_i \equiv v_j$.

In the following example, we show how repeated factors of $\mathbf{f}$ can be eliminated in the new AVM-based ODE relaxation formulation, to yield tighter state relaxations than established relaxations [2,3].

Example 5.3. Let $P := [-0.4, 0.7]$, and $I := [0, 0.06]$, and consider the following instance of (2.3.1) with one state variable x and one parameter $p \in P$:

$$\begin{aligned} \dot{x}(t) &= e^{3x} - e^{2x} - e^{x}, \quad \forall t \in I, \\ x(0) &= p - p^3. \end{aligned} \tag{5.4.24}$$

A standard operator overloading procedure (e.g. employed by EAGO [94,119] or MC++ [96]) may factorize the right-hand side function $f(\xi) : \xi \mapsto e^{3\xi} - e^{2\xi} - e^{\xi}$ using the following factor representation: for any $\xi \in \mathbb{R}$,

$$v_1 = e^{\xi},\ v_2 = v_1^3,\ v_1' = e^{\xi},\ v_3 = (v_1')^2,\ v_1'' = e^{\xi},\ \text{and}\ f(\xi) = v_4 = v_2 - v_3 - v_1''.$$

Observe that v_1, v_1', v_1'' are repeated factors as in Definition 5.4.26. Thus, we used the following factor representation of f by recognizing these repeated factors as one common factor v_1, for constructing the right-hand side functions $(\bar{u}, \bar{o})$ defined in (5.4.10) of the new AVM-based ODE relaxation system:

$$v_1 = e^{\xi},\ v_2 = v_1^3,\ v_3 = v_1^2,\ \text{and}\ f(\xi) = v_4 = v_2 - v_3 - v_1.$$

The intrinsic cubic function was bounded by its convex and concave envelopes [52], and the convex exponential and square functions were bounded above by the affine envelopes.

Thus, the functions $(\bar{u}, \bar{o})$ defined in (5.4.10) were constructed, and the new AVM-based state relaxations for (5.4.24) were computed using the proposed proof-of-concept implementation in Julia. The SBM relaxations [6] were also computed for comparison. Note that the OBM relaxations [3] are identical to the SBM relaxations for $n_x = 1$, as discussed in [3, Remark 4]. Figure 5.2 depicts the original parametric solution $x(0.06, p)$, along with the constructed SBM relaxations and AVM-based ODE relaxations for comparison. Observe that the new AVM-based relaxations are visually tighter than the SBM relaxations, which is consistent with the discussion in Remark 5.4.27. Table 5.2 summarizes the computational times for per-mesh-point evaluation of the two types of relaxations. It can be seen that the AVM-based relaxations took longer time to evaluate, since the convex optimization problems in (5.2.2) were naively solved with IPOPT, while SBM relaxation system's right-hand side was efficiently evaluated in closed form. We expect that this implementation method may be improved with the techniques outlined in Section 5.4.3.

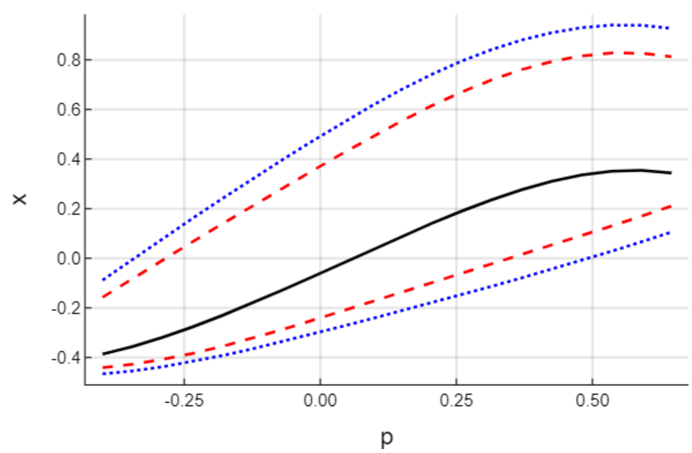

Figure 5.2: The solution $x(0.06, p)$ (*solid black*) of the parametric ODE (5.4.24) from Example 5.3, plotted against p along with corresponding SBM relaxations [2] (*dotted blue*) and the AVM-based relaxations (*dashed red*).

Remark 5.4.27. It has been discussed in [7, 45] that the AVM relaxation method provides

Table 5.2: Average computational times for evaluating state relaxations $(x^{\mathrm{cv}}(0.06, p), x^{\mathrm{cc}}(0.06, p))$ for (5.4.24) in Example 5.3.

State relaxation method	CPU time (seconds)*
SBM relaxations	0.07
New AVM-based relaxations	1.05

* Each CPU time here was averaged over 10 runs, with a sample standard deviation that is much smaller than the reported average.

potentially tighter bounds than the McCormick relaxation method due to repeated terms. This result can be rigorously proved by a straightforward extension of Theorems 5.4.18 and 5.4.20. Then, Theorem 5.4.23 essentially translates such tightness of relaxations from non-dynamic factorable functions to ODE solutions.

In the following example, we show how convex envelopes of multivariate intrinsic functions in $\mathbf{f}$ can be employed by the AVM-based ODE relaxation approach, to yield tighter state relaxations than established relaxations [2, 3].

Example 5.4. Let $P := [0.8, 1.3]$, and $I := [0, 0.1]$, and consider the following instance of (2.3.1) with one state variable x and one parameter $p \in P$:

$$\begin{aligned} \dot{x}(t) &= pe^{-x}(x^4 - 3x^2 - x + 0.4), \\ x(0) &= p - \frac{p^3}{3}. \end{aligned} \tag{5.4.25}$$

Based on Definition 5.4.2, we factorized the right-hand side function $f : (p, \xi) \mapsto pe^{-\xi}(\xi^4 - 3\xi^2 - \xi + 0.4)$ into the following factor representation: for each $(p, \xi) \in P \times \mathbb{R}$,

$$v_1 = pe^{-\xi}, \quad v_2 = \xi^4 - 3\xi^2 - \xi + 0.4, \quad v_3 = v_1 v_2, \quad \text{and} \quad f(p, \xi) = v_3. \tag{5.4.26}$$

Then, we employed the following convex and concave relaxations for bounding the multivariate intrinsic functions above. With $(x^{\mathrm{L}}, x^{\mathrm{U}})$ denoting Harrison state bounds for (5.4.25), it was empirically verified that for each $t \in I$, $0 \leq x^{\mathrm{L}}(t) \leq x^{\mathrm{U}}(t) \leq 1.3$. Furthermore, since $p^{\mathrm{L}} > 0$, the convex envelope $h^{\mathrm{cv,env}}$ of the intrinsic function $h : (p, \xi) \mapsto pe^{-\xi}$ was given by (5.3.2). Since there is no established concave envelope of h, we employed McCormick concave relaxation $h^{\mathrm{cc,MC}}$ as constructed in (5.3.5). For the intrinsic function $g : \xi \mapsto \xi^4 - 3\xi^2 - \xi + 0.4$, we developed convex and concave envelopes on any subinterval of $[0, 1.3]$ (recall $0 \leq x^{\mathrm{L}}(t) \leq x^{\mathrm{U}}(t) \leq 1.3$) following the procedures presented in [5,51], as follows. Define a set $\Xi := \{(\xi, \xi^L, \xi^{\mathrm{U}}) \in \mathbb{R}^3 : 0 \leq \xi^{\mathrm{L}} \leq \xi \leq \xi^{\mathrm{U}} \leq 1.3\}$. Consider the first-order derivative function $g' : \xi \mapsto 4\xi^3 - 6\xi^2 - 1$ of g. We constructed functions $g^{\mathrm{cv,env}}, g^{\mathrm{cc,env}} : \Xi \to \mathbb{R}$ so that for any $[\xi^{\mathrm{L}}, \xi^{\mathrm{U}}] \subseteq [0, 1.3]$, the mappings $g^{\mathrm{cv,env}}(\cdot, \xi^{\mathrm{L}}, \xi^U)$ and $g^{\mathrm{cc,env}}(\cdot, \xi^{\mathrm{L}}, \xi^U)$ are respectively convex and concave envelopes of g on $[\xi^{\mathrm{L}}, \xi^U]$. For each $(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) \in \Xi$, $g^{\mathrm{cv,env}}$ and $g^{\mathrm{cc,env}}$ are thus evaluated as follows:

- if $\xi^{\mathrm{U}} \leq \frac{\sqrt{2}}{2}$,

$$g^{\mathrm{cv,env}}(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) := \frac{g(\xi^{\mathrm{U}}) - g(\xi^{\mathrm{L}})}{\xi^{\mathrm{U}} - \xi^{\mathrm{L}}}(\xi - \xi^{\mathrm{L}}) + g(\xi^{\mathrm{L}}),$$
$$g^{\mathrm{cc,env}}(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) := g(\xi),$$

- if $\xi^{\mathrm{L}} \geq \frac{\sqrt{2}}{2}$,

$$g^{\mathrm{cv,env}}(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) := g(\xi),$$
$$g^{\mathrm{cc,env}}(\xi, \xi^{\mathrm{L}}, \xi^{\mathrm{U}}) := \frac{g(\xi^{\mathrm{U}}) - g(\xi^{\mathrm{L}})}{\xi^{\mathrm{U}} - \xi^{\mathrm{L}}}(\xi - \xi^{\mathrm{L}}) + g(\xi^{\mathrm{L}}),$$

- if $\xi^{\mathrm{L}} < \frac{\sqrt{2}}{2} < \xi^{\mathrm{U}}$, with $\xi^{\mathrm{A}} := \frac{-\xi^{\mathrm{L}}+\sqrt{9-2(\xi^{\mathrm{L}})^2}}{3}$ and $\xi^{\mathrm{B}} := \frac{-\xi^{\mathrm{U}}+\sqrt{9-2(\xi^{\mathrm{U}})^2}}{3}$,

$$g^{\mathrm{cv,env}}(\xi,\xi^{\mathrm{L}},\xi^{\mathrm{U}}) := \begin{cases} \frac{g(\xi^{\mathrm{U}})-g(\xi^{\mathrm{L}})}{\xi^{\mathrm{U}}-\xi^{\mathrm{L}}}(\xi-\xi^{\mathrm{L}})+g(\xi^{\mathrm{L}}), & \text{if } \xi^{\mathrm{U}} \le \xi^{A}, \\ g(\xi), & \text{if } \xi^{\mathrm{U}} > \xi^{A} \text{ and } \xi \ge \xi^{\mathrm{A}}, \\ g'(\xi^{\mathrm{A}})(\xi-\xi^{\mathrm{A}})+g(\xi^{\mathrm{A}}), & \text{if } \xi^{\mathrm{U}} > \xi^{A} \text{ and } \xi < \xi^{\mathrm{A}}, \end{cases}$$

$$g^{\mathrm{cc,env}}(\xi,\xi^{\mathrm{L}},\xi^{\mathrm{U}}) := \begin{cases} \frac{g(\xi^{\mathrm{U}})-g(\xi^{\mathrm{L}})}{\xi^{\mathrm{U}}-\xi^{\mathrm{L}}}(\xi-\xi^{\mathrm{L}})+g(\xi^{\mathrm{L}}), & \text{if } \xi^{\mathrm{L}} \ge \xi^{\mathrm{B}}, \\ g(\xi) & \text{if } \xi^{\mathrm{L}} < \xi^{\mathrm{B}} \text{ and } \xi \le \xi^{\mathrm{B}}, \\ g'(\xi^{\mathrm{B}})(\xi-\xi^{\mathrm{B}})+g(\xi^{\mathrm{B}}), & \text{if } \xi^{\mathrm{L}} < \xi^{\mathrm{B}} \text{ and } \xi > \xi^{\mathrm{B}}. \end{cases}$$

Note that the same function g was also investigated in [3, Example 5], where similar convex and concave envelopes were constructed on a different range of $(\xi^{\mathrm{L}},\xi^{\mathrm{U}})$. Lastly, the bilinear intrinsic function in (5.4.26) was bounded by the McCormick envelope [5].

Functions $(\bar{u},\bar{o})$ defined in (5.4.10) were then constructed based on the factor representation (5.4.26) and the constructed $(g^{\mathrm{cv,env}},g^{\mathrm{cc,env}},h^{\mathrm{cv,env}},h^{\mathrm{cc,MC}})$ and the McCormick envelopes of bilinear terms. Observe that in this case, the right-hand side optimization problems for defining $(\bar{u},\bar{o})$ are in fact linear optimization problems, even though nonlinear relaxations were used for the intrinsic functions. Thus, the LP solver Clp v0.8.3 (available at: https://github.com/jump-dev/Clp.jl) was employed to evaluate $(\bar{u},\bar{o})$. Next, the AVM-based ODE relaxations and SBM relaxations were computed using the proposed proof-of-concept implementation in Julia. Similarly to Example 5.3, the OBM relaxations are identical to the SBM relaxations in this case, since there is only one state variable. Figure 5.3 depicts the resulting state relaxations and shows that the AVM-based relaxations are visually tighter than the SBM relaxations; this is consistent with Results 1 and 2 in Remark 5.4.24. Such tightness benefits from the fact that the AVM-based relaxation formulation utilizes convex envelopes for multivariate intrinsic functions, which are not employed in SBM relaxation method. Table 5.3 shows that the AVM-based relaxations took longer CPU time, which is expected since the implementation naively solves the right-hand side optimization problems in (5.4.10) using the LP solver Clp.

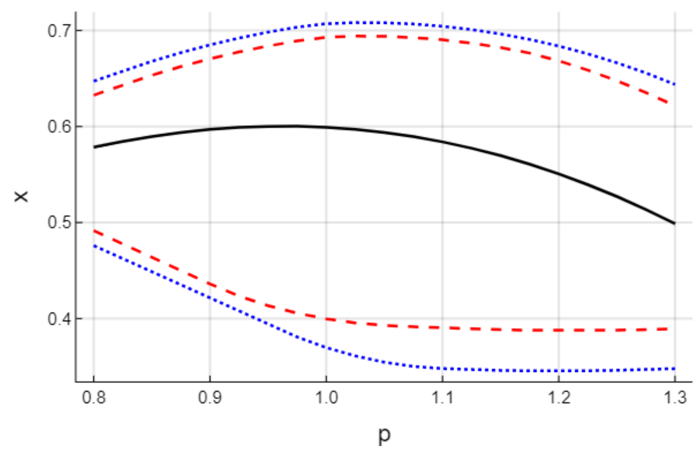

Figure 5.3: The solution $x(0.1, p)$ (*solid black*) of the parametric ODE (5.4.25) from Example 5.4, plotted against p along with corresponding SBM relaxations [2] (*dotted blue*) and the AVM-based state relaxations (*dashed red*).

Table 5.3: Average computational times for evaluating state relaxations $(x^{\mathrm{cv}}(0.1, p), x^{\mathrm{cc}}(0.1, p))$ for (5.4.25) in Example 5.4.

State relaxation method	CPU time (seconds)*
SBM relaxations	0.12
New AVM-based relaxations	0.91

* Each CPU time here was averaged over 10 runs, with a sample standard deviation that is much smaller than the reported average.

In the following example, we apply the AVM-based ODE relaxation approach to a chemical reaction network model whose right-hand side functions have both repeated factors and multivariate intrinsic functions with pre-known convex envelopes. We will show that with a factor representation that has repeated factors, the AVM-based $(\bar{u}_i, \bar{o}_i)$ defined in (5.4.10) may be efficiently evaluated in closed form, yet provide significantly tighter state relaxations than the SBM and OBM relaxations. By recognizing and eliminating the repeated factors, different right-hand side functions $(\bar{u}_i^{\mathrm{NLP}}, \bar{o}_i^{\mathrm{NLP}})$ can be constructed and evaluated with numerical NLP solvers. Using such $(\bar{u}_i^{\mathrm{NLP}}, \bar{o}_i^{\mathrm{NLP}})$ in the AVM-based relaxation formulation yields further tighter state relaxations than using the closed-form $(\bar{u}_i, \bar{o}_i)$, but also requires longer computational time. Then, we construct new $(\bar{u}_i^{\mathrm{LP}}, \bar{o}_i^{\mathrm{LP}})$ that involves solving LP problems. These LPs are constructed from subtangents of the original nonlinear convex relaxations of multivariate intrinsic functions. The results show that by using $(\bar{u}_i^{\mathrm{LP}}, \bar{o}_i^{\mathrm{LP}})$, the AVM-based ODE relaxations can be evaluated much more efficiently than using $(\bar{u}_i^{\mathrm{NLP}}, \bar{o}_i^{\mathrm{NLP}})$, yet without compromising much tightness of the state relaxations.

Example 5.5. This example is from [74, Example 1]. Consider a chemical reaction network $\mathrm{A} \longrightarrow \mathrm{B} \longrightarrow \mathrm{C}$. Assume elementary reactions and Arrhenius rate constants, and denote the concentrations of the chemical species $\mathrm{A}, \mathrm{B}, \mathrm{C}$ as x_1, x_2, x_3, respectively. Then, this reaction network may be modelled using the following ODEs on a time horizon $I := [0, 0.02]$:

$$\begin{aligned}
\dot{x}_1(t) &= -A_1 e^{-E_1/(RT)} x_1, \\
\dot{x}_2(t) &= A_1 e^{-E_1/(RT)} x_1 - A_2 e^{-E_2/(RT)} x_2, \\
\dot{x}_3(t) &= -A_2 e^{-E_2/(RT)} x_2,
\end{aligned} \tag{5.4.27}$$

with $(x_1(0), x_2(0), x_3(0)) = (1.5, 0.5, 0.0)(\frac{\mathrm{mol}}{\mathrm{L}})$. The uppercase letters above denote physical quantities. The temperature T is considered as an uncertain parameter within the bounds

$[T^{\mathrm{L}}, T^{\mathrm{U}}] := [300, 600](\mathrm{K})$. Note that [74, Example 1] considers T as a time-varying control, while here T is assumed to be an uncertain constant along the time horizon. Nevertheless, the setting here may be useful in the context of global optimal control, e.g. assuming that the control T is piecewise constant. The values of the other physical quantities are listed in Table 5.4.

Table 5.4: The values of the constant physical quantities in (5.4.27), for Example 5.5.

Quantities	Values	Units
A_1	2400	$1/s$
A_2	8800	$1/s$
E_1	6.9×10^3	$\mathrm{J/mol}$
E_2	1.69×10^4	$\mathrm{J/mol}$
R	8.314	$\mathrm{J/(mol \cdot K)}$

Now, we show that for this system, how the AVM-based relaxation system's right-hand side functions $(\bar{\mathbf{u}}, \bar{\mathbf{o}})$ defined in (5.4.10) can be evaluated in closed form. For example, we employed the following factor representation of $f_2 : (T, \xi_1, \xi_2) \mapsto A_1 e^{-E_1/(RT)} \xi_1 - A_2 e^{-E_2/(RT)} \xi_2$:

$$\begin{aligned} &v_1 = 1/T, && v_2 = (E_1/R)v_1, && v_3 = \xi_1 e^{-v_2}, && v_1' = 1/T, \\ &v_4 = (E_2/R)v_1', && v_5 = \xi_2 e^{-v_4}, && v_6 = A_1 v_3 - A_2 v_5 = f_2(T, \xi_1, \xi_2). \end{aligned} \tag{5.4.28}$$

Observe that there are repeated factors v_1 and v_1'; we will handle these later. Let $(v_2^{\mathrm{L}}, v_2^{\mathrm{U}})$ and $(v_4^{\mathrm{L}}, v_4^{\mathrm{U}})$ respectively denote the natural interval bounds of v_2 and v_4 in (5.4.28) derived from $(T^{\mathrm{L}}, T^{\mathrm{U}})$. Bound the convex intrinsic functions $T \mapsto 1/T$ from above by the secant line connecting $(T^{\mathrm{L}}, 1/T^{\mathrm{L}})$ and $(T^{\mathrm{U}}, 1/T^{\mathrm{U}})$, and thus define $v_1^{\mathrm{cv}} \equiv v_1^{\mathrm{cv}'} := 1/T$ and $v_1^{\mathrm{cc}} \equiv v_1^{\mathrm{cc}'} := \frac{1}{T^{\mathrm{L}}}\left(-\frac{T-T^{\mathrm{L}}}{T^{\mathrm{U}}} + 1\right)$. We employed $h^{\mathrm{cv,env}}$ in (5.3.2) as the convex envelope of $(\xi_1, v_1) \mapsto$

$\xi_1 e^{-v_1}$ or $(\xi_2, v_3) \mapsto \xi_2 e^{-v_3}$, and employed the corresponding closed-form McCormick concave relaxation $h^{\mathrm{cc,MC}}$ as in (5.3.5). Thus, $\bar{u}_2$ defined in (5.4.10) may be constructed as follows:

$$
\begin{aligned}
\bar{u}_2(t,T,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \min_{\mathbf{v},\xi_1} \quad & A_1 v_3 - A_2 v_5 \\
\text{s.t.} \quad & v_3 \leq h^{\mathrm{cc,MC}}(v_2,\xi_1,v_2^{\mathrm{L}},x_1^{\mathrm{L}}(t),v_2^{\mathrm{U}},x_1^{\mathrm{U}}(t)), \\
& v_3 \geq h^{\mathrm{cv,env}}(v_2,\xi_1,v_2^{\mathrm{L}},x_1^{\mathrm{L}}(t),v_2^{\mathrm{U}},x_1^{\mathrm{U}}(t)), \\
& v_5 \leq h^{\mathrm{cc,MC}}(v_4,\xi_2^{\mathrm{cv}},v_4^{\mathrm{L}},x_2^{\mathrm{L}}(t),v_4^{\mathrm{U}},x_2^{\mathrm{U}}(t)), \\
& v_5 \geq h^{\mathrm{cv,env}}(v_4,\xi_2^{\mathrm{cv}},v_4^{\mathrm{L}},x_2^{\mathrm{L}}(t),v_4^{\mathrm{U}},x_2^{\mathrm{U}}(t)), \\
& v_2 = (E_1/R)v_1 \\
& v_4 = (E_2/R)v_1' \\
& v_1^{\mathrm{cv}} \leq v_1 \leq v_1^{\mathrm{cc}}, \\
& v_1^{\mathrm{cv}'} \leq v_1' \leq v_1^{\mathrm{cc}'}, \\
& \xi_1^{\mathrm{cv}} \leq \xi_1 \leq \xi_1^{\mathrm{cc}}.
\end{aligned}
\tag{5.4.29}
$$

By investigating the monotonicity of $h^{\mathrm{cv,env}}$ and $h^{\mathrm{cv,MC}}$, it was verified that $\bar{u}_2$ above is equivalent to the following closed form:

$$\bar{u}_2(t,T,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv A_1 h^{\mathrm{cv,env}}(v_2^{\mathrm{cc}},\xi_1^{\mathrm{cv}},v_2^{\mathrm{L}},x_1^{\mathrm{L}}(t),v_2^{\mathrm{U}},x_1^{\mathrm{U}}(t)) - A_2 h^{\mathrm{cc,MC}}(v_4^{\mathrm{cv}},\xi_2^{\mathrm{cv}},v_4^{\mathrm{L}},x_2^{\mathrm{L}}(t),v_4^{\mathrm{U}},x_2^{\mathrm{U}}(t)).$$

Similarly, $\bar{o}_2$ defined in (5.4.10) has the following closed form:

$$\bar{o}_2(t,T,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv A_1 h^{\mathrm{cc,MC}}(v_2^{\mathrm{cv}},\xi_1^{\mathrm{cc}},v_2^{\mathrm{L}},x_1^{\mathrm{L}}(t),v_2^{\mathrm{U}},x_1^{\mathrm{U}}(t)) - A_2 h^{\mathrm{cv,env}}(v_4^{\mathrm{cc}},\xi_2^{\mathrm{cc}},v_4^{\mathrm{L}},x_2^{\mathrm{L}}(t),v_4^{\mathrm{U}},x_2^{\mathrm{U}}(t)).$$

Note that this is the case where closed-form solutions of the optimization problems in

(5.4.10) can be easily identified, similarly to the OB relaxations as demonstrated in Example 5.1. Consider the $(v_1^{\mathrm{cv}}, v_1^{\mathrm{cc}}, v_2^{\mathrm{L}}, v_2^{\mathrm{U}}, v_4^{\mathrm{L}}, v_4^{\mathrm{U}})$ defined above, and following the same procedure, we have derived closed-form $(\bar{u}_1, \bar{o}_1, \bar{u}_3, \bar{o}_3)$ as follows:

$$\begin{aligned}
&\bar{u}_1(t,T,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv -A_1 h^{\mathrm{cc,MC}}(\frac{E_1}{R}v_1^{\mathrm{cv}}, \xi_1^{\mathrm{cv}}, v_2^{\mathrm{L}}, x_1^{\mathrm{L}}(t), v_2^{\mathrm{U}}, x_1^{\mathrm{U}}(t)),\\
&\bar{o}_1(t,T,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv -A_1 h^{\mathrm{cv,env}}(\frac{E_1}{R}v_1^{\mathrm{cc}}, \xi_1^{\mathrm{cc}}, v_2^{\mathrm{L}}, x_1^{\mathrm{L}}(t), v_2^{\mathrm{U}}, x_1^{\mathrm{U}}(t)),\\
&\bar{u}_3(t,T,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv A_2 h^{\mathrm{cv,env}}(\frac{E_2}{R}v_1^{\mathrm{cc}}, \xi_2^{\mathrm{cv}}, v_4^{\mathrm{L}}, x_2^{\mathrm{L}}(t), v_4^{\mathrm{U}}, x_2^{\mathrm{U}}(t)),\\
\text{and}\quad &\bar{o}_3(t,T,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv A_2 h^{\mathrm{cc,MC}}(\frac{E_2}{R}v_1^{\mathrm{cv}}, \xi_2^{\mathrm{cc}}, v_4^{\mathrm{L}}, x_2^{\mathrm{L}}(t), v_4^{\mathrm{U}}, x_2^{\mathrm{U}}(t)).
\end{aligned} \tag{5.4.30}$$

However, observe that the factor representation (5.4.28) has repeated factors v_1 and v_1'. By eliminating the repeated v_1', we constructed different $(\bar{u}_2^{\mathrm{NLP}}, \bar{o}_2^{\mathrm{NLP}})$ that are also defined by (5.4.10). Take $\bar{u}_2^{\mathrm{NLP}}$ as an example:

$$\begin{aligned}
\bar{u}_2^{\mathrm{NLP}}(t,T,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \min_{\mathbf{v},\xi_1} \quad & A_1 v_3 - A_2 v_5\\
\text{s.t.} \quad & v_3 \le h^{\mathrm{cc,MC}}(v_2, \xi_1, v_2^{\mathrm{L}}, x_1^{\mathrm{L}}(t), v_2^{\mathrm{U}}, x_1^{\mathrm{U}}(t)), && (5.4.31a)\\
& v_3 \ge h^{\mathrm{cv,env}}(v_2, \xi_1, v_2^{\mathrm{L}}, x_1^{\mathrm{L}}(t), v_2^{\mathrm{U}}, x_1^{\mathrm{U}}(t)),\\
& v_5 \le h^{\mathrm{cc,MC}}(v_4, \xi_2^{\mathrm{cv}}, v_4^{\mathrm{L}}, x_2^{\mathrm{L}}(t), v_4^{\mathrm{U}}, x_2^{\mathrm{U}}(t)), && (5.4.31b)\\
& v_5 \ge h^{\mathrm{cv,env}}(v_4, \xi_2^{\mathrm{cv}}, v_4^{\mathrm{L}}, x_2^{\mathrm{L}}(t), v_4^{\mathrm{U}}, x_2^{\mathrm{U}}(t)),\\
& v_2 = (E_1/R)v_1,\\
& v_4 = (E_2/R)v_1,\\
& v_1^{\mathrm{cv}} \le v_1 \le v_1^{\mathrm{cc}},\\
& \xi_1^{\mathrm{cv}} \le \xi_1 \le \xi_1^{\mathrm{cc}}.
\end{aligned}$$

The counterpart $\bar{o}_2^{\mathrm{NLP}}$ was constructed similarly. Unlike the $\bar{u}_2$ defined in (5.4.29), $\bar{u}_2^{\mathrm{NLP}}$

above does not have a straightforward closed-form optimal objective value, and may be evaluated using numerical NLP solvers. Observe from (5.3.5) that $h^{\mathrm{cc,MC}}$ is equivalent to the following:

$$h^{\mathrm{cc,MC}}(\boldsymbol{\xi},\boldsymbol{\xi}^{\mathrm{L}},\boldsymbol{\xi}^{\mathrm{U}}) \equiv \min\{h^{\mathrm{cc,MC1}}(\boldsymbol{\xi},\boldsymbol{\xi}^{\mathrm{L}},\boldsymbol{\xi}^{\mathrm{U}}), h^{\mathrm{cc,MC2}}(\boldsymbol{\xi},\boldsymbol{\xi}^{\mathrm{L}},\boldsymbol{\xi}^{\mathrm{U}})\},$$

where

$$\begin{aligned} &h^{\mathrm{cc,MC1}} : (\boldsymbol{\xi},\boldsymbol{\xi}^{\mathrm{L}},\boldsymbol{\xi}^{\mathrm{U}}) \mapsto e^{-\xi_1^{\mathrm{U}}}\xi_2 + \xi_2^{\mathrm{U}}\beta^h - \xi_2^{\mathrm{U}}e^{-\xi_1^{\mathrm{U}}} \\ \text{and}\quad &h^{\mathrm{cc,MC2}} : (\boldsymbol{\xi},\boldsymbol{\xi}^{\mathrm{L}},\boldsymbol{\xi}^{\mathrm{U}}) \mapsto e^{-\xi_1^{\mathrm{L}}}\xi_2 + \xi_2^{\mathrm{L}}\beta^h - \xi_2^{\mathrm{L}}e^{-\xi_1^{\mathrm{L}}}. \end{aligned}$$

Observe that $h^{\mathrm{cc,MC}}$ is nonsmooth, while $h^{\mathrm{cc,MC1}}$ and $h^{\mathrm{cc,MC2}}$ are smooth. As discussed in Section 5.4.3, the constraints (5.4.31a) and (5.4.31b) may thus be replaced by the following:

$$\begin{aligned} v_3 &\leq h^{\mathrm{cc,MC1}}(v_2,\xi_1,v_2^{\mathrm{L}},x_1^{\mathrm{L}}(t),v_2^{\mathrm{U}},x_1^{\mathrm{U}}(t)), \\ v_3 &\leq h^{\mathrm{cc,MC2}}(v_2,\xi_1,v_2^{\mathrm{L}},x_1^{\mathrm{L}}(t),v_2^{\mathrm{U}},x_1^{\mathrm{U}}(t)), \\ v_5 &\leq h^{\mathrm{cc,MC1}}(v_4,\xi_2^{\mathrm{cv}},v_4^{\mathrm{L}},x_2^{\mathrm{L}}(t),v_4^{\mathrm{U}},x_2^{\mathrm{U}}(t)), \\ v_5 &\leq h^{\mathrm{cc,MC2}}(v_4,\xi_2^{\mathrm{cv}},v_4^{\mathrm{L}},x_2^{\mathrm{L}}(t),v_4^{\mathrm{U}},x_2^{\mathrm{U}}(t)). \end{aligned}$$

This allows to evaluate $\bar{u}_2^{\mathrm{NLP}}$ with off-the-shelf smooth NLP solvers. The same reformulation was also applied to $\bar{o}_2^{\mathrm{NLP}}$.

We have also constructed a third class of the AVM-based right-hand side functions for f_2; denote these as $(\bar{u}_2^{\mathrm{LP}},\bar{o}_2^{\mathrm{LP}})$. These $(\bar{u}_2^{\mathrm{LP}},\bar{o}_2^{\mathrm{LP}})$ were constructed similarly to $(\bar{u}_2^{\mathrm{NLP}},\bar{o}_2^{\mathrm{NLP}})$, but we replaced the employed nonlinear $(h^{\mathrm{cv,env}},h^{\mathrm{cc,MC1}},h^{\mathrm{cc,MC2}})$ by their subtangents at center of the considered box domains (e.g. a subtangent of $h^{\mathrm{cv,env}}$ at the center of $[v_2^{\mathrm{L}},v_2^{\mathrm{U}}]\times[x_1^{\mathrm{L}}(t),x_1^{\mathrm{U}}(t)]$). Thus, evaluating $(\bar{u}_2^{\mathrm{LP}},\bar{o}_2^{\mathrm{LP}})$ is intuitively more efficient than evaluating $(\bar{u}_2^{\mathrm{NLP}},\bar{o}_2^{\mathrm{NLP}})$, since evaluating $(\bar{u}_2^{\mathrm{LP}},\bar{o}_2^{\mathrm{LP}})$ involves solving LPs while $(\bar{u}_2^{\mathrm{NLP}},\bar{o}_2^{\mathrm{NLP}})$

solve NLPs.

Based on the different constructions of the AVM-based relaxation system's right-hand sides above, Table 5.5 summarizes all state relaxation methods that were applied to the original system (5.4.27) using our proof-of-concept implementation in Julia. Figures 5.4 and 5.5 respectively depict the original parametric solutions $x_2(0.02,T)$ and $x_3(0.02,T)$, along with the resulting state relaxations for comparison. Table 5.6 summarizes the per-mesh-point CPU times for evaluating these state relaxations. From both figures, the OBM relaxations are identical to the SBM relaxations for this example, while the former took longer CPU time than the latter. This is because the SBM method evaluates closed-form right-hand side functions, while our implementation for the OBM relaxations automatically constructs and solves NLPs using numerical solvers at system's right-hand side. It may be possible to derive closed-form right-hand sides for the OBM relaxation system similarly to Example 5.1, which, however, is not the focus of this example. The CF-AVM-based relaxations are as efficient as the SBM relaxations, yet are significantly tighter due to the use of multivariate convex envelopes, which is consistent with Results 1 and 2 in Remark 5.4.24. The NLP-AVM-based method yields the tightest relaxations by eliminating repeated factors and employing convex envelopes, yet requires the longest CPU time. The reason why the NLP-AVM-based relaxations took longer CPU time to evaluate than the OBM relaxations may be that the NLP-AVM-based right-hand side solved convex NLPs with nonlinear convex constraints, while the OBM right-hand side solved convex NLPs with simple box constraints. The LP-AVM-based relaxations are much more efficient than the NLP-AVM-based relaxations, since LPs are more efficient to solve than convex NLPs at ODE right-hand sides. On the other hand, the LP-AVM-based relaxations are no tighter than the NLP-AVM-based relaxations as guaranteed by Result 3 in Remark 5.4.24, since subtangents are outer approximations of the original convex relaxations. Note that there is no general tightness result for comparing the SBM relaxations and the LP-AVM-based relaxations. However, this example indicates that the LP-AVM-based relaxations may still be promising, since they are much more efficient than the NLP-AVM-based relaxations, yet may have extrema that are close to the extrema of the NLP-AVM-based relaxations, as can be seen from the figures. In addition, using multiple subtangents [44] to tighten LP-AVM-based relaxations is also possible. We remark that both implementations for the LP/NLP-AVM-based relaxations may be improved using the techniques outlined in Section 5.4.3.

Table 5.5: Various state relaxation methods that were applied to (5.4.27), for Example 5.5.

Methods	Description
CF-AVM-based method*	New AVM-based method evaluating $(\bar{u}_2, \bar{o}_2)$ in closed form
NLP-AVM-based method*	New AVM-based method evaluating $(\bar{u}_2^{\mathrm{NLP}}, \bar{o}_2^{\mathrm{NLP}})$ with NLP solver IPOPT
LP-AVM-based method*	New AVM-based method evaluating $(\bar{u}_2^{\mathrm{LP}}, \bar{o}_2^{\mathrm{LP}})$ with LP solver Clp
OBM method	From [3], using generalized McCormick relaxations of $\mathbf{f}$
SBM method	From [2], using McCormick relaxations of $\mathbf{f}$

* All these methods evaluate the same closed-form $(\bar{u}_1, \bar{o}_1, \bar{u}_3, \bar{o}_3)$ in (5.4.30).

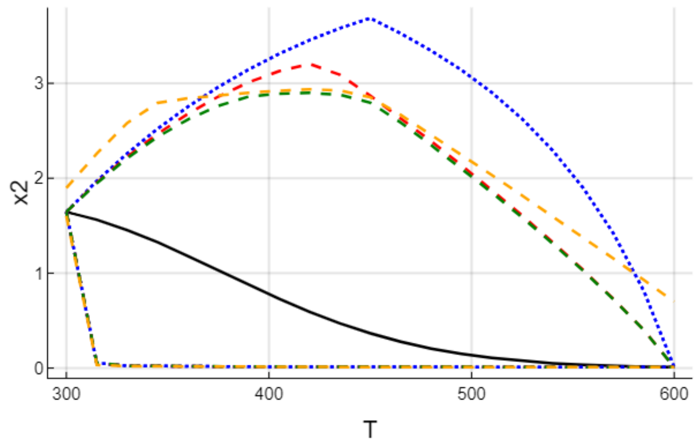

Figure 5.4: The solution $x_2(0.02, T)$ (*solid black*) of the parametric ODE (5.4.27) from Example 5.5, plotted against T along with corresponding SBM relaxations [2] (*dotted blue*), CF-AVM-based relaxations (*dashed red*), NLP-AVM-based relaxations (*dashed green*), and LP-AVM-based relaxations (*dashed orange*). Note that the OBM relaxations [3] are identical to the SBM relaxations in this case, and all state convex relaxations overlap.

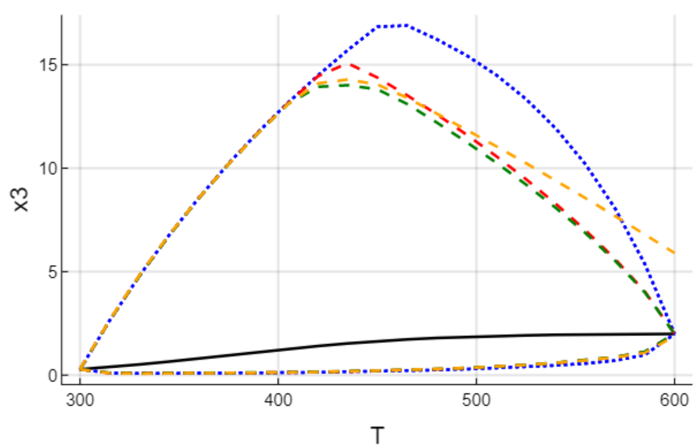

Figure 5.5: The solution $x_3(0.02, T)$ (*solid black*) of the parametric ODE (5.4.27) from Example 5.5, plotted against T along with corresponding SBM relaxations [2] (*dotted blue*), CF-AVM-based relaxations (*dashed red*), NLP-AVM-based relaxations (*dashed green*), and LP-AVM-based relaxations (*dashed orange*). Note that the OBM relaxations [3] are identical to the SBM relaxations in this case.

Table 5.6: Average computational times for evaluating state relaxations $(\mathbf{x}^{\mathrm{cv}}(0.02,T),\mathbf{x}^{\mathrm{cc}}(0.02,T))$ for (5.4.25) in Example 5.5.

State relaxation method	CPU time (seconds)*
CF-AVM-based relaxations	0.07
NLP-AVM-based relaxations	7.15 ± 0.32
LP-AVM-based relaxations	1.75
OBM relaxations	5.23 ± 0.12
SBM relaxations	0.08

* Each CPU time here was averaged over 10 runs. For each CPU time with a significant sample standard deviation (std), the number is reported as "average $\pm$ std".

5.5 Conclusions and future work

This chapter has proposed two extensions of the established OB relaxations [3] introduced in Chapter 4 for solutions of the parametric ODE system (2.3.1), which are useful for efficiently computing tight bounding information in deterministic algorithms of global dynamic optimization. While the OB relaxations may be significantly tighter than the SBM relaxations [2], the previous implementation [3] of the OB relaxations solves convex NLPs using numerical solvers at each time step, which requires expensive computational efforts. Therefore, in the first extension, we considered relaxations $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$ in (5.2.2) with pre-known monotonicity, and showed that by directly identifying closed-form extrema of these relaxations on the box X, we can derive closed-form right-hand side functions of the OB relaxation system. Example 5.1 showed that by using this method, the OB relaxations can be tighter, yet as efficient as the SBM relaxations [2], which may ultimately improve efficiency of an overarching global optimization method. The second extension is a new type of state relaxations termed AVM-based state relaxations, which are computed by solving (2.4.1) with $(\mathbf{u},\mathbf{o})$ satisfying Assumption 5.4.11. This relaxation approach considers a factorable (in the sense of Definition 5.4.2) original right-hand side function **f**, and is motivated by embedding the AVM relaxations as in Definition 5.4.6 of **f** into the OB relaxation system, as discussed in Section 5.4.3. Theorem 5.4.23 and Remark 5.4.24 showed that the new AVM-based relaxations are guaranteed to be at least as tight as both the SBM relaxations and the OBM relaxations. A proof-of-concept implementation of the AVM-based relaxations in Julia was outlined in Section 5.4.5. Section 5.4.6 presented several numerical examples to show that the new AVM-based relaxation approach can effectively handle repeated factors as in Definition 5.4.26 and employ convex envelopes of multivariate intrinsic functions of **f**, to yield significantly tighter state relaxations than the SBM and OBM relaxations. Lastly, Example 5.5 is concerned with the application to a chemical reaction network model, and discussed different types of AVM-based relaxations by constructing closed-form, LP-based, and NLP-based right-hand side functions.

Future work will involve developing more efficient implementation of the AVM-based relaxations, using the techniques outlined in Section 5.4.3. Besides, subgradients of state relaxations with respect to **p** may be developed, to aid local NLP solvers in obtaining lower bounds for the globally optimal objective values in branch-and-bound. The new state relaxations may then be applied to deterministic algorithms of global dynamic optimization.

Chapter 6

Constructing Subgradients for Convex Relaxations of Nonconvex Parametric Systems of ODEs

This chapter, reproduced from the submitted journal article [92], proposes novel subgradient evaluation methods for state relaxations obtained using the state-of-the-art Scott–Barton ODE relaxation framework, for solving lower-bounding problems in the algorithms of deterministic global dynamic optimization. The subgradients are computed as the unique solution of an auxiliary parametric affine ODE system, analogously to classical forward sensitivity evaluation methods for smooth dynamic systems. Unlike established approaches that propagate valid subgradients for nonsmooth dynamic systems, this new method is compatible with existing subgradient evaluation methods for convex functions, and thus allows using well-developed subgradient libraries such as EAGO and MC++ for implementation. Moreover, we show that a subgradient of the objective function of a lower-bounding problem in global dynamic optimization can be directly evaluated using adjoint sensitivity analysis, which may reduce the overall computational effort for an overarching gloal optimization method. Numerical examples are presented to illustrate our new forward subgradient evaluation methods, based on proof-of-concept implementations in Julia. Further implications are discussed.

6.1 Introduction

This chapter focuses on computing *subgradients of state relaxations* for the original parametric ODE system (1.1.2) formalized in Section 2.3. As discussed in Section 1.1, state relaxations (formalized in Definition 2.4.1) are valid underestimators and overestimators of $\mathbf{x}$ in (1.1.2), whose components are respectively convex and concave with respect to $\mathbf{p}$ for each fixed t. Such relaxations are useful in deterministic global optimization algorithms [27–31], for computing lower bounds of the globally optimal objective values of a nonconvex dynamic optimization problem with (1.1.2) embedded. Subgradients provide useful local sensitivity information for convex functions (analogously for concave functions), which are typically required by convex optimization methods such as *level method* and *subgradient method* proposed by Nesterov [40] and general nonsmooth local optimization methods such as *bundle methods* [41–43]. Without subgradients, an overarching global optimization method may fail to compute correct lower bounds when attempting to minimize convex relaxations, which may ultimately lead the method to either report an incorrect global minimum or incorrectly conclude that the problem is infeasible. Moreover, subgradients are useful for constructing piecewise affine relaxations by a finite combination of the corresponding subtangents [34, 44, 45], and each subtangent can be efficiently constructed by a single evaluation of the original convex relaxation and an associated subgradient [46].

Subgradients of state relaxations may also be used for constructing a convex polyhedral enclosure of the reachable set for $\mathbf{x}$ in (1.1.2) [47].

As summarized in Section 2.4, Scott and Barton [2] have developed a general ODE relaxation framework, which computes nonsmooth state relaxations for (1.1.2) by constructing and solving an auxiliary parametric ODE system. This framework requires furnishing relaxations for the initial-value function $\mathbf{x}_0$, *state bounds* (c.f. [2, Definition 2]) for $\mathbf{x}$ that are independent of $\mathbf{p}$, and crucial right-hand side (RHS) functions based on relaxations of $\mathbf{f}$ in (1.1.2). There are thus far two established state relaxation methods [2, 3] in this framework, which describe relaxations as solutions of auxiliary ODEs with different RHS functions. Scott and Barton [2] construct the RHS functions by composing the generalized McCormick relaxations [2] of $\mathbf{f}$ with interval flattening operations; the resulting state relaxations will be referred as *Scott–Barton–McCormick (SBM) relaxations*. Our previous work [3] constructs different RHS functions as optimal-value functions (in the sense of e.g. [114]) with embedded convex optimization problems, which may employ any convex and concave relaxations of $\mathbf{f}$. These state relaxations will be called *optimization-based (OB) relaxations* in this article. While these state relaxations [2, 3] have been shown to exhibit desirable tightness and convergence properties, there are thus far no established methods to evaluate subgradients for these relaxations when nonsmoothness is encountered, which limits the use of these relaxations in global dynamic optimization. Song et al. [4] propose an approach for tractably constructing closed-form affine relaxations via black-box evaluations of an original convex relaxation, without knowing subgradients. However, these affine relaxations are likely to be more conservative than subtangents derived from subgradients, which may in turn provide overly conservative bounds in global optimization. Hence, there is a need to develop new dynamic subgradient evaluation methods for these state relaxations [2, 3].

In general, established methods for evaluating subgradients for convex parametric ODE solutions are somewhat limited. [46, Theorem 3.2] shows that if the ODE RHS is convex, then subgradients of the solution uniquely solve an auxiliary ODE system, whose RHS requires subgradients of the original ODE RHS. This result is applicable to an earlier weaker ODE relaxation method in [75], but not the superior Scott–Barton ODE relaxation framework [2], since the latter framework's RHS is not convex. [46, Theorem 3.3] does not require a convex ODE RHS, but instead differentiability is required throughout. Recently, Yuan and Khan [130] show that for bivariate convex functions, centered finite differences will always converge to a subgradient, and a subgradient may be efficiently evaluated from directional derivatives. However, these results are not extendable to higher dimensions.

When convexity is not exploited, classical forward sensitivity [131, 132] or adjoint sensitivity [90] evaluation methods for smooth dynamic systems are not applicable here, since the mentioned state relaxations are nonsmooth in general. In the field of nonsmooth dynamic sensitivity analysis, several approaches [133–135] extend the smooth sensitivity evaluation methods to nonsmooth dynamic systems defined by smooth functions and discrete events, assuming that the systems are well-behaved around discrete events. [136, Theorem 7.4.1] proposes sufficient conditions under which a parametric ODE system with non-differentiable RHS has differentiable parametric solutions. Pang and Stewart [137] compute *linear Newton approximations* [138] of the solutions of parametric ODEs with RHS that are *semismooth* [139]. Yunt [140] then extends their results to compute linear Newton approximations using adjoints. However, as illustrated in [97], the linear Newton approximations of a convex function may contain elements that are not subgradients, and do not have associated sufficient optimality conditions unlike the gradient. Recently, methods [97, 141, 142] have been proposed to compute *lexicographic derivatives* [143] of the solutions of nonsmooth parametric ODE systems, differential-algebraic equation (DAE) systems, and hybrid systems. These lexicographic derivatives reduce to valid subgradients in a convex case. However, these methods involve solving auxiliary dynamic systems whose RHS functions are generally discontinuous in state variables, and thus require tailored numerical solvers.

In this chapter, we propose a new general dynamic subgradient evaluation framework for any state relaxations obtained using the Scott–Barton ODE relaxation framework [2]. This subgradient framework assumes that the underlying state relaxations do not touch the predefined state bounds during integration, which is guaranteed to be satisfied for a sufficiently small domain of the parameters $\mathbf{p}$, since state relaxations have been shown to converge to the original trajectory $\mathbf{x}$ faster than state bounds as $\mathbf{p}$'s domain shrinks [79]. This framework also requires the convex state relaxations to never overlap with the concave relaxations; this can be easily guaranteed by adding an arbitrarily small perturbation to the original state relaxation system, as will be seen in Section 6.4.4. For each fixed $\mathbf{p}$, we construct subgradients of state relaxations as the unique solution of a forward auxiliary ODE system, to be integrated simultaneously with the state relaxation ODE system. The RHS of this forward sensitivity system is constructed by combining subgradients of relaxations of $\mathbf{f}$ with new flattening operations that play an analogous role to the interval flattening operations at Scott–Barton framework's RHS. While plausible, this result is previously unknown and requires sophisticated justification. Unlike the established dynamic subgradient evaluation method [46, Theorem 3.2], our new method does not require the relaxation system's RHS to be convex, and thus is applicable to Scott–Barton framework.

Unlike the method [97] that requires computing a specific lexicographic derivative of the relaxation system's RHS for each time t, this new method permits any subgradients of relaxations of $\mathbf{f}$ to be used, and thus is compatible with existing subgradient evaluation methods [7,33,35,144,145] for convex functions. Subgradients of McCormick relaxations for example, are readily available from well-developed libraries such as EAGO [94] and MC++ [35,96]. Moreover, we show that if the subgradients of relaxations of $\mathbf{f}$ follow classical chain rule as for gradients, then the resulting forward subgradient propagation system is in fact an affine parametric ODE system (analogously to the forward sensitivity system for smooth ODE systems), which may be easily integrated using off-the-shelf numerical ODE solvers. Based on this, we furthermore show that a subgradient of nonsmooth convex relaxations of objective functions in nonconvex dynamic optimization problems may be efficiently evaluated using adjoints, analogously to classical dynamic adjoint sensitivity analysis [90]. Ruban [135] and Hanneman-Tamás et al. [134] have proposed adjoint sensitivity evaluation methods for nonsmooth dynamic systems, which are under certain regularity assumptions. For example, these methods consider nonsmooth dynamic systems defined using smooth functions and discrete events, and the state variables are assumed to satisfy certain transition conditions around discrete events. Under these conditions, the resulting generalized states are actually continuously differentiable with respect to uncertain parameters. On the other hand, our new adjoint subgradient evaluation method considers an ODE system with nonsmooth RHS but without discrete events; no additional regularity assumptions are required, and yields valid subgradients for nonsmooth convex parametric ODE solutions. Adjoint sensitivity evaluation methods [90] have been shown to be empirically more efficient than forward methods in smooth dynamic optimization for a large number of decision variables. We expect that this extension to nonsmooth subgradient evaluation would speed up computation of subgradients for solving the lower-bounding problems, and ultimately reduce computational effort for deterministic global dynamic optimization.

Based on the general subgradient evaluation framework presented above, we then propose new numerical methods for evaluating subgradients for the SBM relaxations [2] and our newer OB relaxations [3]. For the SBM relaxations, we propose to construct a forward subgradient propagation system based on Mitsos et al.'s vector forward mode subgradient automatic differentiation for McCormick relaxations [35]. Correspondingly, we propose a new adjoint ODE subgradient evaluation system constructed from Beckers et al.'s adjoint mode computation for subgradients of McCormick relaxations, for use in lower bounding in global dynamic optimization. For the OB relaxations, we propose a new forward subgradient propagation system based on a recently established subgradient evaluation method [145, Theorem 5.3.2] for *multivariate McCormick relaxations* [7, Theorem 2]. While the OB relaxation ODE system solves convex optimization problems at RHS, the new forward system's RHS is efficiently evaluated in closed form; no need to solve optimization problems. We have also developed proof-of-concept implementations in Julia [95] for forward subgradient evaluation for both the SBM relaxations and the OB relaxations. Numerical examples show that these new methods indeed yield valid subgradients. We leave implementation for dynamic adjoint subgradient evaluation for future work, since there is no off-the-shelf implementation of adjoint subgradient evaluation for McCormick relaxations and no appropriate adjoint ODE sensitivity solver in Julia. The adjoint solver provided by the package DifferentialEquations [123] appears to be incomplete, and we were unable to adapt it for our systems.

The remainder of this chapter is organized as follows. Section 6.2 summarizes relevant mathematical background including subgradients and directional derivatives of parametric

ODE solutions. The Scott–Barton ODE relaxation framework [2] was already summarized in Section 2.4. Section 6.3 formalizes the problem formulation for this article. Section 6.4 presents a new general subgradient propagation framework, where a forward subgradient evaluation system and an adjoint sensitivity system are constructed. Section 6.5 then proposes new numerical methods for evaluating subgradients of the SBM relaxations [2] and the OB relaxations [3]. Lastly, Section 6.6 presents numerical examples to illustrate the correctness of the proposed forward sensitivity methods, based on a proof-of-concept implementation in Julia.

6.2 Mathematical preliminaries

This section summarizes the mathematical preliminaries of this article. Section 6.2.1 presents the standard definitions of directional derivatives and subgradients in convex analysis [146]. Section 6.2.2 presents an established method [97] for propagating directional derivatives of parametric ODE solutions, which is essential for validating the new subgradient evaluation methods of this article.

6.2.1 Directional derivatives and subgradients

Definition 6.2.1 (from [127])**.** Given an open set $Y \subset \mathbb{R}^n$, a function $\mathbf{h} : Y \to \mathbb{R}^m$, some $\mathbf{y} \in Y$, and some $\mathbf{d} \in \mathbb{R}^n$, the limit

$$\lim_{\alpha \downarrow 0} \frac{\mathbf{h}(\mathbf{y} + \alpha\mathbf{d}) - \mathbf{h}(\mathbf{y})}{\alpha}$$

is called the *directional derivative* of $\mathbf{h}$ at $\mathbf{y}$ if it exists, and is denoted as $\mathbf{h}'(\mathbf{y};\mathbf{d})$. The function $\mathbf{h}$ is *directionally differentiable* at $\mathbf{y}$ if $\mathbf{h}'(\mathbf{y};\mathbf{d})$ exists and is finite for each $\mathbf{d} \in \mathbb{R}^n$.

The function $\mathbf{h}$ is said to be directionally differentiable if it is directionally differentiable at every $\mathbf{y} \in Y$.

The following definitions of *subgradients* and *subdifferentials* are standard in convex analysis [146].

Definition 6.2.2 (adapted from [146])**.** Given an open convex set $Y \subset \mathbb{R}^n$ and a convex function $h^{\mathrm{cv}} : Y \to \mathbb{R}$, $\mathbf{s}^{\mathrm{cv}} \in \mathbb{R}^n$ is a *subgradient* of h^{cv} at $\mathbf{y} \in Y$ if

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

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

$$h^{\mathrm{cc}}(\boldsymbol{\eta}) \leq h^{\mathrm{cc}}(\mathbf{y}) + \langle \mathbf{s}^{\mathrm{cc}}, \boldsymbol{\eta} - \mathbf{y} \rangle, \quad \forall \boldsymbol{\eta} \in Y.$$

The *subdifferential* $\partial h^{\mathrm{cv}}(\mathbf{y}) \subset \mathbb{R}^n$ (resp. $\partial h^{\mathrm{cc}}(\mathbf{y}) \subset \mathbb{R}^n$) is the collection of all subgradients of h^{cv} (resp. h^{cc}) at $\mathbf{y}$.

Theorem 6.2.3 (adapted from [146])**.** Given functions h^{cv} and h^{cc} in Definition 6.2.2, $\mathbf{s}^{\mathrm{cv}} \in \mathbb{R}^n$ is a subgradient of h^{cv} at $\mathbf{y} \in Y$ if

$$[h^{\mathrm{cv}}]'(\mathbf{y}; \mathbf{d}) \geq \langle \mathbf{s}^{\mathrm{cv}}, \mathbf{d} \rangle, \quad \forall \mathbf{d} \in \mathbb{R}^n.$$

Similarly, $\mathbf{s}^{\mathrm{cc}} \in \mathbb{R}^n$ is a subgradient of h^{cc} at $\mathbf{y} \in Y$ if

$$[h^{\mathrm{cc}}]'(\mathbf{y}; \mathbf{d}) \leq \langle \mathbf{s}^{\mathrm{cc}}, \mathbf{d} \rangle, \quad \forall \mathbf{d} \in \mathbb{R}^n.$$

6.2.2 Directional derivatives of parametric ODE solutions

Pang and Stewart [137] showed that the directional derivatives of an ODE solution with respect to its initial condition uniquely solve an auxiliary ODE system. Khan and Barton [97] extended this result to parametric ODE systems whose right-hand side functions are discontinuous in t. The following theorem is adapted from [97, Theorem 4.1].

Theorem 6.2.4. Let $I := [t_0, t_f]$ where $t_0 < t_f$, and let $\tilde{D} \subset \mathbb{R}^n$ and $\tilde{P} \subset \mathbb{R}^m$ be open. Consider a Lipschitz continuous and directionally differentiable function $\mathbf{y}_0 : \tilde{P} \to \tilde{D}$. Suppose that a function $\mathbf{g} : I \times \tilde{P} \times \tilde{D} \to \mathbb{R}^n$ satisfies the following conditions:

1. the mapping $\mathbf{g}(\cdot, \mathbf{p}, \boldsymbol{\eta}) : I \to \mathbb{R}^m$ is measurable for each $\mathbf{p} \in \tilde{P}$ and $\boldsymbol{\eta} \in \tilde{D}$,

2. there exist Lebesgue integrable functions $k_{\mathbf{g}}, m_{\mathbf{g}} : I \to \mathbb{R}_+ \cup \{+\infty\}$ for which

$$\|\mathbf{g}(t, \mathbf{p}, \boldsymbol{\eta})\| \leq m_{\mathbf{g}}(t), \quad \forall t \in I,\ \forall \mathbf{p} \in \tilde{P},\ \forall \boldsymbol{\eta} \in \tilde{D},$$

   and

$$\|\mathbf{g}(t, \mathbf{p}^{\mathrm{A}}, \boldsymbol{\eta}^{\mathrm{A}}) - \mathbf{g}(t, \mathbf{p}^{\mathrm{B}}, \boldsymbol{\eta}^{\mathrm{B}})\| \leq k_{\mathbf{g}}(t)\Big(\|\mathbf{p}^{\mathrm{A}} - \mathbf{p}^{\mathrm{B}}\| + \|\boldsymbol{\eta}^{\mathrm{A}} - \boldsymbol{\eta}^{\mathrm{B}}\|\Big),$$
$$\forall t \in I,\ \forall \mathbf{p}^{\mathrm{A}}, \mathbf{p}^{\mathrm{B}} \in \tilde{P},\ \forall \boldsymbol{\eta}^{\mathrm{A}}, \boldsymbol{\eta}^{\mathrm{B}} \in \tilde{D},$$

3. for each $t \in I$ except in a zero-measure subset $Z_{\mathbf{g}}$, the mapping $\mathbf{g}(t, \cdot, \cdot) : \tilde{P} \times \tilde{D} \to \mathbb{R}^n$ is directionally differentiable.

For each $\mathbf{p} \in \tilde{P}$, with $\mathbf{y}(\cdot, \mathbf{p})$ denoting any solution of the parametric ODE system:

$$\dot{\mathbf{y}}(t, \mathbf{p}) = \mathbf{g}(t, \mathbf{p}, \mathbf{y}(t, \mathbf{p})), \quad \mathbf{y}(t_0, \mathbf{p}) = \mathbf{y}_0(\mathbf{p}), \tag{6.2.1}$$

suppose that there exists a solution $\{\mathbf{y}(t,\mathbf{p}) : t \in I\} \subset \tilde{D}$. Then, for each $t \in I$, the function $\mathbf{y}_t \equiv \mathbf{y}(t,\cdot)$ is well-defined and Lipschitz continuous on $\tilde{P}$ with a Lipschitz constant that is independent of t. Moreover, $\mathbf{y}_t$ is directionally differentiable for each $t \in I$, and for each $\mathbf{p} \in \tilde{P}$ and $\mathbf{d} \in \mathbb{R}^m$, the mapping $t \mapsto [\mathbf{y}_t]'(\mathbf{p};\mathbf{d})$ is the unique solution (in the Carathéodory sense [102]) on I of the ODE system:

$$\dot{\mathbf{z}}(t) = [\hat{\mathbf{g}}_t]'\big((\mathbf{p},\mathbf{y}(t,\mathbf{p}));(\mathbf{d},\mathbf{z}(t))\big), \quad \mathbf{z}(t_0) = [\mathbf{y}_0]'(\mathbf{p};\mathbf{d}), \tag{6.2.2}$$

where $\hat{\mathbf{g}}_t : \tilde{P} \times \tilde{D} \to \mathbb{R}^n$ is defined in terms of $\mathbf{g}$ as follows and is directionally differentiable for each $t \in I \backslash Z_{\mathbf{g}}$: for each $(t,\mathbf{p},\boldsymbol{\eta}) \in I \times \tilde{P} \times \tilde{D}$,

$$\hat{\mathbf{g}}_t(\mathbf{p},\boldsymbol{\eta}) := \begin{cases} \mathbf{g}(t,\mathbf{p},\boldsymbol{\eta}), & \text{if } t \in I\backslash Z_{\mathbf{g}}, \\ \mathbf{0}, & \text{if } t \in Z_{\mathbf{g}}. \end{cases}$$

6.3 Problem statement

Assumption 6.3.1. Suppose that initial relaxations $(\mathbf{x}_0^{\mathrm{cv}},\mathbf{x}_0^{\mathrm{cc}})$ for (2.3.1) satisfy the following: for each $\mathbf{p} \in P$, $\mathbf{x}^{\mathrm{L}}(t_0) \leq \mathbf{x}_0^{\mathrm{cv}}(\mathbf{p}) \leq \mathbf{x}_0^{\mathrm{cc}}(\mathbf{p}) \leq \mathbf{x}^{\mathrm{U}}(t_0)$.

The assumption above is for simplicity of analysis. Under this assumption, the initial conditions in the Scott–Barton framework (2.4.1) becomes $x_i^{\mathrm{cv}}(t_0,\mathbf{p}) = x_{0,i}^{\mathrm{cv}}(\mathbf{p})$ and $x_i^{\mathrm{cc}}(t_0,\mathbf{p}) = x_{0,i}^{\mathrm{cc}}(\mathbf{p})$.

Assumption 6.3.2. Consider the original parametric ODE system (2.3.1) formalized in Assumption 2.3.2. Suppose that appropriate state bounds $(\mathbf{x}^{\mathrm{L}},\mathbf{x}^{\mathrm{U}})$ in Assumption 2.4.5 and initial relaxations $(\mathbf{x}_0^{\mathrm{cv}},\mathbf{x}_0^{\mathrm{cc}})$ in Assumption 6.3.1 are available. Denote the interior of P as $\tilde{P}$. For any Scott–Barton right-hand side functions $(\mathbf{u},\mathbf{o})$, suppose that the mappings $\mathbf{u}(t,\cdot,\cdot,\cdot)$

and $\mathbf{o}(t,\cdot,\cdot,\cdot)$ are directionally differentiable and Lipschitz continuous on $\tilde{P} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly over $t \in I$. Consider state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ for (2.3.1) obtained via (2.4.1), and suppose that for all $t \in (t_0, t_f]$ and $\mathbf{p} \in \tilde{P}$,

$$\mathbf{x}^{\mathrm{L}}(t) < \mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}) < \mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}) < \mathbf{x}^{\mathrm{U}}(t). \tag{6.3.1}$$

Definition 6.3.3. For any state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ for (2.3.1), functions $\mathbf{S}^{\mathrm{cv}}, \mathbf{S}^{\mathrm{cc}} : I \times \tilde{P} \to \mathbb{R}^{n_x \times n_p}$ are called *state relaxation subgradients* if, for each $i \in \{1, ..., n_x\}$, $t \in I$, and $\mathbf{p} \in \tilde{P}$, $[\mathbf{s}^{\mathrm{cv}}_{(i)}(t,\mathbf{p})]^{\mathrm{T}}$ is a subgradient of $x^{\mathrm{cv}}_i(t,\cdot)$ at $\mathbf{p}$, and $[\mathbf{s}^{\mathrm{cc}}_{(i)}(t,\mathbf{p})]^{\mathrm{T}}$ is a subgradient of $x^{\mathrm{cc}}_i(t,\cdot)$ at $\mathbf{p}$.

As discussed in Section 6.1, state relaxation subgradients are useful in deterministic algorithms of global dynamic optimization [2, 12, 18, 108], for computing the required lower bounds by minimizing convex relaxations constructed from state relaxations. To the authors' knowledge, [97] is the only established method which may be used for computing subgradients of state relaxations obtained using the Scott–Barton framework (2.4.1) under Assumption 6.3.2. This method computes valid lexicographic derivatives [143] for nonsmooth parametric ODE solutions via an auxiliary ODE system. These lexicographic derivatives reduce to subgradients in a convex case. However, this method has the following disadvantages. Firstly, the auxiliary ODE system's RHS is in general discontinuous with respect to state variables, and thus tailored ODE solvers are required. Secondly, this method does not allow efficient dynamic sensitivity evaluation using adjoints. Thirdly, the auxiliary ODE system's RHS requires a specific lexicographic derivative of the nonsmooth ODE RHS, which is not compatible with established subgradient evaluation methods [7, 33, 35, 145]. This will increase the difficulty for implementation.

Thus, under Assumption 6.3.2, the goal of this article is to propose a new framework for constructing subgradients $(\mathbf{S}^{\mathrm{cv}}, \mathbf{S}^{\mathrm{cc}})$ for state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ for (2.3.1) obtained using the Scott–Barton framework (2.4.1), and also propose efficient subgradient evaluation methods for the established state relaxations [2, 3] in this framework. These new methods construct new auxiliary ODEs that are easily integrated with off-the-shelf numerical solvers, enable dynamic adjoint subgradient evaluation, and are compatible with established subgradient evaluation methods. Under Assumption 6.3.2, the ODE system (2.4.1) reduces to

$$\begin{aligned}
\dot{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p}) &= \mathbf{u}(t,\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})), \quad \mathbf{x}^{\mathrm{cv}}(t_0,\mathbf{p}) = \mathbf{x}_0^{\mathrm{cv}}(\mathbf{p}), \\
\dot{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p}) &= \mathbf{o}(t,\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})), \quad \mathbf{x}^{\mathrm{cc}}(t_0,\mathbf{p}) = \mathbf{x}_0^{\mathrm{cc}}(\mathbf{p}).
\end{aligned} \tag{6.3.2}$$

We note that our new framework for evaluating state relaxation subgradients will only be valid under (6.3.1), so that $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ can be obtained using (6.3.2). However, the requirement $\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}) > \mathbf{x}^{\mathrm{cv}}(t,\mathbf{p})$ will generally be satisfied if $\mathbf{f}(t,\cdot,\cdot)$ in (2.3.1) is nonlinear for each $t > t_0$, and this can also be guaranteed by adding an arbitrarily small perturbation to $(\mathbf{u}, \mathbf{o}, \mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}})$, as will be shown in Section 6.4.4. As concluded in [79], in the Scott–Barton framework (2.4.1), the state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ converge faster to the original trajectory $\mathbf{x}$ than the state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ as the domain P shrinks. Thus, $\mathbf{x}^{\mathrm{L}}(t) < \mathbf{x}^{\mathrm{cv}}(t,\mathbf{p})$ and $\mathbf{x}^{\mathrm{U}}(t) > \mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})$ in (6.3.1) are guaranteed to be satisfied for a sufficiently small P. Moreover, we suspect that the requirement (6.3.1) could be weakened, but there is no clear way to do this using currently established differential inequalities.

Section 6.4 below presents sophisticated mathematical foundations of the new subgradient evaluation methods. For practical methods for evaluating subgradients of the established state relaxations [2, 3], readers are recommended to look at Sections 6.5 and 6.6.

6.4 Subgradient evaluation framework

In this section, we present a new framework to evaluate subgradients for state relaxations constructed using (6.3.2). This framework constructs state relaxation subgradients $(\mathbf{S}^{\mathrm{cv}}, \mathbf{S}^{\mathrm{cc}})$ as the solutions of a new auxiliary ODE system. In Section 6.6, numerical examples will use this framework to compute subgradients.

6.4.1 Subgradient propagation functions

Our new subgradient evaluation framework requires furnishing crucial *subgradient propagation functions*, for use in an auxiliary ODE system that will be constructed in the next subsection.

Definition 6.4.1. Suppose that Assumption 6.3.2 holds. 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 $(\mathbf{u}, \mathbf{o})$ if the following conditions are satisfied: for any interval $Z \in \mathbb{IR}^{n_x \times n_p}$,

I.1 for each $\mathbf{p} \in \tilde{P}$ and $\mathbf{M}, \mathbf{N} \in Z$, the functions $\mathbf{V}(\cdot, \mathbf{p}, \mathbf{M}, \mathbf{N})$ and $\mathbf{W}(\cdot, \mathbf{p}, \mathbf{M}, \mathbf{N})$ are measurable on I,

I.2 there exist functions $m_Z, k_Z : I \times \tilde{P} \to \mathbb{R}_+ \cup \{+\infty\}$ so that for each $t \in I$, $\mathbf{p} \in \tilde{P}$, and $\mathbf{M}^{\mathrm{A}}, \mathbf{N}^{\mathrm{A}}, \mathbf{M}^{\mathrm{B}}, \mathbf{N}^{\mathrm{B}} \in Z$, $m_Z(\cdot, \mathbf{p})$ and $k_Z(\cdot, \mathbf{p})$ are Lebesgue integrable, and

$$\|\mathbf{V}(t, \mathbf{p}, \mathbf{M}^{\mathrm{A}}, \mathbf{N}^{\mathrm{A}})\| + \|\mathbf{W}(t, \mathbf{p}, \mathbf{M}^{\mathrm{A}}, \mathbf{N}^{\mathrm{A}})\| \leq m_Z(t, \mathbf{p})$$

and

$$\begin{aligned}&\|\mathbf{V}(t, \mathbf{p}, \mathbf{M}^{\mathrm{A}}, \mathbf{N}^{\mathrm{A}}) - \mathbf{V}(t, \mathbf{p}, \mathbf{M}^{\mathrm{B}}, \mathbf{N}^{\mathrm{B}})\| + \|\mathbf{W}(t, \mathbf{p}, \mathbf{M}^{\mathrm{A}}, \mathbf{N}^{\mathrm{A}}) - \mathbf{W}(t, \mathbf{p}, \mathbf{M}^{\mathrm{B}}, \mathbf{N}^{\mathrm{B}})\|\\ &\leq k_Z(t, \mathbf{p})\Big(\|\mathbf{M}^{\mathrm{A}} - \mathbf{M}^{\mathrm{B}}\| + \|\mathbf{N}^{\mathrm{A}} - \mathbf{N}^{\mathrm{B}}\|\Big),\end{aligned}$$

I.3 for a.e. $t \in I$, any $i \in \{1,...,n_x\}$, $\mathbf{p} \in \tilde{P}$, $\mathbf{h} \in \mathbb{R}^{n_p}$, $\mathbf{d}^{\mathrm{A}}, \mathbf{d}^{\mathrm{B}} \in \mathbb{R}^{n_x}$, and $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{n_x \times n_p}$ such that

$$\mathbf{Mh} \leq \mathbf{d}^{\mathrm{A}},$$

$$\mathbf{Nh} \geq \mathbf{d}^{\mathrm{B}},$$

the following conditions hold:

(a) if $\langle \mathbf{m}_{(i)}, \mathbf{h} \rangle = d_i^{\mathrm{A}}$, then

$$\langle \mathbf{v}_{(i)}(t,\mathbf{p},\mathbf{M},\mathbf{N}), \mathbf{h} \rangle \leq [u_{i,t}]'((\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}));(\mathbf{h},\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}})),$$

(b) if $\langle \mathbf{n}_{(i)}, \mathbf{h} \rangle = d_i^{\mathrm{B}}$, then

$$\langle \mathbf{w}_{(i)}(t,\mathbf{p},\mathbf{M},\mathbf{N}), \mathbf{h} \rangle \geq [o_{i,t}]'((\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}));(\mathbf{h},\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}})),$$

where $\mathbf{u}_t \equiv \mathbf{u}(t,\cdot,\cdot,\cdot)$ and $\mathbf{o}_t \equiv \mathbf{o}(t,\cdot,\cdot,\cdot)$.

The conditions in the definition above may look cumbersome. However, Section 6.4.3 below will present a practical method to construct subgradient propagation functions $(\mathbf{V},\mathbf{W})$ for any $(\mathbf{u},\mathbf{o})$ satisfying (2.4.2).

6.4.2 Subgradient evaluation ODE system

As the core result of this article, the following theorem shows that the unique solution of a certain auxiliary parametric ODE system describes valid state relaxation subgradients $(\mathbf{S}^{\mathrm{cv}},\mathbf{S}^{\mathrm{cc}})$ for the state relaxations $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$ obtained using (6.3.2). This auxiliary ODE system employs subgradient propagation functions $(\mathbf{V},\mathbf{W})$.

Theorem 6.4.2. Suppose that Assumption 6.3.2 holds and that valid subgradient propagation functions $(\mathbf{V},\mathbf{W})$ for $(\mathbf{u},\mathbf{o})$ are available. Define functions $\mathbf{S}_0^{\mathrm{cv}},\mathbf{S}_0^{\mathrm{cc}}:\tilde{P}\to\mathbb{R}^{n_x\times n_p}$ so that for each $i\in\{1,\dots,n_x\}$ and $\mathbf{p}\in\tilde{P}$, $[\mathbf{s}_{(i),0}^{\mathrm{cv}}(\mathbf{p})]^{\mathrm{T}}$ is a subgradient of $x_{i,0}^{\mathrm{cv}}$ at $\mathbf{p}$, and $[\mathbf{s}_{(i),0}^{\mathrm{cc}}(\mathbf{p})]^{\mathrm{T}}$ is a subgradient of $x_{i,0}^{\mathrm{cc}}$ at $\mathbf{p}$. Consider the following parametric ODE system:

$$\begin{aligned}\dot{\mathbf{S}}^{\mathrm{cv}}(t,\mathbf{p})&=\mathbf{V}(t,\mathbf{p},\mathbf{S}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{S}^{\mathrm{cc}}(t,\mathbf{p})),\quad \mathbf{S}^{\mathrm{cv}}(t_0,\mathbf{p})=\mathbf{S}_0^{\mathrm{cv}}(\mathbf{p}),\\ \dot{\mathbf{S}}^{\mathrm{cc}}(t,\mathbf{p})&=\mathbf{W}(t,\mathbf{p},\mathbf{S}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{S}^{\mathrm{cc}}(t,\mathbf{p})),\quad \mathbf{S}^{\mathrm{cc}}(t_0,\mathbf{p})=\mathbf{S}_0^{\mathrm{cc}}(\mathbf{p}).\end{aligned}\tag{6.4.1}$$

Then, for each $\mathbf{p}\in\tilde{P}$, local existence and uniqueness of a Carathéodory solution $(\mathbf{S}^{\mathrm{cv}}(\cdot,\mathbf{p}),\mathbf{S}^{\mathrm{cc}}(\cdot,\mathbf{p}))$ of (6.4.1) are guaranteed. Moreover, $(\mathbf{S}^{\mathrm{cv}},\mathbf{S}^{\mathrm{cc}})$ are valid subgradients for state relaxations $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$ as in Definition 6.3.3.

Proof. Since Conditions I.1 and I.2 in Definition 6.4.1 hold, [102, §1, Theorems 1 and 2] imply local existence and uniqueness of a Carathéodory solution of (6.4.1) for each $\mathbf{p}\in\tilde{P}$.

Let $\mathbf{x}_t^{\mathrm{cv}}\equiv\mathbf{x}^{\mathrm{cv}}(t,\cdot)$ and $\mathbf{x}_t^{\mathrm{cc}}\equiv\mathbf{x}^{\mathrm{cc}}(t,\cdot)$. Since $\mathbf{x}_t^{\mathrm{cv}}$ and $\mathbf{x}_t^{\mathrm{cc}}$ are respectively convex and concave on P for each $t\in I$, Theorem 6.2.3 implies that, for proving the claimed result, it is sufficient to show that for all $t\in I$, $\mathbf{p}\in\tilde{P}$, and directions $\mathbf{h}\in\mathbb{R}^{n_p}$,

$$\mathbf{S}^{\mathrm{cv}}(t,\mathbf{p})\mathbf{h}\le[\mathbf{x}_t^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h})\quad\text{and}\quad\mathbf{S}^{\mathrm{cc}}(t,\mathbf{p})\mathbf{h}\ge[\mathbf{x}_t^{\mathrm{cc}}]'(\mathbf{p};\mathbf{h}).$$

Consider any fixed $\mathbf{p}\in\tilde{P}$ and direction $\mathbf{h}\in\mathbb{R}^{n_p}$. Since $\mathbf{x}_0^{\mathrm{cv}}$ and $\mathbf{x}_0^{\mathrm{cc}}$ are respectively convex and concave on P, [146, Theorems 10.4 and 23.4] imply that they are Lipschitz continuous and directionally differentiable on $\tilde{P}$. Furthermore, since Assumption 6.3.2 holds, by applying Theorem 6.2.4 to a reformulated (6.3.2) where the states $\mathbf{x}^{\mathrm{cc}}$ are replaced by their negatives, the mapping $t\mapsto([\mathbf{x}_t^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}),[\mathbf{x}_t^{\mathrm{cc}}]'(\mathbf{p};\mathbf{h}))$ may be described as the unique solution $(\mathbf{z}^{\mathrm{A}},\mathbf{z}^{\mathrm{B}})$ of the following ODE system:

$$\dot{\mathbf{z}}^{\mathrm{A}}(t) = [\mathbf{u}_t]'\big((\mathbf{p},\mathbf{x}_t^{\mathrm{cv}}(\mathbf{p}),\mathbf{x}_t^{\mathrm{cc}}(\mathbf{p}));(\mathbf{h},\mathbf{z}^{\mathrm{A}}(t),\mathbf{z}^{\mathrm{B}}(t))\big),\quad \mathbf{z}^{\mathrm{A}}(t_0) = [\mathbf{x}_0^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}),$$
$$\dot{\mathbf{z}}^{\mathrm{B}}(t) = [\mathbf{o}_t]'\big((\mathbf{p},\mathbf{x}_t^{\mathrm{cv}}(\mathbf{p}),\mathbf{x}_t^{\mathrm{cc}}(\mathbf{p}));(\mathbf{h},\mathbf{z}^{\mathrm{A}}(t),\mathbf{z}^{\mathrm{B}}(t))\big),\quad \mathbf{z}^{\mathrm{B}}(t_0) = [\mathbf{x}_0^{\mathrm{cc}}]'(\mathbf{p};\mathbf{h}).$$

For simplicity of notation, define functions $\mathbf{g}^u,\mathbf{g}^o : I\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$ so that for each $(t,\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}})\in I\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$,

$$\mathbf{g}^u(t,\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}}) := [\mathbf{u}_t]'((\mathbf{p},\mathbf{x}_t^{\mathrm{cv}}(\mathbf{p}),\mathbf{x}_t^{\mathrm{cc}}(\mathbf{p}));(\mathbf{h},\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}})),$$
$$\mathbf{g}^o(t,\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}}) := [\mathbf{o}_t]'((\mathbf{p},\mathbf{x}_t^{\mathrm{cv}}(\mathbf{p}),\mathbf{x}_t^{\mathrm{cc}}(\mathbf{p}));(\mathbf{h},\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}})).$$

Since $(\mathbf{u},\mathbf{o})$ satisfy relevant conditions in Assumption 6.3.2, [127, Theorem 3.1.2] implies that there exists a scalar $l>0$ so that for each $t\in I$ and $\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}},\hat{\mathbf{d}}^{\mathrm{A}},\hat{\mathbf{d}}^{\mathrm{B}}\in\mathbb{R}^{n_x}$,

$$\|\mathbf{g}^u(t,\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}})-\mathbf{g}^u(t,\hat{\mathbf{d}}^{\mathrm{A}},\hat{\mathbf{d}}^{\mathrm{B}})\|+\|\mathbf{g}^o(t,\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}})-\mathbf{g}^o(t,\hat{\mathbf{d}}^{\mathrm{A}},\hat{\mathbf{d}}^{\mathrm{B}})\|\le l\big(\|\mathbf{d}^{\mathrm{A}}-\hat{\mathbf{d}}^{\mathrm{A}}\|+\|\mathbf{d}^{\mathrm{B}}-\hat{\mathbf{d}}^{\mathrm{B}}\|\big). \tag{6.4.2}$$

We proceed in this proof by showing that for all $t\in[t_0,t_f]$, $\mathbf{S}^{\mathrm{cv}}(t,\mathbf{p})\mathbf{h}\le[\mathbf{x}_t^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h})$. That $\mathbf{S}^{\mathrm{cc}}(t,\mathbf{p})\mathbf{h}\ge[\mathbf{x}_t^{\mathrm{cc}}]'(\mathbf{p};\mathbf{h})$ can be proved similarly. The rest of this proof is based on the differential inequality results developed in [39, 91]. According to the construction of $\mathbf{S}_0^{\mathrm{cv}}$, $\mathbf{S}^{\mathrm{cv}}(t_0,\mathbf{p})\mathbf{h}\le[\mathbf{x}_{t_0}^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h})$. For all $t\in(t_0,t_f]$, to arrive at a contradiction, suppose that there exists $\tilde{t}\in(t_0,t_f]$ for which $[x_{i,\tilde{t}}^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h})<\langle\mathbf{s}_{(i)}^{\mathrm{cv}}(\tilde{t},\mathbf{p}),\mathbf{h}\rangle$ for some $i\in\{1,\dots,n_x\}$, and thus define

$$t_1 := \inf\{t\in(t_0,t_f] : \exists\kappa\in\{1,\dots,n_x\}\text{ for which }[x_{\kappa,t}^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h})<\langle\mathbf{s}_{(\kappa)}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{h}\rangle\}\le t_f.$$

Define a function $\boldsymbol{\delta}: I \to \mathbb{R}^{2n_x}$ so that for each $t \in I$,

$$\boldsymbol{\delta}(t) := (\mathbf{S}^{\mathrm{cv}}(t,\mathbf{p})\mathbf{h} - [\mathbf{x}_t^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}), [\mathbf{x}_t^{\mathrm{cc}}]'(\mathbf{p};\mathbf{h}) - \mathbf{S}^{\mathrm{cc}}(t,\mathbf{p})\mathbf{h}).$$

Observe that $\boldsymbol{\delta}$ is absolutely continuous. Applying [39, Lemmata 3.3.4 and 3.3.5] to $\boldsymbol{\delta}$ and t_1, we obtain the following.

Let $\mathbf{1} \in \mathbb{R}^n$ be a vector whose components are all equal to 1. It holds that $t_1 < t_f$, and, for any $t_4 \in (t_1, t_f]$, there exist $j \in \{1, ..., n_x\}$, $\varepsilon > 0$, an absolutely continuous and non-deceasing function $\rho : [t_1, t_4] \to \mathbb{R}$ whose derivative a.e. on $[t_1, t_4]$ is denoted as $\dot{\rho}$, and scalars $t_2, t_3 \in [t_1, t_4]$ with $t_2 < t_3$ such that

$$0 < \rho(t) \leq \varepsilon, \quad \forall t \in [t_1, t_4], \tag{6.4.3}$$

$$\dot{\rho}(t) > 2l\rho(t), \quad \text{a.e. } t \in [t_1, t_4], \tag{6.4.4}$$

$$\mathbf{S}^{\mathrm{cv}}(t,\mathbf{p})\mathbf{h} - \rho(t)\mathbf{1} < [\mathbf{x}_t^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}), \quad \forall t \in [t_2, t_3), \tag{6.4.5}$$

$$\mathbf{S}^{\mathrm{cc}}(t,\mathbf{p})\mathbf{h} + \rho(t)\mathbf{1} > [\mathbf{x}_t^{\mathrm{cc}}]'(\mathbf{p};\mathbf{h}), \quad \forall t \in [t_2, t_3), \tag{6.4.6}$$

$$[x_{j,t_2}^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}) = \langle \mathbf{s}_{(j)}^{\mathrm{cv}}(t_2,\mathbf{p}), \mathbf{h} \rangle, \tag{6.4.7}$$

$$[x_{j,t_3}^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}) = \langle \mathbf{s}_{(j)}^{\mathrm{cv}}(t_3,\mathbf{p}), \mathbf{h} \rangle - \rho(t_3), \tag{6.4.8}$$

$$[x_{j,t}^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}) < \langle \mathbf{s}_{(j)}^{\mathrm{cv}}(t,\mathbf{p}), \mathbf{h} \rangle, \quad \forall t \in (t_2, t_3). \tag{6.4.9}$$

Define functions $\tilde{\mathbf{z}}, \tilde{\mathbf{y}} : I \to \mathbb{R}^{n_x}$ so that for each $t \in I$ and $\kappa \in \{1, ..., n_x\}$,

$$\tilde{z}_\kappa(t) := \max(\langle \mathbf{s}_{(\kappa)}^{\mathrm{cv}}(t,\mathbf{p}), \mathbf{h} \rangle, [x_{\kappa,t}^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h})),$$

$$\tilde{y}_\kappa(t) := \min(\langle \mathbf{s}_{(\kappa)}^{\mathrm{cc}}(t,\mathbf{p}), \mathbf{h} \rangle, [x_{\kappa,t}^{\mathrm{cc}}]'(\mathbf{p};\mathbf{h})).$$

Moreover, since (6.4.7) and (6.4.9) hold, it follows that for all $t \in [t_2, t_3)$,

$$\tilde{\mathbf{z}}(t) \geq \mathbf{S}^{\mathrm{cv}}(t, \mathbf{p})\mathbf{h},$$

$$\tilde{z}_j(t) = \langle \mathbf{s}^{\mathrm{cv}}_{(j)}(t, \mathbf{p}), \mathbf{h} \rangle,$$

$$\tilde{\mathbf{y}}(t) \leq \mathbf{S}^{\mathrm{cc}}(t, \mathbf{p})\mathbf{h}.$$

Since Condition I.3 in Definition 6.4.1 holds, it follows that

$$g^u_j(t, \tilde{\mathbf{z}}(t), \tilde{\mathbf{y}}(t)) \geq \langle \mathbf{v}_{(j)}(t, \mathbf{p}, \mathbf{S}^{\mathrm{cv}}(t, \mathbf{p}), \mathbf{S}^{\mathrm{cc}}(t, \mathbf{p})), \mathbf{h} \rangle, \quad \text{a.e. } t \in [t_2, t_3). \tag{6.4.10}$$

Now, for each $t \in [t_2, t_3)$ and each $\kappa \in \{1, ..., n_x\}$, one of the following cases will occur:

1. if $[x^{\mathrm{cv}}_{\kappa,t}]'(\mathbf{p}; \mathbf{h}) \geq \langle \mathbf{s}^{\mathrm{cv}}_{(\kappa)}(t, \mathbf{p}), \mathbf{h} \rangle$, then

$$\tilde{z}_\kappa(t) - [x^{\mathrm{cv}}_{\kappa,t}]'(\mathbf{p}; \mathbf{h}) = 0, \tag{6.4.11}$$

2. if $[x^{\mathrm{cv}}_{\kappa,t}]'(\mathbf{p}; \mathbf{h}) \leq \langle \mathbf{s}^{\mathrm{cv}}_{(\kappa)}(t, \mathbf{p}), \mathbf{h} \rangle$, then $\tilde{z}_\kappa(t) = \langle \mathbf{s}^{\mathrm{cv}}_{(\kappa)}(t, \mathbf{p}), \mathbf{h} \rangle$; moreover, since (6.4.5) holds,

$$0 < \tilde{z}_\kappa(t) - [x^{\mathrm{cv}}_{\kappa,t}]'(\mathbf{p}; \mathbf{h}) < \rho(t). \tag{6.4.12}$$

The following inequality follows from (6.4.11) and (6.4.12):

$$\|\tilde{\mathbf{z}}(t) - [\mathbf{x}^{\mathrm{cv}}_t]'(\mathbf{p}; \mathbf{h})\| < \rho(t), \quad \forall t \in [t_2, t_3). \tag{6.4.13}$$

Similarly, for all $t \in [t_2, t_3)$,

$$\|\tilde{\mathbf{y}}(t) - [\mathbf{x}^{\mathrm{cc}}_t]'(\mathbf{p}; \mathbf{h})\| < \rho(t), \quad \forall t \in [t_2, t_3). \tag{6.4.14}$$

Combining (6.4.13) and (6.4.14) with (6.4.2) yields

$$|g_j^u(t,\tilde{\mathbf{z}}(t),\tilde{\mathbf{y}}(t)) - g_j^u(t,[\mathbf{x}_t^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}),[\mathbf{x}_t^{\mathrm{cc}}]'(\mathbf{p};\mathbf{h}))| < 2l\rho(t), \quad \text{a.e. } t \in [t_2,t_3). \tag{6.4.15}$$

Applying (6.4.10) yields

$$\langle \mathbf{v}_{(j)}(t,\mathbf{p},\mathbf{S}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{S}^{\mathrm{cc}}(t,\mathbf{p})),\mathbf{h}\rangle < g_j^u(t,[\mathbf{x}_t^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}),[\mathbf{x}_t^{\mathrm{cc}}]'(\mathbf{p};\mathbf{h})) + 2l\rho(t), \quad \text{a.e. } t \in [t_2,t_3).$$

Since $\dot{\rho}(t) > 2l\rho(t)$ for a.e. $t \in [t_2,t_3]$ according to (6.4.4), rearranging the above inequality yields

$$\langle \mathbf{v}_{(j)}(t,\mathbf{p},\mathbf{S}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{S}^{\mathrm{cc}}(t,\mathbf{p})),\mathbf{h}\rangle - g_j^u(t,[\mathbf{x}_t^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}),[\mathbf{x}_t^{\mathrm{cc}}]'(\mathbf{p};\mathbf{h})) - \dot{\rho}(t) < 0, \quad \text{a.e. } t \in [t_2,t_3],$$

which implies that $\left(\langle \mathbf{s}_{(j)}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{h}\rangle - [x_{j,t}^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}) - \rho(t)\right)$ is decreasing on $[t_2,t_3]$, which in turn implies

$$\langle \mathbf{s}_{(j)}^{\mathrm{cv}}(t_2,\mathbf{p}),\mathbf{h}\rangle - [x_{j,t_2}^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}) - \rho(t_2) > \langle \mathbf{s}_{(j)}^{\mathrm{cv}}(t_3,\mathbf{p}),\mathbf{h}\rangle - [x_{j,t_3}^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}) - \rho(t_3). \tag{6.4.16}$$

However, since (6.4.7) and (6.4.8) hold, then (6.4.16) becomes $\rho(t_2) < 0$ which contradicts (6.4.3). Thus $\tilde{t}$ cannot exist.

Since the choices of $\mathbf{p}$ and $\mathbf{h}$ are arbitrary in the proof above, it follows that for all $t \in (t_0,t_f]$, $\mathbf{p} \in \tilde{P}$, and $\mathbf{h} \in \mathbb{R}^{n_p}$,

$$\mathbf{S}^{\mathrm{cv}}(t,\mathbf{p})\mathbf{h} \leq [\mathbf{x}_t^{\mathrm{cv}}]'(\mathbf{p};\mathbf{h}).$$

□

6.4.3 Constructing subgradient propagation functions

In this subsection, we show that for any Scott–Barton right-hand side functions $(\mathbf{u},\mathbf{o})$ in Assumption 6.3.2 that are constructed using (2.4.2), the subgradient propagation functions $(\mathbf{V},\mathbf{W})$ in Theorem 6.4.2 can be constructed by leveraging the subgradient information of $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ in Assumption 2.4.11.

Besides Assumption 2.4.11, suppose that functions $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ in (2.4.2) satisfy the following assumption.

Assumption 6.4.3. Suppose that Assumption 2.4.11 holds. Suppose that the mappings $\tilde{\mathbf{u}}(t,\cdot,\cdot,\cdot)$ and $\tilde{\mathbf{o}}(t,\cdot,\cdot,\cdot)$ in (2.4.2) are directionally differentiable and Lipschitz continuous on $\tilde{P}\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$, uniformly over $t\in I$.

The following proposition will be useful for the subsequent theory development.

Proposition 6.4.4. Suppose that functions $\tilde{\mathbf{u}},\tilde{\mathbf{o}}: I\times P\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$ describe convexity amplifying dynamics as in [2, Definition 5]. Then, these functions also satisfy the following condition. For a.e. $t\in I$, any $\hat{P}\in\mathbb{I}\tilde{P}$, and arbitrary functions $\boldsymbol{\alpha},\boldsymbol{\beta}:\hat{P}\to\mathbb{R}^{n_x}$ for which the following conditions hold:

1. for each $\hat{\mathbf{p}}\in\hat{P}$, $\mathbf{x}^{\mathrm{L}}(t)<\boldsymbol{\alpha}(\hat{\mathbf{p}})\leq\boldsymbol{\beta}(\hat{\mathbf{p}})<\mathbf{x}^{\mathrm{U}}(t)$ and

2. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are respectively convex and concave on $\hat{P}$,

then, the mappings $\mathbf{p}\mapsto\tilde{\mathbf{u}}(t,\mathbf{p},\boldsymbol{\alpha}(\mathbf{p}),\boldsymbol{\beta}(\mathbf{p}))$ and $\mathbf{p}\mapsto\tilde{\mathbf{o}}(t,\mathbf{p},\boldsymbol{\alpha}(\mathbf{p}),\boldsymbol{\beta}(\mathbf{p}))$ are respectively convex and concave on $\hat{P}$.

Proof. This result can be verified by applying [2, Definition 5] of convexity amplifying dynamics of $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ to any $\mathbf{p}^{\mathrm{A}},\mathbf{p}^{\mathrm{B}},\tilde{\mathbf{p}}\in\hat{P}$ and $\lambda\in(0,1)$ for which $\tilde{\mathbf{p}}:=\lambda\mathbf{p}^{\mathrm{A}}+(1-\lambda)\mathbf{p}^{\mathrm{B}}$. □

In concert with the functions $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ in Assumptions 2.4.11 and 6.4.3, we shall consider functions $(\tilde{\mathbf{V}}, \tilde{\mathbf{W}})$ that satisfy the following assumption.

Assumption 6.4.5. Suppose that functions $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ satisfy Assumptions 2.4.11 and 6.4.3. Suppose that functions $\tilde{\mathbf{V}}, \tilde{\mathbf{W}} : I \times \tilde{P} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x \times n_p} \times \mathbb{R}^{n_x \times n_p} \to \mathbb{R}^{n_x \times n_p}$ satisfy the following conditions. Consider a.e. $t \in I$, any $\hat{P} \in \mathbb{I}\tilde{P}$, arbitrary functions $\boldsymbol{\alpha}, \boldsymbol{\beta} : \hat{P} \to \mathbb{R}^{n_x}$, $\mathbf{p}$ in the interior of $\hat{P}$ and $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{n_x \times n_p}$ for which the following conditions hold:

1. for each $\hat{\mathbf{p}} \in \hat{P}$, $\mathbf{x}^{\mathrm{L}}(t) < \boldsymbol{\alpha}(\hat{\mathbf{p}}) \leq \boldsymbol{\beta}(\hat{\mathbf{p}}) < \mathbf{x}^{\mathrm{U}}(t)$,

2. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are respectively convex and concave on $\hat{P}$, and

3. for each $i \in \{1, ..., n_x\}$, $\mathbf{m}_{(i)}$ and $\mathbf{n}_{(i)}$ are respectively subgradients of α_i and β_i at $\mathbf{p}$.

According to Proposition 6.4.4, the mappings $\tilde{\mathbf{u}}(t, \cdot, \boldsymbol{\alpha}(\cdot), \boldsymbol{\beta}(\cdot))$ and $\tilde{\mathbf{o}}(t, \cdot, \boldsymbol{\alpha}(\cdot), \boldsymbol{\beta}(\cdot))$ are respectively convex and concave on $\hat{P}$. Assume that for each $i \in \{1, ..., n_x\}$,

$$\tilde{\mathbf{v}}_{(i)}(t, \mathbf{p}, \boldsymbol{\alpha}(\mathbf{p}), \boldsymbol{\beta}(\mathbf{p}), \mathbf{M}, \mathbf{N}) \text{ and } \tilde{\mathbf{W}}_{(i)}(t, \mathbf{p}, \boldsymbol{\alpha}(\mathbf{p}), \boldsymbol{\beta}(\mathbf{p}), \mathbf{M}, \mathbf{N})$$

are subgradients of $\tilde{u}_i(t, \cdot, \boldsymbol{\alpha}(\cdot), \boldsymbol{\beta}(\cdot))$ and $\tilde{o}_i(t, \cdot, \boldsymbol{\alpha}(\cdot), \boldsymbol{\beta}(\cdot))$ at $\mathbf{p}$, respectively.

Theorem 6.4.8 below shows that valid subgradient propagation functions $(\mathbf{V}, \mathbf{W})$ in Definition 6.4.1 may be constructed by composing functions $(\tilde{\mathbf{V}}, \tilde{\mathbf{W}})$ in Assumption 6.4.5 with new matrix flattening operations as in Definition 6.4.7 below. This theorem requires the following lemma as an intermediate result.

Lemma 6.4.6. Suppose that Assumption 6.4.5 holds. For a.e. $t \in I$, any $\mathbf{p} \in \tilde{P}$, $\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}} \in \mathbb{R}^{n_x}$, $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{n_x \times n_p}$, $\mathbf{d}^{\mathrm{A}}, \mathbf{d}^{\mathrm{B}} \in \mathbb{R}^{n_x}$, and $\mathbf{h} \in \mathbb{R}^{n_p}$ for which the following conditions hold:

1. $\mathbf{x}^{\mathrm{L}}(t) < \boldsymbol{\xi}^{\mathrm{cv}} \leq \boldsymbol{\xi}^{\mathrm{cc}} < \mathbf{x}^{\mathrm{U}}(t)$,

2. for each $i \in \{1,...,n_x\}$ such that $\xi_i^{\mathrm{cv}} < \xi_i^{\mathrm{cc}}$,

$$\langle \mathbf{m}_{(i)}, \mathbf{h} \rangle \leq d_i^{\mathrm{A}} \quad \text{and} \quad \langle \mathbf{n}_{(i)}, \mathbf{h} \rangle \geq d_i^{\mathrm{B}}, \tag{6.4.17}$$

3. for each $j \in \{1,...,n_x\}$ such that $\xi_j^{\mathrm{cv}} = \xi_j^{\mathrm{cc}}$,

$$\mathbf{m}_{(j)} = \mathbf{n}_{(j)}, \quad d_j^{\mathrm{A}} = d_j^{\mathrm{B}}, \quad \text{and} \quad \langle \mathbf{m}_{(j)}, \mathbf{h} \rangle = d_j^{\mathrm{A}}, \tag{6.4.18}$$

then the following inequalities hold:

$$\begin{aligned} \tilde{\mathbf{V}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}, \mathbf{M}, \mathbf{N})\mathbf{h} &\leq [\tilde{\mathbf{u}}_t]'(\mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}; \mathbf{h}, \mathbf{d}^{\mathrm{A}}, \mathbf{d}^{\mathrm{B}}) \\ \text{and} \quad \tilde{\mathbf{W}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}, \mathbf{M}, \mathbf{N})\mathbf{h} &\geq [\tilde{\mathbf{o}}_t]'(\mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}; \mathbf{h}, \mathbf{d}^{\mathrm{A}}, \mathbf{d}^{\mathrm{B}}). \end{aligned} \tag{6.4.19}$$

Proof. Consider a.e. $t \in I$, any $\mathbf{p} \in \tilde{P}$, $\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}} \in \mathbb{R}^{n_x}$, $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{n_x \times n_p}$, $\mathbf{d}^{\mathrm{A}}, \mathbf{d}^{\mathrm{B}} \in \mathbb{R}^{n_x}$, and $\mathbf{h} \in \mathbb{R}^{n_p}$ as in the lemma's statement.

Define functions $\boldsymbol{\alpha}, \boldsymbol{\beta} : \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$ in the following way.

1. Choose vectors $\mathbf{q}_{(i)}, \mathbf{e}_{(i)} \in \mathbb{R}^{n_p}$ for which $\langle \mathbf{q}_{(i)}, \mathbf{h} \rangle = d_i^{\mathrm{A}}$ and $\langle \mathbf{e}_{(i)}, \mathbf{h} \rangle = d_i^{\mathrm{B}}$. For each $i \in \{1,...,n_x\}$ such that $\xi_i^{\mathrm{cv}} < \xi_i^{\mathrm{cc}}$ and each $\tilde{\mathbf{p}} \in \mathbb{R}^{n_p}$, define

$$\begin{aligned} \alpha_i(\tilde{\mathbf{p}}) &:= \max(\xi_i^{\mathrm{cv}} + \langle \mathbf{m}_{(i)}, \tilde{\mathbf{p}} - \mathbf{p} \rangle, \xi_i^{\mathrm{cv}} + \langle \mathbf{q}_{(i)}, \tilde{\mathbf{p}} - \mathbf{p} \rangle) \\ \text{and} \quad \beta_i(\tilde{\mathbf{p}}) &:= \min(\xi_i^{\mathrm{cc}} + \langle \mathbf{n}_{(i)}, \tilde{\mathbf{p}} - \mathbf{p} \rangle, \xi_i^{\mathrm{cc}} + \langle \mathbf{e}_{(i)}, \tilde{\mathbf{p}} - \mathbf{p} \rangle). \end{aligned}$$

Observe that $\alpha_i(\mathbf{p}) = \xi_i^{\mathrm{cv}}$, $\beta_i(\mathbf{p}) = \xi_i^{\mathrm{cc}}$, α_i and β_i are piecewise affine, α_i is convex, β_i is concave, and $\mathbf{m}_{(i)}$ and $\mathbf{n}_{(i)}$ are respectively subgradients of α_i and β_i at $\mathbf{p}$. Besides, since (6.4.17) holds, and since $\langle \mathbf{q}_{(i)}, \mathbf{h} \rangle = d_i^{\mathrm{A}}$ and $\langle \mathbf{e}_{(i)}, \mathbf{h} \rangle = d_i^{\mathrm{B}}$, [129, Example 3.4, Page 260] implies that $[\alpha_i]'(\mathbf{p}; \mathbf{h}) = d_i^{\mathrm{A}}$ and $[\beta_i]'(\mathbf{p}; \mathbf{h}) = d_i^{\mathrm{B}}$.

2. For each $j \in \{1,...,n_x\}$ such that $\xi_j^{\mathrm{cv}} = \xi_j^{\mathrm{cc}}$ and each $\tilde{\mathbf{p}} \in \mathbb{R}^{n_p}$,

$$\begin{aligned} \alpha_j(\tilde{\mathbf{p}}) &:= \xi_j^{\mathrm{cv}} + \langle \mathbf{m}_{(j)}, \tilde{\mathbf{p}} - \mathbf{p} \rangle \\ \text{and} \quad \beta_j(\tilde{\mathbf{p}}) &:= \xi_j^{\mathrm{cc}} + \langle \mathbf{n}_{(j)}, \tilde{\mathbf{p}} - \mathbf{p} \rangle. \end{aligned}$$

Since (6.4.18) holds, it follows that α_j and β_j are the same affine function, $\alpha_j(\mathbf{p}) = \xi_j^{\mathrm{cv}} = \beta_j(\mathbf{p}) = \xi_j^{\mathrm{cc}}$, $[\alpha_j]'(\mathbf{p};\mathbf{h}) = d_j^{\mathrm{A}} = [\beta_j]'(\mathbf{p};\mathbf{h}) = d_j^{\mathrm{B}}$, $\mathbf{m}_{(j)}$ and $\mathbf{n}_{(j)}$ are respectively slopes of the affine functions α_j and β_j.

Moreover, since functions $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are continuous by construction, and since $\mathbf{x}^{\mathrm{L}}(t) < \boldsymbol{\xi}^{\mathrm{cv}} \leq \boldsymbol{\xi}^{\mathrm{cc}} < \mathbf{x}^{\mathrm{U}}(t)$, there exists a set $\hat{P} \in \mathbb{I}\tilde{P}$ such that $\mathbf{p}$ is in the interior of $\hat{P}$ and $\mathbf{x}^{\mathrm{L}}(t) < \boldsymbol{\alpha}(\tilde{\mathbf{p}}) \leq \boldsymbol{\beta}(\tilde{\mathbf{p}}) < \mathbf{x}^{\mathrm{U}}(t)$ for each $\tilde{\mathbf{p}} \in \hat{P}$.

Now, since $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}})$ and $(\tilde{\mathbf{V}}, \tilde{\mathbf{W}})$ satisfy Assumption 6.4.5, according to the construction of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, for all $\tilde{\mathbf{h}} \in \mathbb{R}^{n_p}$

$$\begin{aligned} \tilde{\mathbf{V}}(t, \mathbf{p}, \boldsymbol{\alpha}(\mathbf{p}), \boldsymbol{\beta}(\mathbf{p}), \mathbf{M}, \mathbf{N})\tilde{\mathbf{h}} &\leq [\tilde{\mathbf{u}}_t]'(\mathbf{p}, \boldsymbol{\alpha}(\mathbf{p}), \boldsymbol{\beta}(\mathbf{p}); \tilde{\mathbf{h}}, \boldsymbol{\alpha}'(\mathbf{p};\tilde{\mathbf{h}}), \boldsymbol{\beta}'(\mathbf{p};\tilde{\mathbf{h}})) \\ \text{and} \quad \tilde{\mathbf{W}}(t, \mathbf{p}, \boldsymbol{\alpha}(\mathbf{p}), \boldsymbol{\beta}(\mathbf{p}), \mathbf{M}, \mathbf{N})\tilde{\mathbf{h}} &\geq [\tilde{\mathbf{o}}_t]'(\mathbf{p}, \boldsymbol{\alpha}(\mathbf{p}), \boldsymbol{\beta}(\mathbf{p}); \tilde{\mathbf{h}}, \boldsymbol{\alpha}'(\mathbf{p};\tilde{\mathbf{h}}), \boldsymbol{\beta}'(\mathbf{p};\tilde{\mathbf{h}})). \end{aligned}$$

Then, (6.4.19) is proved by substituting $(\boldsymbol{\alpha}(\mathbf{p}), \boldsymbol{\beta}(\mathbf{p}))$ with $(\boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})$, letting $\tilde{\mathbf{h}} := \mathbf{h}$, and substituting $(\boldsymbol{\alpha}'(\mathbf{p};\mathbf{h}), \boldsymbol{\beta}'(\mathbf{p};\mathbf{h}))$ with $(\mathbf{d}^{\mathrm{A}}, \mathbf{d}^{\mathrm{B}})$.

□

Definition 6.4.7. For each $i \in \{1,...,n_x\}$, define functions $\mathbf{R}^{i,\mathrm{L}}, \mathbf{R}^{i,\mathrm{U}} : \mathbb{R}^{n_x \times n_p} \times \mathbb{R}^{n_x \times n_p} \to \mathbb{R}^{n_x \times n_p} \times \mathbb{R}^{n_x \times n_p}$ by:

1. $\mathbf{R}^{i,\mathrm{L}}(\mathbf{M}^{\mathrm{A}}, \mathbf{N}^{\mathrm{A}}) := (\mathbf{M}^{\mathrm{A}}, \mathbf{N}^{\mathrm{B}})$ where $\mathbf{n}_{(k)}^{\mathrm{B}} := \mathbf{n}_{(k)}^{\mathrm{A}}$, for all $k \in \{1,...,n_x\}$ and $k \neq i$ and $\mathbf{n}_{(i)}^{\mathrm{B}} := \mathbf{m}_{(i)}^{\mathrm{A}}$,

2. $\mathbf{R}^{i,\mathrm{U}}(\mathbf{M}^{\mathrm{A}},\mathbf{N}^{\mathrm{A}}) := (\mathbf{M}^{\mathrm{B}},\mathbf{N}^{\mathrm{A}})$ where $\mathbf{m}^{\mathrm{B}}_{(k)} := \mathbf{m}^{\mathrm{A}}_{(k)}$, for all $k \in \{1,...,n_x\}$ and $k \neq i$, and $\mathbf{m}^{\mathrm{B}}_{(i)} := \mathbf{n}^{\mathrm{A}}_{(i)}$.

Theorem 6.4.8. Suppose that functions $\tilde{\mathbf{u}},\tilde{\mathbf{o}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ satisfy Assumption 6.4.3. Construct functions $\mathbf{u},\mathbf{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ using (2.4.2). Then, the mappings $\mathbf{u}(t,\cdot,\cdot,\cdot)$ and $\mathbf{o}(t,\cdot,\cdot,\cdot)$ are directionally differentiable and Lipschitz continuous on $\tilde{P} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly over $t \in I$. Suppose that the remaining conditions in Assumption 6.3.2 also hold. Consider a set $\tilde{X} \in \mathbb{IR}^{n_x}$ for which $X(t) \subset \tilde{X}, \forall t \in I$. Suppose that functions $(\tilde{\mathbf{V}},\tilde{\mathbf{W}})$ satisfy Assumption 6.4.5 and the following conditions: for any $Z \in \mathbb{IR}^{n_x \times n_p}$,

II.1 for each $\mathbf{p} \in \tilde{P}$ and $\mathbf{M},\mathbf{N} \in Z$, the functions $\tilde{\mathbf{V}}(\cdot,\mathbf{p},\cdot,\cdot,\mathbf{M},\mathbf{N})$ and $\tilde{\mathbf{W}}(\cdot,\mathbf{p},\cdot,\cdot,\mathbf{M},\mathbf{N})$ are *Borel measurable* (c.f. [147]),

II.2 there exist scalars $\tilde{m}_Z,\tilde{k}_Z > 0$ so that for each $t \in I$, $\mathbf{p} \in \tilde{P}$, $\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}} \in \tilde{X}$, and $\mathbf{M}^{\mathrm{A}},\mathbf{N}^{\mathrm{A}},\mathbf{M}^{\mathrm{B}},\mathbf{N}^{\mathrm{B}} \in Z$,

$$\|\tilde{\mathbf{V}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M}^{\mathrm{A}},\mathbf{N}^{\mathrm{A}})\| + \|\tilde{\mathbf{W}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M}^{\mathrm{A}},\mathbf{N}^{\mathrm{A}})\| \leq \tilde{m}_Z \tag{6.4.20}$$

and

$$\begin{aligned}
&\|\tilde{\mathbf{V}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M}^{\mathrm{A}},\mathbf{N}^{\mathrm{A}}) - \tilde{\mathbf{V}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M}^{\mathrm{B}},\mathbf{N}^{\mathrm{B}})\| \\
&+ \|\tilde{\mathbf{W}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M}^{\mathrm{A}},\mathbf{N}^{\mathrm{A}}) - \tilde{\mathbf{W}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M}^{\mathrm{B}},\mathbf{N}^{\mathrm{B}})\| \\
&\leq \tilde{k}_Z(\|\mathbf{M}^{\mathrm{A}} - \mathbf{M}^{\mathrm{B}}\| + \|\mathbf{N}^{\mathrm{A}} - \mathbf{N}^{\mathrm{B}}\|).
\end{aligned} \tag{6.4.21}$$

Define functions $\mathbf{V},\mathbf{W} : I \times \tilde{P} \times \mathbb{R}^{n_x \times n_p} \times \mathbb{R}^{n_x \times n_p} \to \mathbb{R}^{n_x \times n_p}$ so that for each $i \in \{1,...,n_x\}$

and $(t,\mathbf{p},\mathbf{M},\mathbf{N}) \in I \times \tilde{P} \times \mathbb{R}^{n_x \times n_p} \times \mathbb{R}^{n_x \times n_p}$,

$$\begin{aligned} \mathbf{v}_{(i)}(t,\mathbf{p},\mathbf{M},\mathbf{N}) &:= \tilde{\mathbf{v}}_{(i)}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})),\mathbf{R}^{i,\mathrm{L}}(\mathbf{M},\mathbf{N})), \\ \text{and} \quad \mathbf{w}_{(i)}(t,\mathbf{p},\mathbf{M},\mathbf{N}) &:= \tilde{\mathbf{w}}_{(i)}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{U}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})),\mathbf{R}^{i,\mathrm{U}}(\mathbf{M},\mathbf{N})). \end{aligned} \tag{6.4.22}$$

Then, the functions $(\mathbf{V},\mathbf{W})$ are valid subgradient propagation functions for $(\mathbf{u},\mathbf{o})$ as in Definition 6.4.1.

Proof. Since the functions $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ satisfy Assumption 2.4.11, [2, Lemmata 10 and 11] imply that $(\mathbf{u},\mathbf{o})$ are valid Scott–Barton right–hand side functions. Since $(\mathbf{r}^{i,\mathrm{L}},\mathbf{r}^{i,\mathrm{U}})$ are linear mappings, $(\mathbf{u},\mathbf{o})$ also inherits the directional differentiability and uniform Lipschitz continuity of $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ in Assumption 6.4.3. Thus, the mappings $\mathbf{u}(t,\cdot,\cdot,\cdot)$ and $\mathbf{o}(t,\cdot,\cdot,\cdot)$ are directionally differentiable and Lipschitz continuous on $\tilde{P} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly over $t \in I$.

Next, we show that the $(\mathbf{V},\mathbf{W})$ in (6.4.22) are valid subgradient propagation functions by verifying all conditions in Definition 6.4.1. Since for each $\mathbf{p} \in P$, $(\mathbf{x}^{\mathrm{cv}}(\cdot,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(\cdot,\mathbf{p}))$ are continuous, it follows that $(\mathbf{x}^{\mathrm{cv}}(\cdot,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(\cdot,\mathbf{p}))$ are Borel measurable. Since a composition of two Borel measurable functions is also Borel measurable (c.f. [147, Section 2.12(ii)]), Condition II.1 implies that $\mathbf{V}(\cdot,\mathbf{p},\mathbf{M},\mathbf{N})$ and $\mathbf{W}(\cdot,\mathbf{p},\mathbf{M},\mathbf{N})$ satisfy Condition I.1. Moreover, since $\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}) \in X(t) \subset \tilde{X}, \forall (t,\mathbf{p}) \in I \times P$ according to [2, Lemma 1], Condition II.2 implies that Condition I.2 is satisfied with $m_Z(t,\mathbf{p}) := \tilde{m}_Z$ and $k_Z(t,\mathbf{p}) := \tilde{k}_Z$.

Now, we verify Condition I.3. Consider a.e. $t \in I$, any $\mathbf{p} \in \tilde{P}$, $i \in \{1,...,n_x\}$, $\mathbf{M},\mathbf{N} \in \mathbb{R}^{n_x \times n_p}$, $\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}} \in \mathbb{R}^{n_x}$, and $\mathbf{h} \in \mathbb{R}^{n_p}$ such that

$$\begin{aligned} \mathbf{M}\mathbf{h} &\le \mathbf{d}^{\mathrm{A}}, \\ \mathbf{N}\mathbf{h} &\ge \mathbf{d}^{\mathrm{B}}. \end{aligned} \tag{6.4.23}$$

Further assume that $\langle \mathbf{m}_{(i)}, \mathbf{h} \rangle = d_i^{\mathrm{A}}$. Let

$$\begin{aligned}(\hat{\boldsymbol{\xi}}^{\mathrm{cv}}, \hat{\boldsymbol{\xi}}^{\mathrm{cc}}) &:= \mathbf{r}^{i,\mathrm{L}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}), \mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})),\\ (\hat{\mathbf{M}}, \hat{\mathbf{N}}) &:= \mathbf{R}^{i,\mathrm{L}}(\mathbf{M}, \mathbf{N}),\\ (\hat{\mathbf{d}}^{\mathrm{A}}, \hat{\mathbf{d}}^{\mathrm{B}}) &:= \mathbf{r}^{i,\mathrm{L}}(\mathbf{d}^{\mathrm{A}}, \mathbf{d}^{\mathrm{B}}).\end{aligned} \tag{6.4.24}$$

Since $\mathbf{x}^{\mathrm{L}}(t) < \mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}) < \mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}) < \mathbf{x}^{\mathrm{U}}(t)$ by assumption, since (6.4.23) holds, and based on the definitions of $\mathbf{r}^{i,\mathrm{L}}$ and $\mathbf{R}^{i,\mathrm{L}}$, it follows that $\mathbf{x}^{\mathrm{L}}(t) < \hat{\boldsymbol{\xi}}^{\mathrm{cv}}$ and $\hat{\boldsymbol{\xi}}^{\mathrm{cc}} < \mathbf{x}^{\mathrm{U}}(t)$, and for any index $j \in \{1, ..., n_x\}$ with $j \neq i$,

$$\hat{\xi}_j^{\mathrm{cv}} < \hat{\xi}_j^{\mathrm{cc}}, \quad \langle \hat{\mathbf{m}}_{(j)}, \mathbf{h} \rangle \leq \hat{d}_j^{\mathrm{A}}, \quad \langle \hat{\mathbf{n}}_{(j)}, \mathbf{h} \rangle \geq \hat{d}_j^{\mathrm{B}},$$

and

$$\hat{\mathbf{m}}_{(i)} = \hat{\mathbf{n}}_{(i)} = \mathbf{m}_{(i)}, \quad \hat{\xi}_i^{\mathrm{cv}} = \hat{\xi}_i^{\mathrm{cc}} = x_i^{\mathrm{cv}}(t,\mathbf{p}), \quad \hat{d}_i^{\mathrm{A}} = \hat{d}_i^{\mathrm{B}} = d_i^{\mathrm{A}}.$$

Moreover, since $\langle \mathbf{m}_{(i)}, \mathbf{h} \rangle = d_i^{\mathrm{A}}$, it follows that $\langle \hat{\mathbf{m}}_{(i)}, \mathbf{h} \rangle = \hat{d}_i^{\mathrm{A}}$. Thus, Proposition 6.4.4 and Lemma 6.4.6 imply that

$$\langle \tilde{\mathbf{v}}_{(i)}(t, \mathbf{p}, \hat{\boldsymbol{\xi}}^{\mathrm{cv}}, \hat{\boldsymbol{\xi}}^{\mathrm{cc}}, \hat{\mathbf{M}}, \hat{\mathbf{N}}), \mathbf{h} \rangle \leq [\tilde{u}_{i,t}]'((\mathbf{p}, \hat{\boldsymbol{\xi}}^{\mathrm{cv}}, \hat{\boldsymbol{\xi}}^{\mathrm{cc}}); (\mathbf{h}, \hat{\mathbf{d}}^{\mathrm{A}}, \hat{\mathbf{d}}^{\mathrm{B}})).$$

According to (6.4.24),

$$\begin{aligned}&\langle \tilde{\mathbf{v}}_{(i)}(t, \mathbf{p}, \mathbf{r}^{i,\mathrm{L}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}), \mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})), \mathbf{R}^{i,\mathrm{L}}(\mathbf{M}, \mathbf{N})), \mathbf{h} \rangle\\ &\qquad \leq [\tilde{u}_{i,t}]'\big((\mathbf{p}, \mathbf{r}^{i,\mathrm{L}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}), \mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}))); (\mathbf{h}, \mathbf{r}^{i,\mathrm{L}}(\mathbf{d}^{\mathrm{A}}, \mathbf{d}^{\mathrm{B}}))\big).\end{aligned}$$

Moreover, since (2.4.2) and (6.4.22) hold, the inequality above becomes

$$\langle \mathbf{v}_{(i)}(t,\mathbf{p},\mathbf{M},\mathbf{N}),\mathbf{h}\rangle \leq [u_{i,t}]'((\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}));(\mathbf{h},\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}})).$$

If we instead assume that $\langle \mathbf{n}_{(i)},\mathbf{h}\rangle = d_i^{\mathrm{B}}$, it can be shown similarly that

$$\langle \mathbf{w}_{(i)}(t,\mathbf{p},\mathbf{M},\mathbf{N}),\mathbf{h}\rangle \geq [o_{i,t}]'((\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}));(\mathbf{h},\mathbf{d}^{\mathrm{A}},\mathbf{d}^{\mathrm{B}})).$$

Thus, $(\mathbf{V},\mathbf{W})$ satisfy Condition I.3 of Definition 6.4.1. □

Remark 6.4.9. The requirement $\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}) < \mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})$ in Assumption 6.3.2 is essential for the theorem above. To illustrate this point, consider any $i \in \{1,\ldots,n_x\}$ and a index $j \neq i$ for which $x_j^{\mathrm{cv}}(t,\mathbf{p}) = x_j^{\mathrm{cc}}(t,\mathbf{p})$. In this case, for Lemma 6.4.6 to be applicable in the proof above, we must have $d_j^{\mathrm{A}} = d_j^{\mathrm{B}}$. However, Condition I.3 of Definition 6.4.1 requires considering any $d_j^{\mathrm{A}}, d_j^{\mathrm{B}} \in \mathbb{R}$. Thus, coinciding $\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p})$ and $\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})$ cannot be applied.

6.4.4 Accommodating coinciding state relaxations

According to [2, Corollary 1], any state relaxations $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$ obtained using (2.4.1) satisfy $\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}) \leq \mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})$, for each $(t,\mathbf{p}) \in I \times P$. In this subsection, still assuming that valid state relaxations $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$ are obtained using (6.3.2), we show that the requirement $\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}) < \mathbf{x}^{\mathrm{cc}}(t,\mathbf{p})$ in Assumption 6.3.2 can be guaranteed by adding an arbitrarily small perturbation $\varepsilon > 0$ to $(\mathbf{u},\mathbf{o},\mathbf{x}_0^{cv},\mathbf{x}_0^{cc})$ in (6.3.2). This result requires the following additional assumption.

Assumption 6.4.10. Consider functions $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ that satisfy Assumption 2.4.11. Further assume that for each $t \in I$, $\mathbf{p} \in P$, and $\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,A}},\boldsymbol{\xi}^{\mathrm{cc,B}} \in \mathbb{R}^{n_x}$ for which $\boldsymbol{\xi}^{\mathrm{cv,A}} \leq$

$\boldsymbol{\xi}^{\mathrm{cv,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,B}} \leq \boldsymbol{\xi}^{\mathrm{cc,A}}$,

$$\tilde{\mathbf{u}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}) \leq \tilde{\mathbf{u}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}}) \leq \tilde{\mathbf{o}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,B}},\boldsymbol{\xi}^{\mathrm{cc,B}}) \leq \tilde{\mathbf{o}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv,A}},\boldsymbol{\xi}^{\mathrm{cc,A}}).$$

The assumption above is mild and is already required in [39, 79]. This assumption is necessary for the resulting state relaxations to exhibit desirable convergence properties (in the sense of [79]) to $\mathbf{x}$ as P shrinks in methods for deterministic global dynamic optimization. The following proposition establishes strict inequalities between $\mathbf{x}^{\mathrm{cv}}$ and $\mathbf{x}^{\mathrm{cc}}$.

Proposition 6.4.11. Consider functions $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ that satisfy Assumption 6.4.10, and consider functions $(\mathbf{u},\mathbf{o})$ constructed using (2.4.2). Consider the following $(\mathbf{p},\varepsilon)$–dependent ODE system: for each $i \in \{1,...,n_x\}$,

$$\begin{aligned}
\dot{\hat{x}}_i^{\mathrm{cv}}(t,\mathbf{p},\varepsilon) &= u_i(t,\mathbf{p},\hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},\varepsilon),\hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},\varepsilon)) - \varepsilon, \quad \hat{x}_i^{\mathrm{cv}}(t_0,\mathbf{p},\varepsilon) = x_{i,0}^{\mathrm{cv}}(\mathbf{p}) - \varepsilon, \\
\dot{\hat{x}}_i^{\mathrm{cc}}(t,\mathbf{p},\varepsilon) &= o_i(t,\mathbf{p},\hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},\varepsilon),\hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},\varepsilon)) + \varepsilon, \quad \hat{x}_i^{\mathrm{cc}}(t_0,\mathbf{p},\varepsilon) = x_{i,0}^{\mathrm{cc}}(\mathbf{p}) + \varepsilon.
\end{aligned} \tag{6.4.25}$$

For each $\mathbf{p} \in P$ and $\varepsilon \in \mathbb{R}$, any solution $(\hat{\mathbf{x}}^{\mathrm{cv}}(,\cdot,\mathbf{p},\varepsilon),\hat{\mathbf{x}}^{\mathrm{cc}}(,\cdot,\mathbf{p},\varepsilon))$ is understood in the classical sense. Consider an arbitrarily small $\hat{\varepsilon} > 0$ and suppose that for each $\mathbf{p} \in P$ and $t \in I$,

$$\mathbf{x}^{\mathrm{L}}(t) \leq \hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},\hat{\varepsilon}) \quad \text{and} \quad \hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},\hat{\varepsilon}) \leq \mathbf{x}^{U}(t), \tag{6.4.26}$$

and also suppose that

$$\mathbf{x}^{\mathrm{L}}(t) \leq \hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},0) \quad \text{and} \quad \hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},0) \leq \mathbf{x}^{U}(t). \tag{6.4.27}$$

Then, $(\hat{\mathbf{x}}^{\mathrm{cv}}(\cdot,\cdot,0),\hat{\mathbf{x}}^{\mathrm{cc}}(\cdot,\cdot,0))$ and $(\hat{\mathbf{x}}^{\mathrm{cv}}(\cdot,\cdot,\hat{\varepsilon}),\hat{\mathbf{x}}^{\mathrm{cc}}(\cdot,\cdot,\hat{\varepsilon}))$ are both valid state relaxations for

(2.3.1). Moreover,

$$\hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},\hat{\varepsilon}) < \hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},0) \leq \hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},0) < \hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},\hat{\varepsilon}), \quad \forall t \in I,\ \forall \mathbf{p} \in P. \tag{6.4.28}$$

Proof. Since $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ in Assumption 2.4.11 are continuous, the definition of a solution of (6.4.25) is appropriate. Since $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ satisfy Assumption 2.4.11 and $(\mathbf{u},\mathbf{o})$ are constructed using (2.4.2), [2, Lemmata 10 and 11] imply that $(\mathbf{u},\mathbf{o})$ are valid Scott–Barton right-hand side functions. Moreover, it is readily verified that $(\mathbf{u}-\hat{\varepsilon},\mathbf{o}+\hat{\varepsilon})$ are also valid Scott–Barton right-hand side functions. Thus, since (6.4.26) and (6.4.27) are assumed, [2, Corollary 1 and Theorem 3] imply that both $(\hat{\mathbf{x}}^{\mathrm{cv}}(\cdot,\cdot,0),\hat{\mathbf{x}}^{\mathrm{cc}}(\cdot,\cdot,0))$ and $(\hat{\mathbf{x}}^{\mathrm{cv}}(\cdot,\cdot,\hat{\varepsilon}),\hat{\mathbf{x}}^{\mathrm{cc}}(\cdot,\cdot,\hat{\varepsilon}))$ are valid state relaxations for (2.3.1).

Since $(\hat{\mathbf{x}}^{\mathrm{cv}}(\cdot,\cdot,0),\hat{\mathbf{x}}^{\mathrm{cc}}(\cdot,\cdot,0))$ are valid state relaxations, it follows that $\hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},0) \leq \hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},0)$ for each $t \in I$ and $\mathbf{p} \in P$. Now, we prove the remaining inequalities in (6.4.28) using a differential inequality result [77, Theorem III, §12]. Consider any fixed $\mathbf{p} \in P$.

Consider any $t \in I$, $i \in \{1,...,n_x\}$, and $\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}} \in \mathbb{R}^{n_x}$ so that $\hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},\hat{\varepsilon}) \leq \boldsymbol{\xi}^{\mathrm{cv}} \leq \boldsymbol{\xi}^{\mathrm{cc}} \leq \hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},\hat{\varepsilon})$. Further assume that $\xi_i^{\mathrm{cv}} = \hat{x}_i^{\mathrm{cv}}(t,\mathbf{p},\varepsilon)$. Since Assumption 6.4.10 holds, according to the definition of $\mathbf{r}^{i,\mathrm{L}}$,

$$\tilde{u}_i(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})) \geq \tilde{u}_i(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},\hat{\varepsilon}),\hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},\hat{\varepsilon}))).$$

Moreover, since (2.4.2) holds and $\hat{\varepsilon} > 0$,

$$\begin{aligned} u_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) &\geq u_i(t,\mathbf{p},\hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},\hat{\varepsilon}),\hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},\hat{\varepsilon})) \\ &> u_i(t,\mathbf{p},\hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},\hat{\varepsilon}),\hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},\hat{\varepsilon})) - \hat{\varepsilon}. \end{aligned}$$

Similarly, if we instead assume $\xi_i^{\mathrm{cc}} = \hat{x}_i^{\mathrm{cc}}(t,\mathbf{p},\hat{\varepsilon})$, it can be shown that

$$o_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) < o_i(t,\mathbf{p},\hat{\mathbf{x}}^{\mathrm{cv}}(t,\mathbf{p},\hat{\varepsilon}),\hat{\mathbf{x}}^{\mathrm{cc}}(t,\mathbf{p},\hat{\varepsilon})) + \hat{\varepsilon}.$$

Moreover, since $x_{i,0}^{\mathrm{cv}}(\mathbf{p}) - \hat{\varepsilon} < x_{i,0}^{\mathrm{cv}}(\mathbf{p})$ and $x_{i,0}^{\mathrm{cc}}(\mathbf{p}) + \hat{\varepsilon} > x_{i,0}^{\mathrm{cc}}(\mathbf{p})$ and since the ODE solutions $(\hat{\mathbf{x}}^{\mathrm{cv}}(\cdot,\mathbf{p},0),\hat{\mathbf{x}}^{\mathrm{cc}}(\mathbf{p},\cdot,0))$ and $(\hat{\mathbf{x}}^{\mathrm{cv}}(\cdot,\mathbf{p},\hat{\varepsilon}),\hat{\mathbf{x}}^{\mathrm{cc}}(\cdot,\mathbf{p},\hat{\varepsilon}))$ are continuously differentiable, (6.4.28) follows by applying [77, Theorem III, §12] to (6.4.25) reformulated with $\varepsilon := 0$ and with $\varepsilon := \hat{\varepsilon}$, where the states $\hat{\mathbf{x}}^{\mathrm{cc}}$ are replaced by their negatives. $\square$

Note that $\hat{\mathbf{x}}^{\mathrm{cv}}(\cdot,\cdot,\hat{\varepsilon})$ and $\hat{\mathbf{x}}^{\mathrm{cc}}(\cdot,\cdot,\hat{\varepsilon})$ still have desirable convergence properties to $\mathbf{x}$ as P shrinks, if $\hat{\varepsilon}$ is set to be a fixed fraction of the diameter of P. Moreover, the subgradient evaluation results Theorems 6.4.2 and 6.4.8 are still applicable to (6.3.2) if $(\mathbf{u}-\hat{\varepsilon},\mathbf{o}+\hat{\varepsilon})$ is embedded instead of $(\mathbf{u},\mathbf{o})$. The following proposition shows that if an extremely small $\hat{\varepsilon} > 0$ is chosen, subgradients of the "$\hat{\varepsilon}$-perturbed" state relaxations approximate subgradients of the original state relaxations.

Proposition 6.4.12. Consider the setup in Proposition 6.4.11. For any $0 < \varepsilon \leq \hat{\varepsilon}$, consider the resulting state relaxations $(\hat{\mathbf{x}}^{\mathrm{cv}}(\cdot,\cdot,\varepsilon),\hat{\mathbf{x}}^{\mathrm{cc}}(\cdot,\cdot,\varepsilon))$ obtained using (6.4.25), and for each $t \in I$, denote the subdifferential of $\hat{x}_i^{\mathrm{cv}}(t,\cdot,\varepsilon)$ and $\hat{x}_i^{\mathrm{cc}}(t,\cdot,\varepsilon)$ at a $\mathbf{p} \in \tilde{P}$ as $\partial\hat{x}_i^{\mathrm{cv}}(t,\mathbf{p},\varepsilon)$ and $\partial\hat{x}_i^{\mathrm{cc}}(t,\mathbf{p},\varepsilon)$, respectively. Then,

$$\begin{aligned} &\limsup_{\varepsilon\downarrow 0} \partial\hat{x}_i^{\mathrm{cv}}(t,\mathbf{p},\varepsilon) \subset \partial\hat{x}_i^{\mathrm{cv}}(t,\mathbf{p},0) \\ \text{and}\quad &\limsup_{\varepsilon\downarrow 0} \partial\hat{x}_i^{\mathrm{cc}}(t,\mathbf{p},\varepsilon) \subset \partial\hat{x}_i^{\mathrm{cc}}(t,\mathbf{p},0) \end{aligned}$$

(c.f. [148, Definition 4.1] for "lim sup").

Proof. Since functions $(\mathbf{u},\mathbf{o})$ satisfy the first two conditions in Definition 2.4.9, for any

$\mathbf{p} \in \tilde{P}$ and $\varepsilon \in \mathbb{R}$, [99, Theorem 6.1, §III] implies that (6.4.25) has exactly one solution on I. Moreover, since functions $(\mathbf{u}, \mathbf{o}, \mathbf{x}_0^{\mathrm{cv}}, \mathbf{x}_0^{\mathrm{cc}})$ are continuous, [99, Theorem 2.1, §V] implies that the solution $(\hat{\mathbf{x}}^{\mathrm{cv}}, \hat{\mathbf{x}}^{\mathrm{cc}})$ are continuous on $I \times \tilde{P} \times \mathbb{R}$. Thus, for any $t \in I$, as $\varepsilon \downarrow 0$, the mappings $\hat{\mathbf{x}}^{\mathrm{cv}}(t, \cdot, \varepsilon)$ and $\hat{\mathbf{x}}^{\mathrm{cc}}(t, \cdot, \varepsilon)$ converge pointwise to $\hat{\mathbf{x}}^{\mathrm{cv}}(t, \cdot, 0)$ and $\hat{\mathbf{x}}^{\mathrm{cc}}(t, \cdot, 0)$, respectively. Then, the claimed result follows by applying [146, Theorem 24.5] to $(\hat{x}_i^{\mathrm{cv}}(t, \mathbf{p}, \cdot), \hat{x}_i^{\mathrm{cc}}(t, \mathbf{p}, \cdot))$ for each $i \in \{1, ..., n_x\}$ and $(t, \mathbf{p}) \in I \times \tilde{P}$. □

6.4.5 Adjoint subgradient evaluation

In this subsection, we propose to evaluate the subgradients described by Theorem 6.4.2 via an adjoint ODE system. Since adjoint sensitivity evaluation is relatively computational inexpensive, we expect that this will be useful when solving lower-bounding problems in deterministic algorithms of global dynamic optimization. Consider a nonconvex dynamic optimization problem with (2.3.1) embedded:

$$\min_{\mathbf{p} \in P} c(\mathbf{p}) := g(t_f, \mathbf{p}, \mathbf{x}(t_f, \mathbf{p})), \tag{6.4.29}$$

where $\mathbf{p} \in P$ denotes decision variables, $c : \mathbb{R}^{n_p} \to \mathbb{R}$ is an objective function based on a continuous cost function $g : I \times P \times \mathbb{R}^{n_x} \to \mathbb{R}$. A lower-bounding problem for (6.4.29), embedded in an overarching global optimization method, is typically a convex optimization problem whose optimal objective value is a valid lower bound of the optimal objective value of (6.4.29). Consider any state relaxations $\mathbf{x}^{\mathrm{C}} := (\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ for (2.3.1) obtained using (6.3.2). A lower-bounding problem for (6.4.29) is typically constructed as:

$$\min_{\mathbf{p} \in P} c^{\mathrm{cv}}(\mathbf{p}) := g^{\mathrm{cv}}(t_f, \mathbf{p}, \mathbf{x}^{\mathrm{C}}(t_f, \mathbf{p})), \tag{6.4.30}$$

where the function $c^{\text{cv}} : \mathbb{R}^{n_p} \to \mathbb{R}$ is a convex relaxation of c on P, and $g^{\text{cv}} : I \times P \times \mathbb{R}^{2n_x} \to \mathbb{R}$ may be constructed using various adaptations [6, 7, 33, 34] of the McCormick relaxation method [5]. When solving (6.4.30) with nonsmooth NLP solvers, Subgradients of c^{cv} are typically required by these solvers to proceed effectively. In a forward subgradient evaluation method, given state relaxation subgradients

$$\mathbf{S}^{\text{C}} := \begin{bmatrix} \mathbf{S}^{\text{cv}} \\ \mathbf{S}^{\text{cc}} \end{bmatrix}$$

for the state relaxations $\mathbf{x}^{\text{C}}$ obtained by solving (6.4.1), subgradients of c^{cv} may be computed by applying certain subgradient evaluation rules (e.g. [7, 33, 35, 145]) for g^{cv}. However, since the ODE system (6.4.1) has $2n_x \times n_p$ state variables, as n_x and n_p increase, the number of state variables of (6.4.1) will significantly increase, and thus solving (6.4.1) may become intractable. As summarized in [90], reverse (adjoint) sensitivity analysis is more efficient for the situation of computing gradients of a scalar function of state variables with respect to a large number of parameters. In smooth dynamic optimization, an adjoint sensitivity approach allows evaluating derivatives of the objective function without evaluating the partial derivatives of state variables with respect to the uncertain parameters. The following theorem shows that if the subgradients of c^{cv} can be evaluated by applying the classical chain rule as for gradients, and if the forward subgradient evaluation system (6.4.1) is an affine ODE system, then the subgradients of c^{cv} may be evaluated directly using adjoints. The assumptions of this theorem can often be satisfied in practice. For example, given $(\mathbf{S}^{\text{cv}}, \mathbf{S}^{\text{cc}})$, the methods for evaluating subgradients of c^{cv} in [7, 35, 145] satisfy the classical chain rule. As will be seen in Section 6.5, the forward subgradient evaluation systems derived from (6.4.1) for the established state relaxations [2, 3] are affine

ODE systems. Thus, this result essentially extends classical dynamic adjoint sensitivity methods to nonsmooth subgradient evaluation.

Theorem 6.4.13. Consider the lower bounding problem (6.4.30) where the state relaxations $\mathbf{x}^{\mathrm{C}} := (\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ are constructed by solving (6.3.2). Suppose that the following conditions hold for a fixed $\tilde{\mathbf{p}} \in \tilde{P}$.

III.1 Let $(\boldsymbol{\rho}_0, \boldsymbol{\rho})$ be the transpose of a subgradient of $g^{\mathrm{cv}}(t_f, \cdot, \cdot)$ at $(\tilde{\mathbf{p}}, \mathbf{x}^{\mathrm{C}}(t_f, \tilde{\mathbf{p}}))$. Denote the transpose of a subgradient of c^{cv} at $\tilde{\mathbf{p}}$ as $\tilde{\mathbf{s}}$. Suppose that

$$\tilde{\mathbf{s}} := \boldsymbol{\rho}_0 + \boldsymbol{\rho}\mathbf{S}^{\mathrm{C}}(t_f, \tilde{\mathbf{p}}). \tag{6.4.31}$$

III.2 State relaxation subgradients $\mathbf{S}^{\mathrm{C}}(\cdot, \tilde{\mathbf{p}}) := (\mathbf{S}^{\mathrm{cv}}(\cdot, \tilde{\mathbf{p}}), \mathbf{S}^{\mathrm{cc}}(\cdot, \tilde{\mathbf{p}}))$ for $\mathbf{x}^{\mathrm{C}}(\cdot, \tilde{\mathbf{p}})$ as in Definition 6.3.3 are computed by solving (6.4.1) with $\mathbf{p} := \tilde{\mathbf{p}}$, and (6.4.1) also has the following affine form:

$$\begin{aligned} \dot{\mathbf{S}}^{\mathrm{C}}(t, \tilde{\mathbf{p}}) &= \boldsymbol{\Theta}^{\mathrm{A}}(t, \tilde{\mathbf{p}})\mathbf{S}^{\mathrm{C}}(t, \tilde{\mathbf{p}}) + \boldsymbol{\Theta}^{\mathrm{B}}(t, \tilde{\mathbf{p}}), \\ \mathbf{S}^{\mathrm{C}}(t_0, \tilde{\mathbf{p}}) &= \mathbf{S}_0^{\mathrm{C}}(\tilde{\mathbf{p}}). \end{aligned} \tag{6.4.32}$$

where $\boldsymbol{\Theta}^{\mathrm{A}}, \boldsymbol{\Theta}^{\mathrm{B}} : I \times \tilde{P} \to \mathbb{R}^{2n_x \times 2n_x}$ and $\mathbf{S}_0^{\mathrm{C}} := (\mathbf{S}_0^{\mathrm{cv}}, \mathbf{S}_0^{\mathrm{cc}})$ in (6.4.1).

Let $\boldsymbol{\lambda} : I \to \mathbb{R}^{2n_x}$ be a Carathéodory solution (as described in [102]) of the following adjoint ODE system on I:

$$\begin{aligned} (\dot{\boldsymbol{\lambda}}(t))^{\mathrm{T}} &= -(\boldsymbol{\lambda}(t))^{\mathrm{T}}\boldsymbol{\Theta}^{\mathrm{A}}(t, \tilde{\mathbf{p}}), \\ (\boldsymbol{\lambda}(t_f))^{\mathrm{T}} &= \boldsymbol{\rho}. \end{aligned} \tag{6.4.33}$$

Then, the following holds:

$$\tilde{\mathbf{s}} \equiv (\boldsymbol{\lambda}(t_0))^{\mathrm{T}}\mathbf{S}_0^{\mathrm{C}}(\tilde{\mathbf{p}}) + \boldsymbol{\rho}_0 + \int_{t_0}^{t_f} (\boldsymbol{\lambda}(t))^{\mathrm{T}}\boldsymbol{\Theta}^{\mathrm{B}}(t,\tilde{\mathbf{p}})\mathrm{d}t. \tag{6.4.34}$$

Proof. Define a function $\boldsymbol{\gamma}: I \times \mathbb{R}^{n_p} \times \mathbb{R}^{2n_x} \to \mathbb{R}^{2n_x}$ so that for each $(t,\mathbf{p},\mathbf{z}) \in I \times \mathbb{R}^{n_p} \times \mathbb{R}^{2n_x}$,

$$\boldsymbol{\gamma}(t,\mathbf{p},\mathbf{z}) := \boldsymbol{\Theta}^{\mathrm{A}}(t,\tilde{\mathbf{p}})(\mathbf{z} - \mathbf{x}^{\mathrm{C}}(t,\tilde{\mathbf{p}})) + \boldsymbol{\Theta}^{\mathrm{B}}(t,\tilde{\mathbf{p}})(\mathbf{p} - \tilde{\mathbf{p}}).$$

Consider the following affine ODE system on I for each $\mathbf{p} \in \mathbb{R}^{n_p}$:

$$\begin{aligned} \dot{\mathbf{y}}(t,\mathbf{p}) &= \boldsymbol{\gamma}(t,\mathbf{p},\mathbf{y}(t,\mathbf{p})), \\ \mathbf{y}(t_0,\mathbf{p}) &= \mathbf{S}_0^{\mathrm{C}}(\tilde{\mathbf{p}})(\mathbf{p} - \tilde{\mathbf{p}}). \end{aligned} \tag{6.4.35}$$

For a.e. $t \in I$, since $\boldsymbol{\gamma}(t,\cdot,\cdot)$ is affine, the Clarke's *generalized Jacobian* (c.f. [136, Definition 2.6.1]) of $\boldsymbol{\gamma}(t,\cdot,\cdot)$ at each $(\mathbf{p},\mathbf{z}) \in \mathbb{R}^{n_p} \times \mathbb{R}^{2n_x}$ reduces to a singleton. Thus, [136, Theorem 7.4.1] implies that the function $\mathbf{y}(t_f,\cdot)$ is *strictly differentiable* (c.f. [136, Proposition 2.2.1]) on $\mathbb{R}^{n_p}$. Furthermore, according to [97, Corollary 4.3], for each $\mathbf{p} \in \mathbb{R}^{n_p}$, the mapping $t \mapsto \frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t,\mathbf{p})$ may be described as the solution of the following ODE system on I:

$$\dot{\mathbf{H}}(t) = \boldsymbol{\Theta}^{\mathrm{A}}(t,\tilde{\mathbf{p}})\mathbf{H}(t) + \boldsymbol{\Theta}^{\mathrm{B}}(t,\tilde{\mathbf{p}}), \quad \mathbf{H}(t_0) = \mathbf{S}_0^{\mathrm{C}}(\tilde{\mathbf{p}}), \tag{6.4.36}$$

which indicates that $\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\mathbf{p}}(t,\mathbf{p})$ is independent of $\mathbf{p}$, and thus the mapping $\mathbf{y}(t,\cdot)$ is affine for each $t \in I$. Observe that (6.4.36) and (6.4.32) are the same ODE system, which implies that

$$\mathbf{S}^{\mathrm{C}}(t_f,\tilde{\mathbf{p}}) \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t_f,\tilde{\mathbf{p}}). \tag{6.4.37}$$

Now define $\tilde{g}:\mathbb{R}^{n_p}\times\mathbb{R}^{n_x}\to\mathbb{R}$ so that for each $\mathbf{p}\in\mathbb{R}^{n_p}$ and $\mathbf{z}\in\mathbb{R}^{n_x}$,

$$\tilde{g}(\mathbf{p},\mathbf{z}) := \boldsymbol{\rho}(\mathbf{z}-\mathbf{x}^{\mathrm{C}}(t_f,\tilde{\mathbf{p}}))+\boldsymbol{\rho}_0(\mathbf{p}-\tilde{\mathbf{p}}),$$

and define $\tilde{c}:\mathbb{R}^{n_p}\to\mathbb{R}$ so that for each $\mathbf{p}\in\mathbb{R}^{n_p}$,

$$\tilde{c}(\mathbf{p})=\tilde{g}(\mathbf{p},\mathbf{y}(t_f,\mathbf{p})).$$

Since $\mathbf{y}(t_f,\cdot)$ is differentiable and $\tilde{g}$ is differentiable by construction, $\tilde{c}$ is differentiable and

$$\frac{\mathrm{d}\tilde{c}}{\mathrm{d}\mathbf{p}}(\tilde{\mathbf{p}})=\boldsymbol{\rho}_0+\boldsymbol{\rho}\frac{\partial\mathbf{y}}{\partial\mathbf{p}}(t_f,\tilde{\mathbf{p}}).$$

Since (6.4.31) and (6.4.37) hold,

$$\frac{\mathrm{d}\tilde{c}}{\mathrm{d}\mathbf{p}}(\tilde{\mathbf{p}})\equiv\tilde{\mathbf{s}}=\boldsymbol{\rho}_0+\boldsymbol{\rho}\frac{\partial\mathbf{y}}{\partial\mathbf{p}}(t_f,\tilde{\mathbf{p}}). \tag{6.4.38}$$

Now, we show that $\frac{\mathrm{d}\tilde{c}}{\mathrm{d}\mathbf{p}}(\tilde{\mathbf{p}})$ can be evaluated using adjoints. The remainder of this proof is similar to the proof for adjoint sensitivity analysis for differentiable dynamic systems (c.f. [90, Section 2.1]), but with essential differences, since $\mathbf{y}(t,\tilde{\mathbf{p}})$ is not differentiable everywhere in t.

Consider an absolutely continuous function $\boldsymbol{\lambda}:I\to\mathbb{R}^{2n_x}$. It follows that for each $\mathbf{p}\in\mathbb{R}^{n_p}$,

$$\begin{aligned}\tilde{c}(\mathbf{p})&=\tilde{g}(\mathbf{p},\mathbf{y}(t_f,\mathbf{p}))\\&=\boldsymbol{\rho}(\mathbf{y}(t_f,\mathbf{p})-\mathbf{x}^{\mathrm{C}}(t_f,\tilde{\mathbf{p}}))+\boldsymbol{\rho}_0(\mathbf{p}-\tilde{\mathbf{p}})\\&\quad-\int_{t_0}^{t_f}(\boldsymbol{\lambda}(t))^{\mathrm{T}}\dot{\mathbf{y}}(t,\mathbf{p})\mathrm{d}t+\int_{t_0}^{t_f}(\boldsymbol{\lambda}(t))^{\mathrm{T}}\boldsymbol{\gamma}(t,\mathbf{p},\mathbf{y}(t,\mathbf{p}))\mathrm{d}t.\end{aligned} \tag{6.4.39}$$

Note that the integral terms above are understood as Lebesgue integrals. Moreover, [147, Corollary 5.4.3] implies that

$$\int_{t_0}^{t_f} (\boldsymbol{\lambda}(t))^{\mathrm{T}} \dot{\mathbf{y}}(t,\mathbf{p})\mathrm{d}t = (\boldsymbol{\lambda}(t_f))^{\mathrm{T}} \mathbf{y}(t_f,\mathbf{p}) - (\boldsymbol{\lambda}(t_0))^{\mathrm{T}} \mathbf{y}(t_0,\mathbf{p}) - \int_{t_0}^{t_f} (\dot{\boldsymbol{\lambda}}(t))^{\mathrm{T}} \mathbf{y}(t,\mathbf{p})\mathrm{d}t. \quad (6.4.40)$$

By construction, all integrands in (6.4.40) and (6.4.39) are differentiable with respect to $\mathbf{p} \in \mathbb{R}^{n_p}$. According to [147, Corollary 2.8.7], plugging (6.4.40) into (6.4.39) and differentiating at $\mathbf{p} := \tilde{\mathbf{p}}$ yields

$$\begin{aligned}\frac{\mathrm{d}\tilde{c}}{\mathrm{d}\mathbf{p}}(\tilde{\mathbf{p}}) &= \boldsymbol{\rho}\frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t_f,\tilde{\mathbf{p}}) + \boldsymbol{\rho}_0 - (\boldsymbol{\lambda}(t_f))^{\mathrm{T}}\frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t_f,\tilde{\mathbf{p}}) + (\boldsymbol{\lambda}(t_0))^{\mathrm{T}}\frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t_0,\tilde{\mathbf{p}}) \\ &\quad + \int_{t_0}^{t_f} (\dot{\boldsymbol{\lambda}}(t))^{\mathrm{T}}\frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t,\tilde{\mathbf{p}})\mathrm{d}t + \int_{t_0}^{t_f} (\boldsymbol{\lambda}(t))^{\mathrm{T}}\Big(\frac{\partial \boldsymbol{\gamma}}{\partial \mathbf{p}}(t,\tilde{\mathbf{p}},\mathbf{y}(t,\tilde{\mathbf{p}})) + \frac{\partial \boldsymbol{\gamma}}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t,\tilde{\mathbf{p}},\mathbf{y}(t,\tilde{\mathbf{p}}))\Big)\mathrm{d}t.\end{aligned}$$

By construction of $\boldsymbol{\gamma}$ and the initial values in (6.4.36), it follows that

$$\begin{aligned}\frac{\mathrm{d}\tilde{c}}{\mathrm{d}\mathbf{p}}(\tilde{\mathbf{p}}) &= \boldsymbol{\rho}\frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t_f,\tilde{\mathbf{p}}) + \boldsymbol{\rho}_0 - (\boldsymbol{\lambda}(t_f))^{\mathrm{T}}\frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t_f,\tilde{\mathbf{p}}) + (\boldsymbol{\lambda}(t_0))^{\mathrm{T}}\mathbf{S}_0^{\mathrm{C}}(\tilde{\mathbf{p}}) \\ &\quad + \int_{t_0}^{t_f} (\dot{\boldsymbol{\lambda}}(t))^{\mathrm{T}}\frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t,\tilde{\mathbf{p}})\mathrm{d}t + \int_{t_0}^{t_f} (\boldsymbol{\lambda}(t))^{\mathrm{T}}\Big(\boldsymbol{\Theta}^{\mathrm{B}}(t,\tilde{\mathbf{p}}) + \boldsymbol{\Theta}^{\mathrm{A}}(t,\tilde{\mathbf{p}})\frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t,\tilde{\mathbf{p}})\Big)\mathrm{d}t.\end{aligned} \quad (6.4.41)$$

Now, for a.e. $t \in I$, let

$$\begin{aligned}(\dot{\boldsymbol{\lambda}}(t))^{\mathrm{T}} &= -(\boldsymbol{\lambda}(t))^{\mathrm{T}}\boldsymbol{\Theta}^{\mathrm{A}}(t,\tilde{\mathbf{p}}), \\ (\boldsymbol{\lambda}(t_f))^{\mathrm{T}} &= \boldsymbol{\rho}.\end{aligned} \quad (6.4.42)$$

Then, (6.4.41) reduces to

$$\tilde{\mathbf{s}} \equiv \frac{\mathrm{d}\tilde{c}}{\mathrm{d}\mathbf{p}}(\tilde{\mathbf{p}}) = (\boldsymbol{\lambda}(t_0))^{\mathrm{T}}\mathbf{S}_0^{\mathrm{C}}(\tilde{\mathbf{p}}) + \boldsymbol{\rho}_0 + \int_{t_0}^{t_f} (\boldsymbol{\lambda}(t))^{\mathrm{T}}\boldsymbol{\Theta}^{\mathrm{B}}(t,\tilde{\mathbf{p}})\mathrm{d}t.$$

□

Remark 6.4.14. For each $t \in I$, the integrand $(\boldsymbol{\lambda}(t))^{\mathrm{T}}\boldsymbol{\Theta}^{\mathrm{B}}(t,\tilde{\mathbf{p}})$ in (6.4.34) may be computed based on the states $(\boldsymbol{\lambda}(t))^{\mathrm{T}}$. Thus, $\int_{t_0}^{t_f}(\boldsymbol{\lambda}(t))^{\mathrm{T}}\boldsymbol{\Theta}^{\mathrm{B}}(t,\tilde{\mathbf{p}})\mathrm{d}t$ may be computed simultaneously with numerical integration of (6.4.33).

6.5 New subgradients of established state relaxations

Based on the general subgradient evaluation results established in the previous sections, we now propose new numerical methods for evaluating subgradients of the two established state relaxation methods in the Scott–Barton framework (2.4.1): the Scott–Barton–McCormick (SBM) relaxations [2] and the optimization-based (OB) relaxations [3]. These subgradient evaluation methods assume that the underlying state relaxations do not touch the predefined state bounds; i.e. (2.4.1) reduces to (6.3.2). Roughly, a forward subgradient evaluation ODE system for the SBM relaxations will be constructed from Mitsos et al.'s subgradients [35] of McCormick relaxations [5, 6], and an adjoint ODE sensitivity system for the SBM relaxations will be constructed from Beckers et al.'s method for adjoint mode computation of subgradients for McCormick relaxations. A forward subgradient evaluation system for the OB relaxations will be constructed from a subgradient evaluation method for the multivariate McCormick relaxations [7] proposed in [145].

6.5.1 Subgradients of Scott–Barton–McCormick relaxations

As mentioned in Section 2.4, Scott and Barton [2] construct functions $(\tilde{\mathbf{u}},\tilde{\mathbf{o}})$ in Assumption 2.4.11 using the generalized McCormick (gMC) relaxations [6] of $\mathbf{f}$ in (2.3.1), and then construct appropriate $(\mathbf{u},\mathbf{o})$ in (2.4.1) via (2.4.2). [2, Corollary 1 and Theorem 3] show that by such construction, the unique solution $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$ of (2.4.1) is a valid state relaxation for

(2.3.1). Mitsos et al. [35] propose a forward subgradient evaluation method for the standard McCormick (MC) relaxations, analogously to the forward mode of automatic differentiation [149, 150]. The gMC relaxation method generalizes the standard MC relaxations by allowing previously established relaxations to be inputs. Similarly, the subgradient evaluation rules of Mitsos et al. may also be generalized to evaluate subgradients of the gMC relaxations, by taking in valid subgradients of previously established relaxations. The following proposition formalizes this result for the gMC relaxations of $\mathbf{f}$ in (2.3.1).

Proposition 6.5.1. Denote the generalized McCormick relaxations [6, Definition 15] of $\mathbf{f}$ in (2.3.1) on P as $\tilde{\mathbf{u}}^{\mathrm{gMC}}, \tilde{\mathbf{o}}^{\mathrm{gMC}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$. Encode the subgradient evaluation procedure of [35] for $(\tilde{\mathbf{u}}^{\mathrm{gMC}}, \tilde{\mathbf{o}}^{\mathrm{gMC}})$ as functions $\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}} : I \times \tilde{P} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x \times n_p} \times \mathbb{R}^{n_x \times n_p} \to \mathbb{R}^{n_x \times n_p}$, by finitely composing the addition rule for subgradients in [35, Proposition 2.9], the multiplication rule for subgradients in [35, Theorem 3.3], and the univariate composition rule for subgradients in [35, Theorem 3.2]. Then, Assumption 6.4.5 is satisfied with $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}}) := (\tilde{\mathbf{u}}^{\mathrm{gMC}}, \tilde{\mathbf{o}}^{\mathrm{gMC}})$ and $(\tilde{\mathbf{V}}, \tilde{\mathbf{W}}) := (\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$.

Proof. Firstly, we verify Assumptions 2.4.11 and 6.4.3 that are required in Assumption 6.4.5. According to [2, 6], the functions $(\tilde{\mathbf{u}}^{\mathrm{gMC}}, \tilde{\mathbf{o}}^{\mathrm{gMC}})$ satisfy Assumption 2.4.11 and have the uniform Lispchitz continuity required in Assumption 6.4.3. To establish the directional differentiability requirement of Assumption 6.4.3, observe from [6] that the functions (c.f. [6, Definition 15]) for computing relaxations for each binary addition, binary multiplication, and univariate composition are Lipschitz continuous and directionally differentiable. Then [127, Theorem 3.1.2] implies that these functions are *B-differentiable* (c.f. [127, Section 3.1]). According to [127, Theorem 3.1.1], $(\tilde{\mathbf{u}}^{\mathrm{gMC}}, \tilde{\mathbf{o}}^{\mathrm{gMC}})$ are also B-differentiable since they are finite compositions of B-differentiable functions. This implies that $(\tilde{\mathbf{u}}^{\mathrm{gMC}}, \tilde{\mathbf{o}}^{\mathrm{gMC}})$ are directionally differentiable.

The remaining conditions for $(\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$ except Assumptions 2.4.11 and 6.4.3 in Assumption 6.4.5 can be verified based on [35]. $\square$

The following lemma establishes important bounds and measurability of the constructed $(\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$. These properties guarantee that a subgradient propagation ODE system which will be constructed in Corollary 6.5.3 has solutions.

Lemma 6.5.2. Conditions II.1 and II.2 of Theorem 6.4.8 are satisfied with the substitution $(\tilde{\mathbf{V}}, \tilde{\mathbf{W}}) := (\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$.

Proof. Denote the subgradient evaluation rules [35] for binary addition, binary multiplications, and univariate composition as functions $(\tilde{\mathbf{v}}^{+}, \tilde{\mathbf{w}}^{+})$, $(\tilde{\mathbf{v}}^{\times}, \tilde{\mathbf{w}}^{\times})$, and $(\tilde{\mathbf{v}}^{\mathrm{Uni}}, \tilde{\mathbf{w}}^{\mathrm{Uni}})$, respectively. Let $\mathbf{w}^{\mathrm{cv}} := (w_1^{\mathrm{cv}}, w_2^{\mathrm{cv}})$ and $\mathbf{w}^{\mathrm{cc}} := (w_1^{\mathrm{cc}}, w_2^{\mathrm{cc}})$. Let $\mathbf{s}_{w_1}^{\mathrm{cv}}, \mathbf{s}_{w_2}^{\mathrm{cv}}, \mathbf{s}_{w_3}^{\mathrm{cv}}, \mathbf{s}_{w_1}^{\mathrm{cc}}, \mathbf{s}_{w_2}^{\mathrm{cc}}, \mathbf{s}_{w_3}^{\mathrm{cc}} \in \mathbb{R}^{n_p}$ be arbitrary row vectors. Let

$$\mathbf{S}_{\mathbf{w}}^{\mathrm{cv}} := \begin{bmatrix} \mathbf{s}_{w_1}^{\mathrm{cv}} \\ \mathbf{s}_{w_2}^{\mathrm{cv}} \end{bmatrix} \quad \text{and} \quad \mathbf{S}_{\mathbf{w}}^{\mathrm{cc}} := \begin{bmatrix} \mathbf{s}_{w_1}^{\mathrm{cc}} \\ \mathbf{s}_{w_2}^{\mathrm{cc}}, \end{bmatrix}$$

Observe from [35] that the subgradient evaluation functions $(\tilde{\mathbf{v}}^{+}, \tilde{\mathbf{v}}^{\times}, \tilde{\mathbf{v}}^{\mathrm{Uni}})$ have the following affine forms:

$$\tilde{\mathbf{v}}^{+} : (t, \mathbf{w}^{\mathrm{cv}}, \mathbf{w}^{\mathrm{cc}}, \mathbf{S}_{\mathbf{w}}^{\mathrm{cv}}, \mathbf{S}_{\mathbf{w}}^{\mathrm{cc}}) \mapsto \tilde{\mathbf{\Theta}}^{+}(t, \mathbf{w}^{\mathrm{cv}}, \mathbf{w}^{\mathrm{cc}}) \begin{bmatrix} \mathbf{S}_{\mathbf{w}}^{\mathrm{cv}} \\ \mathbf{S}_{\mathbf{w}}^{\mathrm{cc}} \end{bmatrix},$$

$$\tilde{\mathbf{v}}^{\times} : (t, \mathbf{w}^{\mathrm{cv}}, \mathbf{w}^{\mathrm{cc}}, \mathbf{S}_{\mathbf{w}}^{\mathrm{cv}}, \mathbf{S}_{\mathbf{w}}^{\mathrm{cc}}) \mapsto \tilde{\mathbf{\Theta}}^{\times}(t, \mathbf{w}^{\mathrm{cv}}, \mathbf{w}^{\mathrm{cc}}) \begin{bmatrix} \mathbf{S}_{\mathbf{w}}^{\mathrm{cv}} \\ \mathbf{S}_{\mathbf{w}}^{\mathrm{cc}} \end{bmatrix},$$

$$\text{and} \quad \tilde{\mathbf{v}}^{\mathrm{Uni}} : (t, w_3^{\mathrm{cv}}, w_3^{\mathrm{cc}}, \mathbf{s}_{w_3}^{\mathrm{cv}}, \mathbf{s}_{w_3}^{\mathrm{cc}}) \mapsto \tilde{\mathbf{\Theta}}^{\mathrm{Uni}}(t, w_3^{\mathrm{cv}}, w_3^{\mathrm{cc}}) \begin{bmatrix} \mathbf{s}_{w_3}^{\mathrm{cv}} \\ \mathbf{s}_{w_3}^{\mathrm{cc}} \end{bmatrix},$$

where $\tilde{\mathbf{\Theta}}^{+}, \tilde{\mathbf{\Theta}}^{\times} : I \times \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^{1\times 4}$ and $\tilde{\mathbf{\Theta}}^{\mathrm{Uni}} : I \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^{1\times 2}$ are coefficient matrix-valued functions. Also observe that the functions $(\tilde{\mathbf{\Theta}}^{+}, \tilde{\mathbf{\Theta}}^{\times}, \tilde{\mathbf{\Theta}}^{\mathrm{Uni}})$ are piecewise continuous in the sense of Filippov [102], and thus these functions are bounded on any bounded domain, and are Borel measurable. Similarly, the functions $(\tilde{\mathbf{w}}^{+}, \tilde{\mathbf{w}}^{\times}, \tilde{\mathbf{w}}^{\mathrm{Uni}})$ also have similar affine structure and the corresponding coefficient matrix functions are bounded and Borel measurable. Since $(\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$ are constructed by finitely composing these functions $(\tilde{\mathbf{v}}^{+}, \tilde{\mathbf{v}}^{\times}, \tilde{\mathbf{v}}^{\mathrm{Uni}}, \tilde{\mathbf{w}}^{+}, \tilde{\mathbf{w}}^{\times}, \tilde{\mathbf{w}}^{\mathrm{Uni}})$, they have the following affine forms:

$$\begin{aligned} \tilde{\mathbf{V}}^{\mathrm{gMC}} &: (t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}, \mathbf{M}, \mathbf{N}) \mapsto \tilde{\mathbf{\Theta}}^{\mathrm{cv,A}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix} + \tilde{\mathbf{\Theta}}^{\mathrm{cv,B}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \\ \text{and} \quad \tilde{\mathbf{W}}^{\mathrm{gMC}} &: (t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}, \mathbf{M}, \mathbf{N}) \mapsto \tilde{\mathbf{\Theta}}^{\mathrm{cc,A}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix} + \tilde{\mathbf{\Theta}}^{\mathrm{cc,B}}(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}), \end{aligned} \tag{6.5.1}$$

where $\tilde{\mathbf{\Theta}}^{\mathrm{cv,A}}, \tilde{\mathbf{\Theta}}^{\mathrm{cc,A}}, \tilde{\mathbf{\Theta}}^{\mathrm{cv,B}}, \tilde{\mathbf{\Theta}}^{\mathrm{cc,B}} : I \times \tilde{P} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x \times 2n_x}$. Since $(\tilde{\mathbf{\Theta}}^{+}, \tilde{\mathbf{\Theta}}^{\times}, \tilde{\mathbf{\Theta}}^{\mathrm{Uni}})$ are bounded and Borel measurable, $(\tilde{\mathbf{\Theta}}^{\mathrm{cv,A}}, \tilde{\mathbf{\Theta}}^{\mathrm{cc,A}}, \tilde{\mathbf{\Theta}}^{\mathrm{cv,B}}, \tilde{\mathbf{\Theta}}^{\mathrm{cc,B}})$ are bounded on the bounded domain $I \times \tilde{P} \times \tilde{X} \times \tilde{X}$, and are Borel measurable according to [147, 2.12(ii) and Theorem 2.1.5]. Thus, $(\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$ in (6.5.1) are Borel measurable, which indicates that Condition II.1 is satisfied. Moreover, since $Z \in \mathbb{IR}^{n_x \times n_p}$ is also bounded, $(\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$ satisfy Condition II.2. $\square$

The following corollary shows that we may combine $(\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$ above with the flattening operations in Definitions 2.4.3 and 6.4.7 to construct valid subgradient propagation functions $(\mathbf{V}, \mathbf{W})$ for $(\tilde{\mathbf{u}}^{\mathrm{gMC}}, \tilde{\mathbf{o}}^{\mathrm{gMC}})$ in (6.4.1). Then, the unique solution $(\mathbf{S}^{cv}, \mathbf{S}^{cc})$ of (6.4.1) is guaranteed to comprise valid subgradients of the SBM relaxations obtained using (6.3.2).

Corollary 6.5.3. Consider functions $(\tilde{\mathbf{u}}^{\mathrm{gMC}}, \tilde{\mathbf{o}}^{\mathrm{gMC}}, \tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$ as in Proposition 6.5.1. Define functions $\mathbf{u}, \mathbf{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ using (2.4.2) with $(\tilde{\mathbf{u}}, \tilde{\mathbf{o}}) := (\tilde{\mathbf{u}}^{\mathrm{gMC}}, \tilde{\mathbf{o}}^{\mathrm{gMC}})$. Then, the mappings $\mathbf{u}(t, \cdot, \cdot, \cdot)$ and $\mathbf{o}(t, \cdot, \cdot, \cdot)$ are directionally differentiable and Lipschitz continuous on $\tilde{P} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly over $t \in I$. Suppose that the remaining conditions of Assumption 6.3.2 also hold. Define functions $\mathbf{V}, \mathbf{W} : I \times \tilde{P} \times \mathbb{R}^{n_x \times n_p} \times \mathbb{R}^{n_x \times n_p} \to \mathbb{R}^{n_x \times n_p}$ using (6.4.22) with $(\tilde{\mathbf{V}}, \tilde{\mathbf{W}}) := (\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$. Then, (6.4.1) has local existence and uniqueness of solutions, and the unique solution $(\mathbf{S}^{\mathrm{cv}}, \mathbf{S}^{\mathrm{cc}})$ comprises valid state relaxation subgradients of $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$.

Proof. Since Proposition 6.5.1 and Lemma 6.5.2 hold, Theorem 6.4.8 implies that $(\mathbf{V}, \mathbf{W})$ are valid subgradient propagation functions for $(\mathbf{u}, \mathbf{o})$. Then, the claimed result follows from Theorem 6.4.2.

□

Remark 6.5.4. Since the functions $(\tilde{\mathbf{V}}^{\mathrm{gMC}}, \tilde{\mathbf{W}}^{\mathrm{gMC}})$ have the affine forms as in (6.5.1), and since $(\mathbf{R}^{i,\mathrm{L}}, \mathbf{R}^{i,\mathrm{U}})$ are affine functions, the functions $(\mathbf{V}, \mathbf{W})$ in the corollary above also have similar affine forms to (6.5.1), as follows. For each $i \in \{1, \ldots, n_x\}$, define identity matrices $\Phi^i, \Psi^i \in \mathbb{R}^{2n_x \times 2n_x}$, and then replace the $(n_x + i)^{\mathrm{th}}$ row of $\mathbf{\Phi}^i$ by $[\underbrace{0, \ldots, 0}_{i-1}, 1, \underbrace{0, \ldots, 0}_{2n_x - i}]$, and replace the i^{th} row of $\mathbf{\Psi}^i$ by $[\underbrace{0, \ldots, 0}_{n_x + i - 1}, 1, \underbrace{0, \ldots, 0}_{n_x - i}]$. Then, for each $i \in \{1, \ldots, n_x\}$, observe that $\mathbf{v}_{(i)}(t,\mathbf{p},\mathbf{M},\mathbf{N})$ and $\mathbf{w}_{(i)}(t,\mathbf{p},\mathbf{M},\mathbf{N})$ are equivalent to the following:

$$\begin{aligned}
\mathbf{v}_{(i)}(t,\mathbf{p},\mathbf{M},\mathbf{N}) &\equiv \tilde{\boldsymbol{\theta}}_{(i)}^{\mathrm{cv,A}}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\mathbf{p})))\boldsymbol{\Phi}^{i}\begin{bmatrix}\mathbf{M}\\ \mathbf{N}\end{bmatrix} \\
&\quad + \tilde{\boldsymbol{\theta}}_{(i)}^{\mathrm{cv,B}}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\mathbf{p}))), \\
\text{and}\quad \mathbf{w}_{(i)}(t,\mathbf{p},\mathbf{M},\mathbf{N}) &\equiv \tilde{\boldsymbol{\theta}}_{(i)}^{\mathrm{cc,A}}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{U}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\mathbf{p})))\boldsymbol{\Psi}^{i}\begin{bmatrix}\mathbf{M}\\ \mathbf{N}\end{bmatrix} \\
&\quad + \tilde{\boldsymbol{\theta}}_{(i)}^{\mathrm{cc,B}}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{U}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\mathbf{p}))).
\end{aligned} \tag{6.5.2}$$

Thus, the constructed subgradient propagation system (6.4.1) is actually an affine parametric ODE system, which may be easily integrated by off-the-shelf numerical ODE solvers.

Corollary 6.5.3 and Remark 6.5.4 above show that subgradients of the SBM relaxations may be computed by solving a forward affine parametric ODE system constructed from Mitsos et al.'s vector forward mode subgradient automatic differentiation [35]. Beckers et al. [144] propose a method for adjoint mode computation of subgradients for McCormick relaxations, which is empirically more efficient than Mitsos et al.'s subgradient evaluation method for a large number of parameters. Based on these results, analogously to the adjoint sensitivity analysis for smooth dynamic systems [90], the following corollary and remark propose a new adjoint subgradient evaluation method for the objective function c^{cv} of the lower bounding problem (6.4.30) in global dynamic optimization. This method may reduce the computational effort required to evaluate the subgradients required by nonsmooth optimizers for minimizing c^{cv}, and may thus speed up an overarching global dynamic optimization method.

Corollary 6.5.5. For each $\mathbf{p} \in P$, consider computing the objective value $c^{\mathrm{cv}}(\mathbf{p})$ of (6.4.30) by first constructing the SBM relaxations [2] $\mathbf{x}^{\mathrm{C}}(t_f,\mathbf{p}) := (\mathbf{x}^{\mathrm{cv}}(t_f,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t_f,\mathbf{p}))$ for which $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$ satisfy (6.3.1), and then evaluating $g^{\mathrm{cv}}(t_f,\mathbf{p},\mathbf{x}^{\mathrm{C}}(t_f,\mathbf{p}))$ using the generalized McCormick relaxation method [6]. Consider the functions and quantities in (6.5.2), and define functions $\mathbf{\Theta}^{\mathrm{A}},\mathbf{\Theta}^{\mathrm{B}} : I\times\tilde{P} := \mathbb{R}^{2n_x}\times\mathbb{R}^{2n_x}$ so that for each $i\in\{1,...,n_x\}$ and $(t,\mathbf{p})\in I\times\tilde{P}$,

$$\begin{aligned}
\boldsymbol{\theta}^{\mathrm{A}}_{(i)}(t,\mathbf{p}) &:= \tilde{\boldsymbol{\theta}}^{\mathrm{cv,A}}_{(i)}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\mathbf{p})))\mathbf{\Phi}^i,\\
\boldsymbol{\theta}^{\mathrm{A}}_{(n_x+i)}(t,\mathbf{p}) &:= \tilde{\boldsymbol{\theta}}^{\mathrm{cc,A}}_{(i)}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{U}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\mathbf{p})))\mathbf{\Psi}^i,\\
\boldsymbol{\theta}^{\mathrm{B}}_{(i)}(t,\mathbf{p}) &:= \tilde{\boldsymbol{\theta}}^{\mathrm{cv,B}}_{(i)}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\mathbf{p}))),\\
\text{and}\quad \boldsymbol{\theta}^{\mathrm{B}}_{(n_x+i)}(t,\mathbf{p}) &:= \tilde{\boldsymbol{\theta}}^{\mathrm{cc,B}}_{(i)}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{U}}(\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\boldsymbol{\xi}^{\mathrm{cc}}(t,\mathbf{p}))).
\end{aligned}$$

Consider any fixed $\tilde{\mathbf{p}}\in\tilde{P}$. Let $(\boldsymbol{\rho}_0,\boldsymbol{\rho})\in\mathbb{R}^{n_p}\times\mathbb{R}^{2n_x}$ be the transpose of a subgradient of $g^{\mathrm{cv}}(t_f,\cdot,\cdot)$ at $(\tilde{\mathbf{p}},\mathbf{x}^{\mathrm{C}}(t,\tilde{\mathbf{p}}))$. Let $\mathbf{S}^{\mathrm{C}}_0 := (\mathbf{S}^{\mathrm{cv}}_0,\mathbf{S}^{\mathrm{cc}}_0)$ as in (6.4.1). Let $\boldsymbol{\lambda} : I\to\mathbb{R}^{2n_x}$ be a Carathéodory solution (c.f. [102]) of the following reverse ODE system on I:

$$\begin{aligned}
(\dot{\boldsymbol{\lambda}}(t))^{\mathrm{T}} &= -(\boldsymbol{\lambda}(t))^{\mathrm{T}}\mathbf{\Theta}^{\mathrm{A}}(t,\tilde{\mathbf{p}}),\\
(\boldsymbol{\lambda}(t_f))^{\mathrm{T}} &= \boldsymbol{\rho}.
\end{aligned}\tag{6.5.3}$$

Then, the quantity (equation (6.5.4) below) is the transpose of a subgradient of c^{cv} at $\tilde{\mathbf{p}}$:

$$\tilde{\mathbf{s}} := (\boldsymbol{\lambda}(t_0))^{\mathrm{T}}\mathbf{S}^{\mathrm{C}}_0(\tilde{\mathbf{p}})+\boldsymbol{\rho}_0+\int_{t_0}^{t_f}(\boldsymbol{\lambda}(t))^{\mathrm{T}}\mathbf{\Theta}^{\mathrm{B}}(t,\tilde{\mathbf{p}})\mathrm{d}t.\tag{6.5.4}$$

Proof. Since g^{cv} is constructed using the generalized McCormick relaxation method, Condition III.1 of Theorem 6.4.13 is satisfied when using the subgradient propagation method in [35] for computing $\tilde{\mathbf{s}}$ based on $(\boldsymbol{\rho}_0,\boldsymbol{\rho},\mathbf{S}^{\mathrm{C}}(t_f,\tilde{\mathbf{p}}))$. According to the discussion in Remark 6.5.4, Condition III.2 is also satisfied when using the dynamic forward subgradient propagation system constructed in Corollary 6.5.3 for computing $\mathbf{S}^{\mathrm{C}}(t_f,\tilde{\mathbf{p}})$. Then, the claimed results follow from Theorem 6.4.13. $\square$

Remark 6.5.6. In the corollary above, the subgradients $(\boldsymbol{\rho}_0, \boldsymbol{\rho})$ and the term $(\boldsymbol{\lambda}(t))^{\mathrm{T}}\boldsymbol{\Theta}^{\mathrm{B}}(t, \tilde{\mathbf{p}})$ may be computed efficiently using an established method for adjoint mode subgradient computation [144] for McCormick relaxations. The term $(\boldsymbol{\lambda}(t))^{\mathrm{T}}\boldsymbol{\Theta}^{\mathrm{A}}(t, \tilde{\mathbf{p}})$ may be computed efficiently by an easy modification of this adjoint subgradient computation method, which incorporates the operations $(\boldsymbol{\Phi}_i, \boldsymbol{\Psi}_i)$.

6.5.2 Subgradients of optimization-based relaxations

This subsection first summarizes our recent OB state relaxations [3], and then presents a new dynamic forward subgradient evaluation method for the relaxations.

Formulation of optimization-based relaxations

Define a function $\mathbf{v} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ so that for all $i \in \{1, ..., n_x\}$ and $\boldsymbol{\alpha}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}} \in \mathbb{R}^{n_x}$,

$$v_i(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) := \tfrac{1}{2}[(\alpha_i + 1)\xi_i^{\mathrm{cc}} - (\alpha_i - 1)\xi_i^{\mathrm{cv}}]. \tag{6.5.5}$$

Consider functions $\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}} : I \times P \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ that satisfy [3, Assumption 3]. Roughly, for each $t \in I$, the mappings $\mathbf{f}^{\mathrm{cv}}(t, \cdot, \cdot)$ and $\mathbf{f}^{\mathrm{cc}}(t, \cdot, \cdot)$ are, respectively, convex and concave relaxations of $\mathbf{f}(t, \cdot, \cdot)$ in (2.3.1) on $P \times X(t)$. [3] construct the following $\mathbf{u}, \mathbf{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ for use in (2.4.1): for each $i \in \{1, ..., n_x\}$ and $(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) \in I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$,

$$\begin{aligned} u_i(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) &:= \min_{\boldsymbol{\alpha} \in [-1,1]^{n_x}} f_i^{\mathrm{cv}}(t, \mathbf{p}, \mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})) \text{ subject to } \alpha_i = -1, \\ \text{and} \quad o_i(t, \mathbf{p}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}}) &:= \max_{\boldsymbol{\alpha} \in [-1,1]^{n_x}} f_i^{\mathrm{cc}}(t, \mathbf{p}, \mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\xi}^{\mathrm{cv}}, \boldsymbol{\xi}^{\mathrm{cc}})) \text{ subject to } \alpha_i = +1. \end{aligned} \tag{6.5.6}$$

Then, [3, Lemmata 1 and 2] show that these functions $(\mathbf{u},\mathbf{o})$ are valid Scott–Barton right-hand side functions, and [3, Theorems 3 and 4] show that the unique solution $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$ of (2.4.1) comprises valid state relaxations for (2.3.1). Now, define functions $\tilde{\mathbf{u}}^{\mathrm{OB}},\tilde{\mathbf{o}}^{\mathrm{OB}}: I\times P\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$ so that for each $i\in\{1,...,n_x\}$ and $(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})\in I\times P\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$,

$$\begin{aligned}
&\tilde{u}_i^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \min_{\boldsymbol{\alpha}\in[-1,1]^{n_x}} f_i^{\mathrm{cv}}(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}))\\
\text{and}\quad &\tilde{o}_i^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) := \max_{\boldsymbol{\alpha}\in[-1,1]^{n_x}} f_i^{\mathrm{cc}}(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})).
\end{aligned}\tag{6.5.7}$$

Observe that

$$\begin{aligned}
&u_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \tilde{u}_i^{\mathrm{OB}}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}))\\
\text{and}\quad &o_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \tilde{o}_i^{\mathrm{OB}}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{U}}(\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})).
\end{aligned}\tag{6.5.8}$$

On a set $S:=\{(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})\in I\times P\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}:\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}\in X(t)\text{ and }\boldsymbol{\xi}^{\mathrm{cv}}\leq\boldsymbol{\xi}^{\mathrm{cc}}\}$, the functions $(\tilde{u}_i^{\mathrm{OB}},\tilde{o}_i^{\mathrm{OB}})$ in (6.5.7) reduce to

$$\begin{aligned}
&\tilde{u}_i^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \min_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi}),\\
\text{and}\quad &\tilde{o}_i^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \max_{\boldsymbol{\xi}\in[\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}]} f_i^{\mathrm{cc}}(t,\mathbf{p},\boldsymbol{\xi}).
\end{aligned}\tag{6.5.9}$$

Dynamic subgradient evaluation method

Lemma 6.5.8 below shows that under mild assumptions on $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$, the functions $(\tilde{\mathbf{u}}^{\mathrm{OB}},\tilde{\mathbf{o}}^{\mathrm{OB}})$ in (6.5.7) satisfy Assumption 6.4.3.

Assumption 6.5.7. Suppose that functions $\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}}: I\times P\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$ satisfy the following conditions:

IV.1 The mappings $\mathbf{f}^{\mathrm{cv}}(t,\cdot,\cdot)$ and $\mathbf{f}^{\mathrm{cc}}(t,\cdot,\cdot)$ are Lipschitz continuous on $\tilde{P}\times\mathbb{R}^{n_x}$, uniformly over $t\in I$.

IV.2 There exist functions $\mathbf{f}^{\mathrm{cv},1},\mathbf{f}^{\mathrm{cv},2},...,\mathbf{f}^{\mathrm{cv},k},\mathbf{f}^{\mathrm{cc},1},\mathbf{f}^{\mathrm{cc},2},...,\mathbf{f}^{\mathrm{cc},\kappa}: I\times P\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$ so that for each $t\in I$, each $\mathbf{f}^{\mathrm{cv},j}(t,\cdot,\cdot)$ and $\mathbf{f}^{\mathrm{cc},j}(t,\cdot,\cdot)$ is continuously differentiable, and is respectively convex and concave on $P\times\mathbb{R}^{n_x}$. Moreover, for each $i\in\{1,...,n_x\}$ and $(t,\mathbf{p},\boldsymbol{\xi})\in I\times P\times\mathbb{R}^{n_x}$,

$$\begin{aligned} f_i^{\mathrm{cv}}(t,\mathbf{p},\boldsymbol{\xi}) &\equiv \max(f_i^{\mathrm{cv},1}(t,\mathbf{p},\boldsymbol{\xi}),f_i^{\mathrm{cv},2}(t,\mathbf{p},\boldsymbol{\xi}),...,f_i^{\mathrm{cv},k}(t,\mathbf{p},\boldsymbol{\xi}))\\ \text{and}\quad f_i^{\mathrm{cc},}(t,\mathbf{p},\boldsymbol{\xi}) &\equiv \min(f_i^{\mathrm{cc},1}(t,\mathbf{p},\boldsymbol{\xi}),f_i^{\mathrm{cc},2}(t,\mathbf{p},\boldsymbol{\xi}),...,f_i^{\mathrm{cc},\kappa}(t,\mathbf{p},\boldsymbol{\xi})). \end{aligned}$$

Lemma 6.5.8. Consider functions $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$ that satisfy [3, Assumption 3] and Assumption 6.5.7. Then, Assumptions 2.4.11 and 6.4.3 are satisfied with $(\tilde{\mathbf{u}},\tilde{\mathbf{o}}):=(\tilde{\mathbf{u}}^{\mathrm{OB}},\tilde{\mathbf{o}}^{\mathrm{OB}})$ in (6.5.7).

Proof. Since Condition IV.1 holds, by an analogous argument to the proof of [3, Proposition 2], $(\tilde{\mathbf{u}}^{\mathrm{OB}},\tilde{\mathbf{o}}^{\mathrm{OB}})$ have the uniform Lipschitz continuity required in Assumption 6.4.3. Since (6.5.9) holds, [7, Theorem 2] implies that $(\tilde{\mathbf{u}}^{\mathrm{OB}},\tilde{\mathbf{o}}^{\mathrm{OB}})$ satisfy Assumption 2.4.11.

Now, we prove the directional differentiability in Assumption 6.4.3. Consider any $(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})\in I\times P\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$. Since Condition IV.2 holds, we may reformulate $\tilde{u}_i^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})$ as

$$\begin{aligned} \tilde{u}_i^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}) \equiv \min_{\gamma\in\mathbb{R},\boldsymbol{\alpha}\in[-1,1]^{n_x}}\ & \gamma\\ \text{subject to}\quad & f_i^{\mathrm{cv},m}(t,\mathbf{p},\mathbf{v}(\boldsymbol{\alpha},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}))\leq\gamma,\quad \forall m\in\{1,...,k\}. \end{aligned} \tag{6.5.10}$$

Observe that since $\mathbf{v}$ is affine in $\boldsymbol{\alpha}$ and due to the convexity of $f_i^{\mathrm{cv},m}$, the mapping $f_i^{\mathrm{cv},m}(t,\mathbf{p},\mathbf{v}(\cdot,\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}))$ is convex. Since $\tilde{u}_i^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})$ inherently minimizes a finite convex function on a box domain as in (6.5.7), $\tilde{u}_i^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})$ in the form of (6.5.10) is finite. Observe that the domain of the right-hand side optimization problem in (6.5.10) has nonempty interior, and the optimal solution set is nonempty and bounded. Moreover, since $f_i^{cv,m}(t,\cdot,\cdot)$ is assumed to be continuously differentiable, [151, Theorem 2] implies that $\tilde{u}_i^{\mathrm{OB}}(t,\cdot,\cdot,\cdot)$ is directionally differentiable. By a similar argument, $\tilde{o}_i^{\mathrm{OB}}(t,\cdot,\cdot,\cdot)$ is also directionally differentiable. □

Remark 6.5.9. Appropriate functions $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$ in the lemma above may be constructed by first constructing Lipschitz continuous and piecewise differentiable (in the sense of e.g. Scholtes [127]) relaxations of $\mathbf{f}(t,\cdot,\cdot)$ on $P\times X(t)$, such as McCormick relaxations [5] and αBB relaxations [9]. Then, an appropriate convex extension ([129, Proposition 3.1.4]) of such relaxations from $P\times X(t)$ to $P\times\mathbb{R}^{n_x}$ may be applied.

In concert with $(\tilde{\mathbf{u}}^{\mathrm{OB}},\tilde{\mathbf{o}}^{\mathrm{OB}})$, the following proposition constructs new functions $(\tilde{\mathbf{V}}^{\mathrm{OB}},\tilde{\mathbf{W}}^{\mathrm{OB}})$ that satisfy Assumption 6.4.5, based on a recently established subgradient evaluation method [145, Theorem 5.3.2] for multivariate McCormick relaxations (mMC) [7]. The mMC relaxation method computes convex and concave relaxations for multivariate composite functions by solving convex optimization problems. Unlike the original subgradient evaluation method for mMC relaxations that solves dual NLPs proposed in [7, Theorem 4], once the optimization problems for computing the mMC relaxations are solved, [145, Theorem 5.3.2] evaluates the subgradients in closed form, and no NLPs need to be solved.

Proposition 6.5.10. Consider functions $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$ that satisfy [3, Assumption 3] and the functions $(\tilde{\mathbf{u}}^{\mathrm{OB}},\tilde{\mathbf{o}}^{\mathrm{OB}})$ defined in (6.5.7). For any $\sigma\in\mathbb{R}$, let σ^+ denote $\sigma^+=\max(0,\sigma)$, and let σ^- denote $\sigma^-=\min(0,\sigma)$. Define functions $\tilde{\mathbf{V}}^{\mathrm{OB}},\tilde{\mathbf{W}}^{\mathrm{OB}}: I\times\tilde{P}\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x\times n_p}\times\mathbb{R}^{n_x\times n_p}\to\mathbb{R}^{n_x\times n_p}$ so that for each $i\in\{1,...,n_x\}$, $t\in I$, $\mathbf{p}\in\tilde{P}$, $\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}\in X(t)$ for which $\boldsymbol{\xi}^{\mathrm{cv}}\leq\boldsymbol{\xi}^{\mathrm{cc}}$, and $\mathbf{M},\mathbf{N}\in\mathbb{R}^{n_x\times n_p}$, $\tilde{\mathbf{v}}_{(i)}^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M},\mathbf{N})$ and $\tilde{\mathbf{w}}_{(i)}^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M},\mathbf{N})$ are defined by the following procedure:

1. Compute an optimal solution $\boldsymbol{\xi}^{*,\tilde{u}}$ of the optimization problem for defining $\tilde{u}_i^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})$ in (6.5.9), and compute an optimal solution $\boldsymbol{\xi}^{*,\tilde{o}}$ of the optimization problem for defining $\tilde{o}_i^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})$ in (6.5.9).

2. Compute a subgradient $(\mathbf{s}^{\mathbf{p},\tilde{u}},\mathbf{s}^{*,\tilde{u}}) \in \mathbb{R}^{n_p}\times\mathbb{R}^{n_x}$ of $f_i^{\mathrm{cv}}(t,\cdot,\cdot)$ at $(\mathbf{p},\boldsymbol{\xi}^{*,\tilde{u}})$, and compute a subgradient $(\mathbf{s}^{\mathbf{p},\tilde{o}},\mathbf{s}^{*,\tilde{o}}) \in \mathbb{R}^{n_p}\times\mathbb{R}^{n_x}$ of $f_i^{\mathrm{cc}}(t,\cdot,\cdot)$ at $(\mathbf{p},\boldsymbol{\xi}^{*,\tilde{o}})$.

3. Set

$$\tilde{\mathbf{v}}_{(i)}^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M},\mathbf{N}) := \mathbf{s}^{\mathbf{p},\tilde{u}} + \sum_{j=1}^{n_x}\Big([s_j^{*,\tilde{u}}]^{+}\mathbf{m}_{(j)} + [s_j^{*,\tilde{u}}]^{-}\mathbf{n}_{(j)}\Big)$$
$$\text{and}\quad \tilde{\mathbf{w}}_{(i)}^{\mathrm{OB}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M},\mathbf{N}) := \mathbf{s}^{\mathbf{p},\tilde{o}} + \sum_{j=1}^{n_x}\Big([s_j^{*,\tilde{o}}]^{-}\mathbf{m}_{(j)} + [s_j^{*,\tilde{o}}]^{+}\mathbf{n}_{(j)}\Big).$$

Then, Assumption 6.4.5 is satisfied with $(\tilde{\mathbf{u}},\tilde{\mathbf{o}}) := (\tilde{\mathbf{u}}^{\mathrm{OB}},\tilde{\mathbf{o}}^{\mathrm{OB}})$ and $(\tilde{\mathbf{V}},\tilde{\mathbf{W}}) := (\tilde{\mathbf{V}}^{\mathrm{OB}},\tilde{\mathbf{W}}^{\mathrm{OB}})$.

Proof. By Lemma 6.5.8, $(\tilde{\mathbf{u}}^{\mathrm{OB}},\tilde{\mathbf{o}}^{\mathrm{OB}})$ satisfy Assumptions 2.4.11 and 6.4.3. Since $(\tilde{u}_i^{\mathrm{OB}},\tilde{o}_i^{\mathrm{OB}})$ reduce to (6.5.9) on the set S, $(\tilde{\mathbf{V}}^{\mathrm{OB}},\tilde{\mathbf{W}}^{\mathrm{OB}})$ satisfying Assumption 6.4.5 can be verified by applying [145, Theorem 5.3.2] to $\tilde{u}_i^{\mathrm{OB}}(t,\cdot,\cdot,\cdot)$ and $-\tilde{o}_i^{\mathrm{OB}}(t,\cdot,\cdot,\cdot)$ for a.e. $t \in I$ and $i \in \{1,...,n_x\}$. □

The following corollary shows that we may construct valid subgradient propagation functions $(\mathbf{V},\mathbf{W})$ for $(\mathbf{u},\mathbf{o})$ defined in (6.5.6), based on the $(\tilde{\mathbf{V}}^{\mathrm{OB}},\tilde{\mathbf{W}}^{\mathrm{OB}})$ constructed in the previous proposition. Using such $(\mathbf{V},\mathbf{W})$, the subgradient evaluation ODE system (6.4.1) yields valid state relaxation subgradients $(\mathbf{S}^{\mathrm{cv}},\mathbf{S}^{\mathrm{cc}})$ for the OB relaxations $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$. Even though the state relaxation system's right-hand side nominally requires solving convex optimization problems as in (6.5.6), the new subgradient propagation system's right-hand side only employs closed-form functions which can be evaluated efficiently. However, we have to generally assume that these functions have the measurability and bounds required in

Conditions II.1 and II.2 in Theorem 6.4.8. These properties are essential for guaranteeing the local existence and uniqueness of solutions of (6.4.1).

Corollary 6.5.11. Consider functions $(\mathbf{f}^{\mathrm{cv}},\mathbf{f}^{\mathrm{cc}})$ that satisfy [3, Assumption 3] and Assumption 6.5.7, and consider functions $(\mathbf{u},\mathbf{o})$ defined in (6.5.6). Then, the mappings $\mathbf{u}(t,\cdot,\cdot,\cdot)$ and $\mathbf{o}(t,\cdot,\cdot,\cdot)$ are directionally differentiable and Lipschitz continuous on $\tilde{P}\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$, uniformly over $t\in I$. Suppose that the remaining conditions in Assumption 6.3.2 also hold. Consider the notations (σ^+,σ^-) and functions $(\tilde{\mathbf{V}}^{\mathrm{OB}},\tilde{\mathbf{W}}^{\mathrm{OB}})$ as in Proposition 6.5.10, and suppose that Conditions II.1 and II.2 of Theorem 6.4.8 are satisfied with $(\tilde{\mathbf{V}},\tilde{\mathbf{W}}):=(\tilde{\mathbf{V}}^{\mathrm{OB}},\tilde{\mathbf{W}}^{\mathrm{OB}})$. Define functions $\hat{\mathbf{V}},\hat{\mathbf{W}}:I\times\tilde{P}\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x\times n_p}\times\mathbb{R}^{n_x\times n_p}\to\mathbb{R}^{n_x\times n_p}$ so that for each $i\in\{1,\dots,n_x\}$, $t\in I$, $\mathbf{p}\in\tilde{P}$, $\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}\in X(t)$ for which $\boldsymbol{\xi}^{\mathrm{cv}}\leq\boldsymbol{\xi}^{\mathrm{cc}}$, and $\mathbf{M},\mathbf{N}\in\mathbb{R}^{n_x\times n_p}$, $\hat{\mathbf{v}}_{(i)}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M},\mathbf{N})$ and $\hat{\mathbf{w}}_{(i)}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M},\mathbf{N})$ are defined by the following procedure:

1. Compute an optimal solution $\boldsymbol{\xi}^{*,u}$ of the optimization problem for defining $u_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})$ in (6.5.6), and compute an optimal solution $\boldsymbol{\xi}^{*,o}$ of the optimization problem for defining $o_i(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}})$ in (6.5.6).

2. Compute a subgradient $(\mathbf{s}^{\mathbf{p},u},\mathbf{s}^{*,u})\in\mathbb{R}^{n_p}\times\mathbb{R}^{n_x}$ of $f_i^{\mathrm{cv}}(t,\cdot,\cdot)$ at $(\mathbf{p},\boldsymbol{\xi}^{*,u})$, and compute a subgradient $(\mathbf{s}^{\mathbf{p},o},\mathbf{s}^{*,o})\in\mathbb{R}^{n_p}\times\mathbb{R}^{n_x}$ of $f_i^{\mathrm{cc}}(t,\cdot,\cdot)$ at $(\mathbf{p},\boldsymbol{\xi}^{*,o})$.

3. Set

$$\hat{\mathbf{v}}_{(i)}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M},\mathbf{N}) := \mathbf{s}^{\mathbf{p},u} + \sum_{j=1,j\neq i}^{n_x}\Big([s_j^{*,u}]^{+}\mathbf{m}_{(j)} + [s_j^{*,u}]^{-}\mathbf{n}_{(j)}\Big) + ([s_i^{*,u}]^{+} + [s_i^{*,u}]^{-})\mathbf{m}_{(i)}$$

and

$$\hat{\mathbf{w}}_{(i)}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M},\mathbf{N}) := \mathbf{s}^{\mathbf{p},o} + \sum_{j=1,j\neq i}^{n_x}\Big([s_j^{*,o}]^{-}\mathbf{m}_{(j)} + [s_j^{*,o}]^{+}\mathbf{n}_{(j)}\Big) + ([s_i^{*,o}]^{-} + [s_i^{*,o}]^{+})\mathbf{n}_{(i)}.$$

Define functions $\mathbf{V},\mathbf{W} : I \times \tilde{P} \times \mathbb{R}^{n_x\times n_p} \times \mathbb{R}^{n_x\times n_p} \to \mathbb{R}^{n_x\times n_p}$ by setting

$$\mathbf{V}(t,\mathbf{p},\mathbf{M},\mathbf{N}) := \hat{\mathbf{V}}(t,\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}),\mathbf{M},\mathbf{N})$$

and

$$\mathbf{W}(t,\mathbf{p},\mathbf{M},\mathbf{N}) := \hat{\mathbf{W}}(t,\mathbf{p},\mathbf{x}^{\mathrm{cv}}(t,\mathbf{p}),\mathbf{x}^{\mathrm{cc}}(t,\mathbf{p}),\mathbf{M},\mathbf{N}).$$

Then, (6.4.1) has local existence and uniqueness of solutions, and its unique solution $(\mathbf{S}^{\mathrm{cv}},\mathbf{S}^{\mathrm{cc}})$ comprises valid state relaxation subgradients of $(\mathbf{x}^{\mathrm{cv}},\mathbf{x}^{\mathrm{cc}})$.

Proof. Observe that

$$\hat{\mathbf{V}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M},\mathbf{N}) \equiv \tilde{\mathbf{V}}^{\mathrm{OB}}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}),\mathbf{R}^{i,\mathrm{L}}(\mathbf{M},\mathbf{N}))$$

and

$$\hat{\mathbf{W}}(t,\mathbf{p},\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}},\mathbf{M},\mathbf{N}) \equiv \tilde{\mathbf{W}}^{\mathrm{OB}}(t,\mathbf{p},\mathbf{r}^{i,\mathrm{L}}(\boldsymbol{\xi}^{\mathrm{cv}},\boldsymbol{\xi}^{\mathrm{cc}}),\mathbf{R}^{i,\mathrm{L}}(\mathbf{M},\mathbf{N})).$$

Since (6.5.8) holds, Lemma 6.5.8, Proposition 6.5.10, and Theorem 6.4.8 imply that $(\mathbf{V},\mathbf{W})$ are valid subgradient propagation functions for $(\mathbf{u},\mathbf{o})$. Then, the claimed result follows from Theorem 6.4.2. □

Remark 6.5.12. Observe that the functions $(\mathbf{V},\mathbf{W})$ constructed in the corollary above have similar affine forms to (6.5.2), which implies that the constructed forward subgradient propagation system for the OB relaxations is an affine ODE system. Thus, in principle, there is also a corresponding adjoint subgradient evaluation system for the OB relaxations, according to Theorem 6.4.13. However, implementing this adjoint method would require developing new adjoint counterparts of the forward subgradient evaluation functions $(\tilde{\mathbf{V}}^{\mathrm{OB}}, \tilde{\mathbf{W}}^{\mathrm{OB}})$.

6.6 Implementation and examples

6.6.1 Implementation

Proof-of-concept implementations were developed in Julia v1.4.2 [95] to compute subgradients for the SBM relaxations [2] and the OB relaxations [3], according to Corollaries 6.5.3 and 6.5.11. These implementations simultaneously integrate state bounding systems, state relaxation systems, and subgradient propagation systems. Harrison's bounding method [69] is employed, which computes state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ via an auxiliary ODE system whose right-hand side is derived from natural interval extensions [48]. The SBM state relaxation system [2] introduced in Section 2.4 is constructed. The corresponding subgradient propagation system is thus constructed according to Corollary 6.5.3, whose right-hand side employs subgradients of generalized McCormick relaxations by Mitsos et al. [35]. We have developed a new implementation of the OB relaxations [3] in Julia. Similarly to to the MATLAB implementation used in [3], the new Julia implementation automatically constructs $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$ as the McCormick relaxations of $\mathbf{f}$ in (2.3.1), and then the right-hand side functions $(\mathbf{u}, \mathbf{o})$ are evaluated naively by solving the convex optimization problems in (5.2.2) using numerical NLP solvers. The local optimizer IPOPT v3.13.2 [120] is employed in the Julia implementation with a convergence tolerance of 10^{-8}. The subgradients of $(\mathbf{f}^{\mathrm{cv}}, \mathbf{f}^{\mathrm{cc}})$, as described by Mitsos et al. [35], are supplied to IPOPT in place of gradients. Refer to [3,124] for further discussion about using IPOPT to solve nonsmooth optimization problems. The corresponding subgradient propagation system is thus constructed according to Corollary 6.5.11. EAGO v0.4.1 [94] is used to automatically compute natural interval extensions, generalized McCormick relaxations, and Mitsos et al.'s subgradients via operator overloading. We employ the ODE solver `BS3()` from the package DifferentialEquations v6.15.0 [123] with an absolute tolerance of 10^{-6} and a relative tolerance of 10^{-6} to solve all ODE systems. We employ JuMP v0.21.3 [121] as an interface with IPOPT. All computation in this section was performed on a Dell desktop computer with two 3.00 GHz Intel Core i7-9700 CPUs and 16.0 GB of RAM.

6.6.2 Numerical examples

Using the implementations described in Section 6.6.1, this subsection presents two numerical examples to illustrate our new forward dynamic subgradient evaluation methods proposed in Corollaries 6.5.3 and 6.5.11.

The following example is adapted from [2, Example 1], and shows that the new subgradient propagation system proposed in Corollary 6.5.3 appears to yield valid subgradients for the SBM relaxations [2].

Example 6.1. Let $P := [-3.0, 3.0] \times [0.21, 0.5]$ and $I := [0, 4.0]$, and consider the following instance of (2.3.1):

$$\begin{aligned} \dot{x}_1(t) &= -(2+\sin(p_1/3))x_1^2 + p_2 x_1 x_2, & x_1(0) &= 1.0, \\ \dot{x}_2(t) &= \sin(p_1/3)x_1^2 - p_2|x_1 x_2|, & x_2(0) &= 0.5. \end{aligned} \tag{6.6.1}$$

The SBM relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ for (6.6.1) were generated numerically by applying our Julia implementation, and it was observed that $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ satisfied (6.3.1). We arbitrarily picked two mesh points $\mathbf{p}^{\kappa} \in \tilde{P}$, and then $(\mathbf{S}^{\mathrm{cv}}(t_f, \mathbf{p}^{\kappa}), \mathbf{S}^{\mathrm{cc}}(t_f, \mathbf{p}^{\kappa}))$ were generated numerically at the two mesh points $\mathbf{p}^\kappa$ by applying our Julia implementation of the new subgradient propagation system in Corollary 6.5.3. Figure 6.1 presents two cross-sectional plots, which depict the original parametric solution of (6.6.1), the SBM relaxations, and two candidate subtangent cross-sections derived from $(\mathbf{S}^{\mathrm{cv}}(t_f,\mathbf{p}^\kappa),\mathbf{S}^{\mathrm{cc}}(t_f,\mathbf{p}^\kappa))$ and $(\mathbf{x}^{\mathrm{cv}}(t_f,\mathbf{p}^\kappa),\mathbf{x}^{\mathrm{cc}}(t_f,\mathbf{p}^\kappa))$. Observe that these candidates are indeed valid subtangent cross-sections in both smooth and nonsmooth cases, which implies that our new dynamic subgradient propagation system appears to yield valid subgradients for the SBM relaxations, as guaranteed by Corollary 6.5.3.

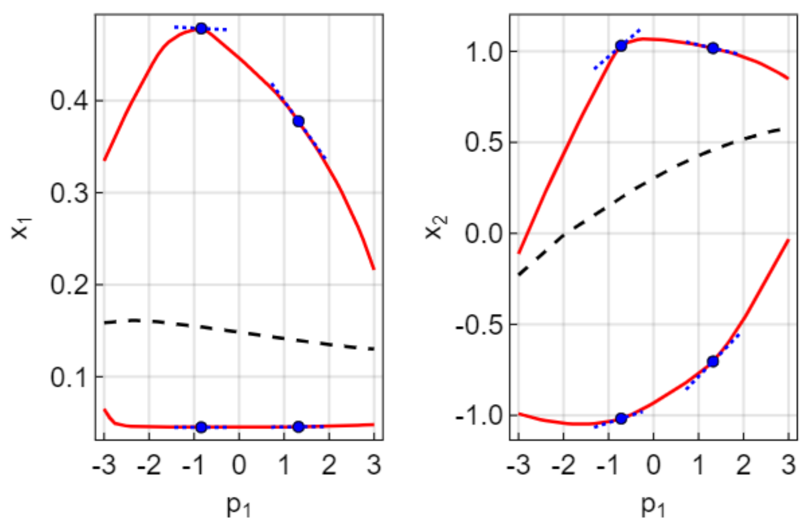

Figure 6.1: A cross-section at $p_2 := 0.4$ of the solution $x_1(4.0,\cdot)$ (*left, dashed black*) and $x_2(4.0,\cdot)$ (*right, dashed black*) of the ODE system (6.6.1) from Example 6.1, along with corresponding SBM relaxations (*solid red*) and subtangent cross-sections (*dotted blue*) derived from the subgradients at two reference points (*blue circle*)

The following example is a model of catalytic cracking of gas oil from [152, Example 15.3.5], which was studied in [108] for global optimization. This example shows that the new subgradient propagation system proposed in Corollary 6.5.11 appears to yield valid subgradients for the OB relaxations [3] as summarized in Section 6.5.2.

Example 6.2. Let $P := [11.0, 14.0] \times [7.0, 10.0] \times [1.0, 3.0]$ and $I := [0, 0.95]$, and consider the following instance of (2.3.1):

$$\begin{aligned} \dot{x}_1(t) &= -(p_1 + p_3)x_1^2, \quad x_1(0) = 1.0, \\ \dot{x}_2(t) &= p_1 x_1^2 - p_2 x_2, \qquad x_2(0) = 0.0. \end{aligned} \tag{6.6.2}$$

The OB relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ for (6.6.2) were generated numerically by applying our Julia implementation, and it was observed that $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ satisfied (6.3.1). We arbitrarily picked two mesh points $\mathbf{p}^{\kappa} \in \tilde{P}$, and then $(\mathbf{S}^{\mathrm{cv}}(t_f, \mathbf{p}^{\kappa}), \mathbf{S}^{\mathrm{cc}}(t_f, \mathbf{p}^{\kappa}))$ were generated numerically at the two mesh points $\mathbf{p}^{\kappa}$ by applying our Julia implementation of the new subgradient propagation system according to Corollary 6.5.11. Figure 6.2 presents two cross-sectional plots, which depict the original parametric solution of (6.6.2), the OB relaxations, and two candidate subtangent cross-sections derived from $(\mathbf{S}^{\mathrm{cv}}(t_f, \mathbf{p}^{\kappa}), \mathbf{S}^{\mathrm{cc}}(t_f, \mathbf{p}^{\kappa}))$ and $(\mathbf{x}^{\mathrm{cv}}(t_f, \mathbf{p}^{\kappa}), \mathbf{x}^{\mathrm{cc}}(t_f, \mathbf{p}^{\kappa}))$. Observe that these candidates are indeed valid subtangent cross-sections in both smooth and nonsmooth cases, which implies that our new dynamic subgradient propagation system appears to yield valid subgradients for the OB relaxations, as guaranteed by Corollary 6.5.11.

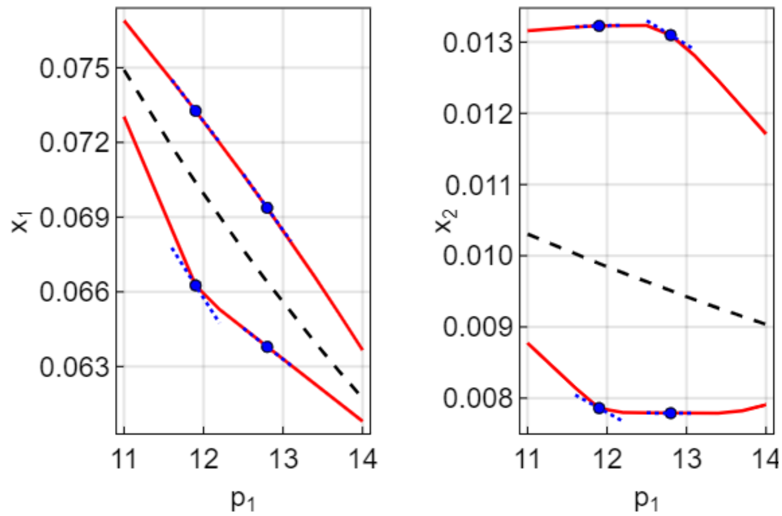

Figure 6.2: A cross-section at $(p_2, p_3) := (8.5, 2.0)$ of the solution $x_1(0.95, \cdot)$ (*left, dashed black*) and $x_2(0.95, \cdot)$ (*right, dashed black*) of the ODE system (6.6.2) from Example 6.2, along with corresponding OB relaxations (*solid red*) and subtangent cross-sections (*dotted blue*) derived from the subgradients at two reference points (*blue circle*)

6.7 Conclusions and future work

This article has proposed new methods for evaluating subgradients of state-of-the-art ODE relaxations [2, 3] in the Scott–Barton ODE relaxation framework (2.4.1), which enable computing lower bounds with these ODE relaxations in deterministic algorithms of global dynamic optimization. These methods assume that the underlying state relaxations $(\mathbf{x}^{\mathrm{cv}}, \mathbf{x}^{\mathrm{cc}})$ do not touch the predefined state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$, which is guaranteed to be satisfied for a sufficiently small domain of the uncertain parameters. Unlike a recent sensitivity evaluation approach [97] for nonsmooth dynamic systems, the new forward subgradient propagation systems constructed in Corollaries 6.5.3 and 6.5.11 are affine ODE systems, which may be easily integrated using off-the-shelf numerical solvers. Besides, the RHS of these ODE systems are constructed based on arbitrary subgradients of the relaxations embedded in the relaxation systems' RHS, which may be easily computed by established subgradient evaluation methods. Theorem 6.4.13 and Corollary 6.5.5 for the first time show that a subgradient of the objective function of a lower-bounding problem in global dynamic optimization can be directly evaluated using dynamic adjoint sensitivity evaluation approaches, which would speed up the lower-bounding procedure in global dynamic optimization.

Future work may involve developing subgradient evaluation methods for the Scott–Barton ODE relaxation framework without satisfying (6.3.1). This is important since at a node of the branch-and-bound tree in global optimization, the satisfaction of (6.3.1) may not be known *a priori*. Future work will also involve developing an implementation for the dynamic adjoint subgradient evaluation method in Corollary 6.5.5.

Chapter 7

Bounding Convex Relaxations of Process Models from Below by Tractable Black-Box Sampling

As mentioned in Section 1.1, when computing lower bounds in global dynamic/non-dynamic optimization methods by minimizing convex relaxations, subgradients are typically required by nonsmooth local optimization solvers to proceed effectively. However, due to limitations in convex analysis theory, there are currently no methods for computing subgradients for certain useful convex relaxations, such as state relaxations obtained using the Scott–Barton framework as summarized in Section 2.4. Recall that the dynamic subgradient evaluation methods for these state relaxations proposed in Chapter 6 require certain assumptions. In the case where convex relaxations are known to be correct but subgradients are unavailable, this chapter, reproduced from the published journal article [4], proposes a new approach for tractably constructing useful, correct affine underestimators and lower bounds of the original convex relaxations just by black-box sampling. No additional assumptions are required, and no subgradients or gradients must be computed at any point. The new affine underestimators are shown to converge rapidly to an original nonconvex function as domain shrinks, which is beneficial in global optimization. Variants of these methods are presented to account for numerical error or noise in the sampling procedure. Notably, [4, Example 7] successfully employed the SBM relaxations [2] for solving a dynamic parameter estimation problem, without access to dynamic subgradients. The associated article [4] was written in collaboration with colleagues, but this chapter only presents the contributions of the author of this thesis.

7.1 Introduction

This chapter considers general nonconvex dynamic/non-dynamic optimization problems. Models of chemical processes may exhibit nonconvexity [30], thus complicating simulation and optimization. Nonconvexity in process models may arise, for instance, due to process dynamics [1,2], standard thermodynamic relationships such as the Peng-Robinson equation of state, discrete switches between design choices [153] or operating regimes [154], or established correlations used to model individual process units such as compressors [155].

Though stochastic global search algorithms can be useful for nonconvex process optimization [156], deterministic methods for global (nonconvex) optimization typically guarantee that an ε-optimal solution will be located and verified in finite time under relatively mild assumptions [30, 157]. Typical branch-and-bound methods for deterministic global optimization proceed by generating upper and lower bounds on the globally minimal objective value on various subdomains. Upper bounds are generally computed by off-the-shelf local nonlinear programming (NLP) solvers, whereas lower bounds are computed by constructing *convex relaxations* of the original problem, and minimizing these convex relaxations using local NLP solvers. Several established methods construct useful convex relaxations automatically; for nontrivial composite nonconvex functions, these methods include the Auxiliary Variable Method underlying BARON [30, 60], McCormick relaxations [5, 35] and later variants [7, 33, 34], αBB relaxations [9, 49], and convex envelopes of edge-concave relaxations [158]. Beyond the context of global optimization, convex relaxations are also used to provide useful enclosures for reachable sets of uncertain dynamic process models [81, 159].

When computing lower bounds in global optimization methods by minimizing convex relaxations, typical local NLP solvers require gradient information to proceed effectively, or subgradients in the case of nonsmoothness. The McCormick relaxations of functions are often nonsmooth, for example [35]. However, for certain useful convex relaxations such as the dynamic relaxations of [2], there are currently no methods to compute subgradients due to limitations in convex analysis theory. As shown in [130], if subgradients are unavailable, then finite differencing methods may also fail to approximate subgradients well even in the absence of numerical error. Moreover, when testing new types of convex relaxation in a global optimization setting, automatic differentiation tools for computing subgradients correctly may be unavailable or difficult to implement in the employed software platform.

Thus, in a global optimization setting, this chapter considers convex relaxations that are known to be correct, but are only available via black-box evaluation. Hence, subgradients are unavailable, and we must resort to derivative-free techniques [160] instead. The main contribution of this chapter is to show that, for a convex function of n variables defined on a box, without any further assumptions, a guaranteed closed-form affine underestimator of this function may be constructed by performing at most $(2n+1)$ black-box evaluations of the function, and assembling the results tractably using a new variant of centered finite differencing. Since global optimization methods typically require numerous lower bound evaluations, computational tractability becomes particularly important. Being affine, this new underestimator has a closed-form minimum on the domain box. These results are generalized to account for uncertainty or errors in the function evaluations, and to relax convex optimization problems by sampling the objective function and constraints. We also show that, under mild assumptions, our sampling-based underestimators converge to the original function rapidly in the sense of second-order pointwise convergence [36]. Thus, our sampling approach may be deployed effectively in methods for global optimization and reachable set estimation when subgradients are difficult or impossible to evaluate. Notably, [4, Example 7] successfully employed the dynamic SBM relaxations [2] (summarized in Section 2.4) for solving a dynamic parameter estimation problem to global optimality, without access to dynamic subgradients. In the vein of [9, Section 2.7], our approach also permits global optimization solvers to handle sums of known nonconvex functions and black-box convex functions, by using sampling to handle the latter. Another application of affine relaxations involves sidestepping the nonsmoothness of McCormick relaxations (c.f. [33]) during lower-bounding problems in global optimization.

We note that, as discussed by [160], black-box methods are only recommended when gradients or subgradients really are unavailable. By construction, our new black-box affine underestimators are guaranteed to underestimate the original convex relaxations they are sampled from, and are guaranteed to underestimate affine relaxations constructed from actual subgradients.

Larson et al. [161] recently considered a similar problem. In the context of mixed-integer nonlinear programming, Larson et al. sample a convex function several times in a box, deduce secant information from these evaluations, and combine these secants to construct a nonconvex discontinuous piecewise-affine underestimator of this function. This underestimator may be evaluated by solving a mixed-integer linear program with a number of constraints that grows exponentially with problem dimension. Unlike [161], we seek to obtain underestimators that are convex and may be tractably constructed and evaluated. Nevertheless, our results are related; we employ intermediate results obtained by [161] in our proofs, and it can be shown that our new affine underestimators are also underestimators of the discontinuous piecewise-affine relaxations of Larson et al. when the same sampled points are chosen.

The remainder of this chapter is structured as follows. Section 7.2 presents the new affine relaxation approach for a convex function via black-box sampling, and Sections 7.3, 7.4, and 7.5 summarizes Song's contributions in theory development of [4]. Section 7.3 establishes tightness and convergence properties of the new affine underestimator. Section 7.4 derives an extension of the new affine relaxation approach to account for numerical error or noise in the sampling procedure. Section 7.5 presents another extension of the new approach, which employs a sampled set that is not centered within the considered box domain.

7.2 New affine relaxation formulation

The new affine relaxation approach focuses on a generic convex function defined on a box domain, as formalized in the following definition.

Definition 7.2.1. Consider vectors $\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}} \in \mathbb{R}^n$ for which $\mathbf{x}^{\mathrm{L}} \leq \mathbf{x}^{\mathrm{U}}$, the nonempty box $X := \{\boldsymbol{\xi} \in \mathbb{R}^n : \mathbf{x}^{\mathrm{L}} \leq \boldsymbol{\xi} \leq \mathbf{x}^{\mathrm{U}}\}$, and a convex function $f : X \to \mathbb{R}$. With slight abuse with notation,

$\mathbf{f}$ and $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ in this chapter do not represent the original ODE right-hand side and state bounds of $\mathbf{x}$ in (2.3.1).

Under this problem setup, we suppose that black-box evaluations of f are possible, and we wish to generate a guaranteed affine relaxation of f on X without assuming any additional knowledge of f. To proceed, we sample this function $(2n+1)$ times in a star-shaped stencil that is formalized in the following definition. This definition also describes certain intermediate quantities that will be useful when formulating our results.

Definition 7.2.2. Consider the problem setup in Definition 7.2.1, and define the following additional sets and quantities:

- a nondegenerate index set $I := \{i \in \{1, \ldots, n\} : x_i^{\mathrm{L}} < x_i^{\mathrm{U}}\}$ (note that I here is not the time horizon of ODE systems as in chapters above),
- the midpoint $\mathbf{w}^{(0)} := \frac{1}{2}(\mathbf{x}^{\mathrm{L}} + \mathbf{x}^{\mathrm{U}})$ of X,
- let $\mathbf{e}^{(i)} \in \mathbb{R}^n$ denote the i^{th} unit coordinate vector in $\mathbb{R}^n$, for each $i \in I$: a step length $\alpha_i \in (0,1]$, and two vectors

$$\mathbf{w}^{(\pm i)} := \mathbf{w}^{(0)} \pm \frac{\alpha_i}{2}(x_i^{\mathrm{U}} - x_i^{\mathrm{L}})\,\mathbf{e}^{(i)},$$

- a sample set

$$W := \{\mathbf{w}^{(0)}\} \cup \{\mathbf{w}^{(+i)} : i \in I\} \cup \{\mathbf{w}^{(-i)} : i \in I\} \subset X,$$

- function values $y_0 := f(\mathbf{w}^{(0)})$ and $y_{\pm i} := f(\mathbf{w}^{(\pm i)})$ for each $i \in I$,

- a vector $\mathbf{b} \in \mathbb{R}^n$ for which, for each $i \in \{1,\ldots,n\}$:

$$b_i := \begin{cases} \dfrac{y_{+i} - y_{-i}}{\alpha_i(x_i^{\mathrm{U}} - x_i^{\mathrm{L}})} = \dfrac{y_{+i} - y_{-i}}{\|\mathbf{w}^{(+i)} - \mathbf{w}^{(-i)}\|_\infty} & \text{if } i \in I \\ 0 & \text{if } i \notin I, \end{cases}$$

- a scalar $c \in \mathbb{R}$ for which

$$c := y_0 - \frac{1}{2}\sum_{i \in I}\left(\frac{y_{+i} + y_{-i} - 2y_0}{\alpha_i}\right),$$

- and a scalar $f^{\mathrm{L}} \in \mathbb{R}$ for which

$$f^{\mathrm{L}} := y_0 - \sum_{i \in I}\left(\frac{\max(y_{+i}, y_{-i}) - y_0}{\alpha_i}\right).$$

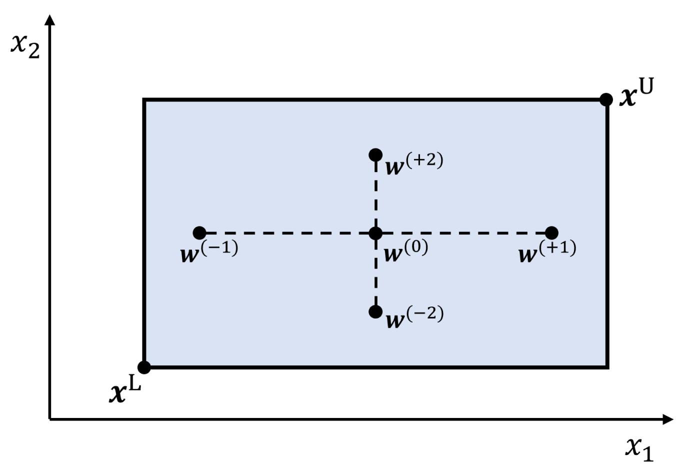

Figure 7.1: Illustration of the domain of the function f described in Definition 7.2.2.

Some of these quantities are depicted in Figure 7.1. The vector $\mathbf{b}$ has been employed in established derivative-free optimization approaches, where it is known as a *centered simplex gradient* of f at $\mathbf{w}^{(0)}$ sampled along the coordinate vectors [160]. If f were smooth, then $\mathbf{b}$ would also be a standard centered finite difference approximation of the gradient $\nabla f(\mathbf{w}^{(0)})$. The sum in the definition of c resembles a standard finite difference approximation of the (possibly nonexistent) second-order partial derivative $\frac{\partial^2 f}{\partial x_i^2}(\mathbf{w}^{(0)})$, except with a different denominator.

Then, as proved in [4, Theorem 1 and Corollary 1], $\mathbf{f}^{\mathrm{aff}} : \mathbf{x} \mapsto c + \langle \mathbf{b}, \mathbf{x} - \mathbf{w}^{(0)} \rangle$ is the new affine underestimator of the convex function f on X, and f^{L} is a lower bound of f on X. For illustration, Figure 7.2 (reproduced from [4, Fig. 6]) depicts an original nonconvex function, convex relaxations on various subdomains, and corresponding new sampling-based affine relaxations derived from the convex relaxations.

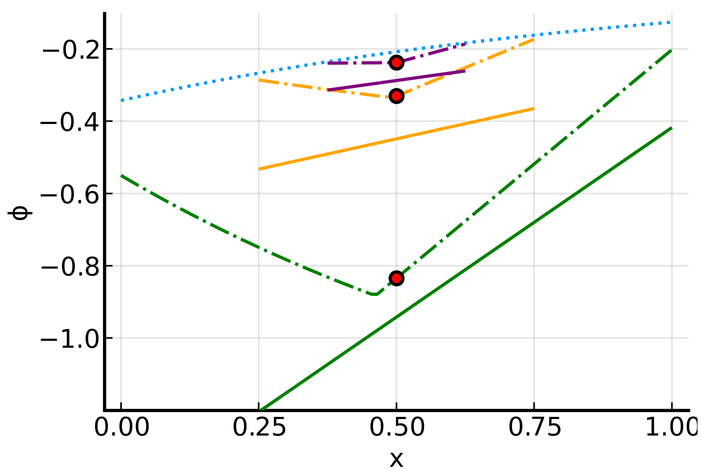

Figure 7.2: Illustration of an original function ϕ (*dotted blue*), along with convex relaxations (*dash-dotted*) on various subdomains, and corresponding new sampling-based affine relaxations (*solid*), from [4, Fig. 6]. One sample data point $(\mathbf{w}^{(0)}, y_0)$ (*circle*) is depicted in each case.

7.3 Tightness and convergence properties

[4, Theorem 1 and Corollary 1] show that the constructed $\mathbf{f}^{\mathrm{aff}}$ and $\mathbf{f}^{\mathrm{L}}$ are guaranteed to be respectively an affine relaxation and a lower bound of f on X in Definition 7.2.1. A tighter sampling-based affine relaxation formulation in a univariate case is discussed in [4, Section 3.3]. This section then summarizes two theoretical results developed by the author of this thesis. Theorem 7.3.1 below shows that the sampling-based affine relaxation is tighter as each α_i in Definition 7.2.2 shrinks. Theorem 7.3.4 shows that under mild assumptions, the new affine relaxation is guaranteed to exhibit second-order pointwise convergence [36] to an original nonconvex function, which help mitigate the cluster effects [37, 38] in global optimization.

Theorem 7.3.1. Consider the problem setup in Definition 7.2.1 and the auxiliary quantities in Definition 7.2.2, except with two choices of α_i for each $i \in I$, denoted with superscripts "A" and "B". Suppose that $\alpha_i^{\mathrm{A}} \leq \alpha_i^{\mathrm{B}}$ for each $i \in I$. Let quantities computed with α_i^{A} in place of α_i be given the superscript "A", and let quantities computed with α_i^{B} in place of α_i be given the superscript "B". Then, for each $\mathbf{x} \in X$,

$$c^{\mathrm{A}} + \langle \mathbf{b}^{\mathrm{A}}, \mathbf{x} - \mathbf{w}^{(0)} \rangle \geq c^{\mathrm{B}} + \langle \mathbf{b}^{\mathrm{B}}, \mathbf{x} - \mathbf{w}^{(0)} \rangle,$$

and $f^{\mathrm{L,A}} \geq f^{\mathrm{L,B}}$.

Proof. For each $i \in I$, define a function $h_i : (0,1] \times [x_i^{\mathrm{L}}, x_i^{\mathrm{U}}] \to \mathbb{R}$ for which

$$h_i : (\alpha, \xi) \mapsto \frac{y_{+i}(\alpha) - y_{-i}(\alpha)}{\alpha} \left(\frac{\xi - w_i^{(0)}}{x_i^{\mathrm{U}} - x_i^{\mathrm{L}}} \right) - \frac{y_{+i}(\alpha) + y_{-i}(\alpha) - 2y_0}{2\alpha},$$

where $y_{+i}(\alpha)$ and $y_{-i}(\alpha)$ are defined according to Definition 7.2.2 except with $\alpha_i := \alpha$. Thus, for each $\alpha \in (0,1]$,

$$h_i(\alpha, x_i^{\mathrm{L}}) = \frac{y_0 - y_{+i}(\alpha)}{\alpha} \qquad \text{and} \qquad h_i(\alpha, x_i^{\mathrm{U}}) = \frac{y_0 - y_{-i}(\alpha)}{\alpha}.$$

[146, Theorem 23.1] implies that both mappings $h_i(\cdot, x_i^{\mathrm{L}})$ and $h_i(\cdot, x_i^{\mathrm{U}})$ are non-increasing, and so

$$h_i(\alpha_i^{\mathrm{B}}, x_i^{\mathrm{L}}) \leq h_i(\alpha_i^{\mathrm{A}}, x_i^{\mathrm{L}}) \quad \text{and} \quad h_i(\alpha_i^{\mathrm{B}}, x_i^{\mathrm{U}}) \leq h_i(\alpha_i^{\mathrm{A}}, x_i^{\mathrm{U}}).$$

Thus, noting that $h_i(\alpha_i^{\mathrm{B}}, \cdot)$ and $h_i(\alpha_i^{\mathrm{A}}, \cdot)$ are affine on $[x_i^{\mathrm{L}}, x_i^{\mathrm{U}}]$, it follows that

$$h_i(\alpha_i^{\mathrm{B}}, \xi) \leq h_i(\alpha_i^{\mathrm{A}}, \xi), \qquad \forall \xi \in [x_i^{\mathrm{L}}, x_i^{\mathrm{U}}].$$

Hence, for each $\mathbf{x} \in X$, we have

$$(c^{\mathrm{A}} + \langle \mathbf{b}^{\mathrm{A}}, \mathbf{x} - \mathbf{w}^{(0)} \rangle) - (c^{\mathrm{B}} + \langle \mathbf{b}^{\mathrm{B}}, \mathbf{x} - \mathbf{w}^{(0)} \rangle) = \sum_{i \in I} \big(h_i(\alpha_i^{\mathrm{A}}, x_i) - h_i(\alpha_i^{\mathrm{B}}, x_i) \big) \geq 0,$$

as claimed. Thus,

$$f^{\mathrm{L,A}} = \inf_{\mathbf{x} \in X} (c^{\mathrm{A}} + \langle \mathbf{b}^{\mathrm{A}}, \mathbf{x} - \mathbf{w}^{(0)} \rangle) \geq \inf_{\mathbf{x} \in X} (c^{\mathrm{B}} + \langle \mathbf{b}^{\mathrm{B}}, \mathbf{x} - \mathbf{w}^{(0)} \rangle) = f^{\mathrm{L,B}},$$

as claimed. □

Now, consider the following setup involving an underlying nonconvex function and a *scheme of convex relaxations.*

Definition 7.3.2. Consider a nonempty open set $Z \subset \mathbb{R}^n$, a nonempty compact set $Q \subset Z$, and a nonconvex function $g : Z \to \mathbb{R}$. Let $\mathbb{I}Q$ denote the collection of boxes of the form $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^{\mathrm{L}} \leq \mathbf{x} \leq \mathbf{x}^{\mathrm{U}}\}$ that are subsets of Q. For any box $X \in \mathbb{I}Q$, define the *width* of X as

$$\mathrm{wid}\, X := \max\{\|\mathbf{x} - \boldsymbol{\xi}\|_\infty : \mathbf{x}, \boldsymbol{\xi} \in X\}.$$

Consider a *scheme of convex relaxations* $\{g^{\mathrm{cv},X}\}_{X \in \mathbb{I}Q}$ of g for which, for any $X \in \mathbb{I}Q$,

$g^{\mathrm{cv},X} : X \to \mathbb{R}$ is a convex relaxation of g on X.

With this foundation, we now modify the setup of Definition 7.2.2 as follows by substituting $g^{\mathrm{cv},X}$ in place of f.

Definition 7.3.3. Under Definition 7.3.2, consider the quantities in Definition 7.2.2 for each $X \in \mathbb{I}Q$, with $g^{\mathrm{cv},X}$ in place of f. Denote the dependence of these quantities on X explicitly with a superscript (e.g. $\mathbf{x}^{\mathrm{L},X}$, I^X, $\mathbf{w}^{(0),X}$, α_i^X). Let $g^{\mathrm{aff},X} : X \to \mathbb{R}$ denote the corresponding affine relaxation of $g^{\mathrm{cv},X}$ on X provided by [4, Theorem 1]. Thus, for each $\mathbf{x} \in X$,

$$g^{\mathrm{aff},X}(\mathbf{x}) := c^X + \langle \mathbf{b}^X, \mathbf{x} - \mathbf{w}^{(0),X} \rangle. \tag{7.3.1}$$

The following theorem shows that if a smooth function g has a scheme of convex relaxations with second-order pointwise convergence in the sense of [36, Definition 10], then our new sampling-based affine relaxations of this scheme will inherit this second-order pointwise convergence.

Theorem 7.3.4. Consider the setup of Definitions 7.3.2 and 7.3.3. Suppose that the function g is twice-continuously differentiable on Z, and that there is a scalar $\tau^{\mathrm{cv}} > 0$ for which

$$\sup_{\mathbf{x} \in X} \left(g(\mathbf{x}) - g^{\mathrm{cv},X}(\mathbf{x}) \right) \leq \tau^{\mathrm{cv}} (\mathrm{wid}\, X)^2, \quad \forall X \in \mathbb{I}Q. \tag{7.3.2}$$

Then, there exists another scalar $\tau^{\mathrm{aff}} > 0$ for which

$$\sup_{\mathbf{x} \in X} \left(g(\mathbf{x}) - g^{\mathrm{aff},X}(\mathbf{x}) \right) \leq \tau^{\mathrm{aff}} (\mathrm{wid}\, X)^2, \quad \forall X \in \mathbb{I}Q. \tag{7.3.3}$$

Proof. For each $X \in \mathbb{I}Q$, consider a subgradient $\mathbf{s}^{(0),X}$ of $g^{\mathrm{cv},X}$ at $\mathbf{w}^{(0),X}$ (with $s_i^{(0),X}$ arbitrarily chosen to be 0 whenever $i \notin I^X$), and let $g^{\mathrm{sub},X} : X \to \mathbb{R}$ denote the corresponding subtangent, so that:

$$g^{\mathrm{sub},X}(\mathbf{x}) = y_0^X + \langle \mathbf{s}^{(0),X}, \mathbf{x} - \mathbf{w}^{(0),X} \rangle, \qquad \forall \mathbf{x} \in X.$$

Since $\mathbf{w}^{(0),X}$ is in any *centrally-scaled interval* in X in the sense of [46, Definition 4.1], since g is twice-continuously differentiable, and since (7.3.2) holds, [46, Theorem 4.2] shows that there is a scalar $\tau^{\mathrm{sub}} > 0$ (independent of X) for which

$$\sup_{\mathbf{x} \in X} \big(g(\mathbf{x}) - g^{\mathrm{sub},X}(\mathbf{x}) \big) \leq \tau^{\mathrm{sub}} (\mathrm{wid}\, X)^2. \tag{7.3.4}$$

According to Theorem 7.3.1, for any $X \in \mathbb{I}Q$ and $\mathbf{x} \in X$, $g^{\mathrm{aff},X}(\mathbf{x})$ decreases as any α_i^X increases. Thus, in the remainder of this proof, it suffices to consider only the case in which $\alpha_i^X := 1$ for each $X \in \mathbb{I}Q$ and each $i \in I^X$.

Now, consider any fixed $X \in \mathbb{I}Q$ and $i \in I^X$. Since g is twice-continuously differentiable, Taylor's Theorem (as described by [162, Theorem 2.1]) implies that there exists a point $\mathbf{d}^{\mathrm{A}} := \lambda^{\mathrm{A}} \mathbf{w}^{(+i),X} + (1 - \lambda^{\mathrm{A}}) \mathbf{w}^{(0),X}$ for some $0 < \lambda^{\mathrm{A}} < 1$ for which

$$g(\mathbf{w}^{(+i),X}) - g(\mathbf{w}^{(0),X}) = \frac{1}{2} \frac{\partial g}{\partial x_i}(\mathbf{d}^{\mathrm{A}}) \, (x_i^{\mathrm{U},X} - x_i^{\mathrm{L},X}). \tag{7.3.5}$$

Similarly, there exists a point $\mathbf{d}^{\mathrm{B}} := \lambda^{\mathrm{B}} \mathbf{w}^{(-i),X} + (1 - \lambda^{\mathrm{B}}) \mathbf{w}^{(0),X}$ for some $0 < \lambda^{\mathrm{B}} < 1$ for which

$$g(\mathbf{w}^{(-i),X}) - g(\mathbf{w}^{(0),X}) = -\frac{1}{2} \frac{\partial g}{\partial x_i}(\mathbf{d}^{\mathrm{B}}) \, (x_i^{\mathrm{U},X} - x_i^{\mathrm{L},X}). \tag{7.3.6}$$

Adding (7.3.5) and (7.3.6) yields

$$\begin{aligned} g(\mathbf{w}^{(+i),X}) + g(\mathbf{w}^{(-i),X}) - 2g(\mathbf{w}^{(0),X}) &= \frac{1}{2}\left(\frac{\partial g}{\partial x_i}(\mathbf{d}^{\mathrm{A}}) - \frac{\partial g}{\partial x_i}(\mathbf{d}^{\mathrm{B}})\right)(x_i^{\mathrm{U},X} - x_i^{\mathrm{L},X}) \\ &\leq \frac{1}{2}\left|\frac{\partial g}{\partial x_i}(\mathbf{d}^{\mathrm{A}}) - \frac{\partial g}{\partial x_i}(\mathbf{d}^{\mathrm{B}})\right| \mathrm{wid}\, X. \end{aligned} \tag{7.3.7}$$

Now, Taylor's Theorem implies there is a point $\mathbf{d}^{\mathrm{C}} := \lambda^{\mathrm{C}}\mathbf{d}^{\mathrm{A}} + (1-\lambda^{\mathrm{C}})\mathbf{d}^{\mathrm{B}}$ for some $0 < \lambda^{\mathrm{C}} < 1$ for which

$$\frac{\partial g}{\partial x_i}(\mathbf{d}^{\mathrm{A}}) - \frac{\partial g}{\partial x_i}(\mathbf{d}^{\mathrm{B}}) = \frac{\partial^2 g}{\partial x_i^2}(\mathbf{d}^{\mathrm{C}})\,(d_i^{\mathrm{A}} - d_i^{\mathrm{B}}).$$

Moreover, since g is twice-continuously differentiable and Q is compact, there exists $\tau^{\mathrm{H}} > 0$ (independent of X) for which

$$\left|\frac{\partial^2 g}{\partial x_j^2}(\mathbf{x})\right| \leq \tau^{\mathrm{H}}, \qquad \forall j \in \{1,\ldots,n\}, \quad \forall \mathbf{x} \in Q.$$

Thus,

$$\left|\frac{\partial g}{\partial x_i}(\mathbf{d}^{\mathrm{A}}) - \frac{\partial g}{\partial x_i}(\mathbf{d}^{\mathrm{B}})\right| \leq \tau^{\mathrm{H}}\mathrm{wid}\, X. \tag{7.3.8}$$

Combining (7.3.7) and (7.3.8) yields

$$g(\mathbf{w}^{(+i),X}) + g(\mathbf{w}^{(-i),X}) - 2g(\mathbf{w}^{(0),X}) \leq \frac{1}{2}\tau^{\mathrm{H}}(\mathrm{wid}\, X)^2. \tag{7.3.9}$$

Since $g^{\mathrm{cv},X}$ is a convex relaxation of g on X, we have

$$y_{+i}^{X} - g(\mathbf{w}^{(+i),X}) \leq 0 \qquad \text{and} \qquad y_{-i}^{X} - g(\mathbf{w}^{(-i),X}) \leq 0. \tag{7.3.10}$$

Moreover, (7.3.2) implies

$$2g(\mathbf{w}^{(0),X}) - 2y_0^X \leq 2\tau^{\mathrm{cv}}(\mathrm{wid}\, X)^2. \tag{7.3.11}$$

Adding (7.3.10) to (7.3.11) yields

$$\left(y_{+i}^X + y_{-i}^X - 2y_0^X\right) - \left(g(\mathbf{w}^{(+i),X}) + g(\mathbf{w}^{(-i),X}) - 2g(\mathbf{w}^{(0),X})\right) \leq 2\tau^{\mathrm{cv}}(\mathrm{wid}\, X)^2.$$

Moreover, since (7.3.9) holds and since $g^{\mathrm{cv},X}$ is convex, we have

$$0 \leq y_{+i}^X + y_{-i}^X - 2y_0^X \leq (2\tau^{\mathrm{cv}} + \tfrac{1}{2}\tau^{\mathrm{H}})(\mathrm{wid}\, X)^2. \tag{7.3.12}$$

Combining (7.3.12) for each $i \in I^X$ with the definition of c^X, and recalling that $\alpha_i = 1$ for each $i \in I^X$ by assumption, we obtain

$$0 \leq y_0^X - c^X = \frac{1}{2}\sum_{i \in I^X}\left(y_{+i}^X + y_{-i}^X - 2y_0^X\right) \leq n(\tau^{\mathrm{cv}} + \tfrac{1}{4}\tau^{\mathrm{H}})(\mathrm{wid}\, X)^2. \tag{7.3.13}$$

Next, consider any fixed $i \in I^X$ and any subgradient $\mathbf{s}^{(0),X}$ of $g^{\mathrm{cv},X}$ at $\mathbf{w}^{(0),X}$. Thus,

$$\pm\frac{1}{2}(x_i^{\mathrm{U},X} - x_i^{\mathrm{L},X})\, s_i^{(0),X} \leq y_{\pm i}^X - y_0. \tag{7.3.14}$$

The positive branch of (7.3.14) implies that

$$s_i^{(0),X} \leq \frac{2(y_{+i}^X - y_0^X)}{x_i^{\mathrm{U},X} - x_i^{\mathrm{L},X}}.$$

Thus, according to (7.3.12) and the definition of b_i^X,

$$\begin{aligned}
s_i^{(0),X} - b_i^X &\leq \frac{2(y_{+i}^X - y_0^X) - (y_{+i}^X - y_{-i}^X)}{x_i^{\mathrm{U},X} - x_i^{\mathrm{L},X}} \\
&= \frac{y_{+i}^X + y_{-i}^X - 2y_0^X}{x_i^{\mathrm{U},X} - x_i^{\mathrm{L},X}} \\
&\leq \frac{2\tau^{\mathrm{cv}} + \frac{\tau^{\mathrm{H}}}{2}}{(x_i^{\mathrm{U},X} - x_i^{\mathrm{L},X})}(\mathrm{wid}\,X)^2
\end{aligned}$$

Similarly, combining the negative branch of (7.3.14) with the definition of b_i^X and (7.3.12) yields

$$s_i^{(0),X} - b_i^X \geq -\frac{2\tau^{\mathrm{cv}} + \frac{\tau^{\mathrm{H}}}{2}}{(x_i^{\mathrm{U},X} - x_i^{\mathrm{L},X})}(\mathrm{wid}\,X)^2.$$

Combining the above results yields:

$$|s_i^{(0),X} - b_i^X| \leq \frac{2\tau^{\mathrm{cv}} + \frac{\tau^{\mathrm{H}}}{2}}{(x_i^{\mathrm{U},X} - x_i^{\mathrm{L},X})}(\mathrm{wid}\,X)^2. \tag{7.3.15}$$

Since (7.3.13) and (7.3.15) hold, for any $X \in \mathbb{I}Q$, $\mathbf{s}^{(0),X}$, and $\mathbf{x} \in X$, we have

$$\begin{aligned}
|g^{\mathrm{sub},X}(\mathbf{x}) - g^{\mathrm{aff},X}(\mathbf{x})| &= |y_0^X - c^X + \langle \mathbf{s}^{(0),X} - \mathbf{b}^X, \mathbf{x} - \mathbf{w}^{(0),X}\rangle| \\
&\leq |y_0^X - c^X| + |\langle \mathbf{s}^{(0),X} - \mathbf{b}^X, \mathbf{x} - \mathbf{w}^{(0),X}\rangle| \\
&\leq n(\tau^{\mathrm{cv}} + \frac{1}{4}\tau^{\mathrm{H}})(\mathrm{wid}\,X)^2 + \sum_{i \in I^X}\left(|s_i^{(0),X} - b_i^X||x_i - w_i^{(0),X}|\right) \\
&\leq n(\tau^{\mathrm{cv}} + \tfrac{1}{4}\tau^{\mathrm{H}})(\mathrm{wid}\,X)^2 + \sum_{i \in I^X}\left((2\tau^{\mathrm{cv}} + \tfrac{1}{2}\tau^{\mathrm{H}})(\mathrm{wid}\,X)^2\right) \\
&\leq \big(3n\tau^{\mathrm{cv}} + \tfrac{3}{4}n\tau^{\mathrm{H}}\big)(\mathrm{wid}\,X)^2.
\end{aligned}$$

Thus, since (7.3.4) holds, we have

$$\begin{aligned}\sup_{\mathbf{x}\in X}\left(g(\mathbf{x})-g^{\mathrm{aff},X}(\mathbf{x})\right) &\leq \sup_{\mathbf{x}\in X}\left(g(\mathbf{x})-g^{\mathrm{aff},X}(\mathbf{x})\right)+\sup_{\mathbf{x}\in X}\left(g^{\mathrm{aff},X}(\mathbf{x})-g^{\mathrm{sub},X}(\mathbf{x})\right)\\ &\leq (\tau^{\mathrm{sub}}+3n\tau^{\mathrm{cv}}+\tfrac{3}{4}n\tau^{\mathrm{H}})(\mathrm{wid}\,X)^2.\end{aligned}$$

Thus, the claimed result holds with $\tau^{\mathrm{aff}} := \tau^{\mathrm{sub}}+3n\tau^{\mathrm{cv}}+\frac{3}{4}n\tau^{\mathrm{H}}$. □

7.4 Accounting for noise of sampling

This section modifies [4, Theorem 1 and Corollary 1] to account for a case in which the function evaluations y_0 and $y_{\pm i}$ in Definition 7.2.2 are unavailable directly. Instead, we suppose that corresponding approximations $\tilde{y}_0$ and $\tilde{y}_{\pm i}$ are available, and it is known that these approximations are valid to within a known absolute tolerance $\varepsilon > 0$.

These results permit us to construct guaranteed affine underestimators and lower bounds for convex functions when there is empirical noise or numerical error in these functions' evaluations, or when validated arithmetic or outward rounding are employed. Explicitly accounting for noise in this context can be crucial; if neglecting noise leads us to construct an approximate convex relaxation that is not in fact a valid relaxation, then we may inadvertently relax a feasible optimization problem into an infeasible approximate relaxation, and incorrectly conclude that the original problem was infeasible.

Corollary 7.4.1. Consider the setup of Definition 7.2.2, and suppose that there exist values $\varepsilon > 0$, $\tilde{y}_0 \in \mathbb{R}$, and $\tilde{y}_{\pm i} \in \mathbb{R}$ for each $i \in I$, for which:

$$|\tilde{y}_0 - y_0| \leq \varepsilon, \quad |\tilde{y}_{+i} - y_{+i}| \leq \varepsilon, \quad \text{and} \quad |\tilde{y}_{-i} - y_{-i}| \leq \varepsilon \qquad \forall i \in I. \tag{7.4.1}$$

Define a vector $\tilde{\mathbf{b}} \in \mathbb{R}^n$ so that for each $i \in \{1, ..., n\}$,

$$\tilde{b}_i := \begin{cases} \dfrac{\tilde{y}_{+i} - \tilde{y}_{-i}}{\alpha_i(x_i^{\mathrm{U}} - x_i^{\mathrm{L}})} & \text{if } i \in I \\ \\ 0 & \text{if } i \notin I, \end{cases}$$

and a scalar $\tilde{c} \in \mathbb{R}$ for which

$$\tilde{c} := \tilde{y}_0 - \varepsilon - \sum_{i \in I} \left(\frac{\tilde{y}_{+i} + \tilde{y}_{-i} - 2\tilde{y}_0 + 4\varepsilon}{2\alpha_i} \right).$$

Then $f(\mathbf{x}) \geq \tilde{c} + \langle \tilde{\mathbf{b}}, \mathbf{x} - \mathbf{w}^{(0)} \rangle$ for each $\mathbf{x} \in X$.

Proof. Consider the following affine functions $f^{\mathrm{aff}}, \tilde{f}^{\mathrm{aff}} : [\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}] \to \mathbb{R}$:

$$f^{\mathrm{aff}} : \mathbf{x} \mapsto c + \langle \mathbf{b}, \mathbf{x} - \mathbf{w}^{(0)} \rangle \quad \text{and} \quad \tilde{f}^{\mathrm{aff}} : \mathbf{x} \mapsto \tilde{c} + \langle \tilde{\mathbf{b}}, \mathbf{x} - \mathbf{w}^{(0)} \rangle.$$

According to [4, Theorem 1], f^{aff} is an affine relaxation of f on X. The claimed result will be proved by showing that

$$\tilde{f}^{\mathrm{aff}}(\mathbf{x}) \leq f^{\mathrm{aff}}(\mathbf{x}), \quad \forall \mathbf{x} \in X.$$

Now, for each $i \in I$, consider affine functions $\tilde{h}_i, h_i : [x_i^{\mathrm{L}}, x_i^{\mathrm{U}}] \to \mathbb{R}$ for which, for each $\xi \in [x_i^{\mathrm{L}}, x_i^{\mathrm{U}}]$,

$$\begin{aligned} h_i(\xi) &= \frac{y_{+i} - y_{-i}}{\alpha_i(x_i^{\mathrm{U}} - x_i^{\mathrm{L}})}(\xi - w_i^{(0)}) - \frac{y_{+i} + y_{-i} - 2y_0}{2\alpha_i}, \\ \tilde{h}_i(\xi) &= \frac{\tilde{y}_{+i} - \tilde{y}_{-i}}{\alpha_i(x_i^{\mathrm{U}} - x_i^{\mathrm{L}})}(\xi - w_i^{(0)}) - \frac{\tilde{y}_{+i} + \tilde{y}_{-i} - 2\tilde{y}_0}{2\alpha_i}. \end{aligned}$$

Observe that the mapping $h_i(\cdot) - \tilde{h}_i(\cdot)$ is also affine, and thus attains its extreme values on X at x_i^{L} and x_i^{U}. Direct evaluation yields:

$$\begin{aligned} h_i(x_i^{\mathrm{L}}) - \tilde{h}_i(x_i^{\mathrm{L}}) &= \frac{\tilde{y}_{+i} - y_{+i} + y_0 - \tilde{y}_0}{\alpha_i}, \\ h_i(x_i^{\mathrm{U}}) - \tilde{h}_i(x_i^{\mathrm{U}}) &= \frac{\tilde{y}_{-i} - y_{-i} + y_0 - \tilde{y}_0}{\alpha_i}. \end{aligned}$$

Moreover, it follows from (7.4.1) that for each $\xi \in [x_i^{\mathrm{L}}, x_i^{\mathrm{U}}]$,

$$|h_i(\xi) - \tilde{h}(\xi)| \leq \frac{2\varepsilon}{\alpha_i}. \tag{7.4.2}$$

Observe that for each $\mathbf{x} \in X$,

$$f^{\mathrm{aff}}(\mathbf{x}) - \tilde{f}^{\mathrm{aff}}(\mathbf{x}) = y_0 - \tilde{y}_0 + \varepsilon + \sum_{i \in I} \left(h_i(x_i) - \tilde{h}_i(x_i) + \tfrac{2\varepsilon}{\alpha_i} \right).$$

Moreover, since (7.4.1) and (7.4.2) hold,

$$y_0 - \tilde{y}_0 \geq -\varepsilon \quad \text{and} \quad h_i(x_i) - \tilde{h}(x_i) \geq -\frac{2\varepsilon}{\alpha_i} \quad \forall i \in I.$$

Thus,

$$f^{\mathrm{aff}}(\mathbf{x}) - \tilde{f}^{\mathrm{aff}}(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in X,$$

as required.

□

Corollary 7.4.2. Consider the setup of Corollary 7.4.1, and define a quantity:

$$\tilde{f}^{\mathrm{L}} := \tilde{y}_0 - \varepsilon - \sum_{i \in I} \left(\frac{\max(\tilde{y}_{+i}, \tilde{y}_{-i}) - \tilde{y}_0 + 2\varepsilon}{\alpha_i} \right).$$

Then $f(\mathbf{x}) \geq \tilde{f}^{\mathrm{L}}$ for each $\mathbf{x} \in X$.

Proof. Observe that $\tilde{f}^{\mathrm{L}} = \min\{\tilde{c} + \langle \tilde{\mathbf{b}}, \mathbf{x} - \mathbf{w}^{(0)} \rangle : \mathbf{x} \in X\}$. The claimed result then follows immediately from Corollary 7.4.1. □

In the limit $\varepsilon \to 0^+$, these results converge to the corresponding noise-free results [4, Theorem 1 and Corollary 1].

The proof of Corollary 7.4.1 also suggests the following way to choose the step length quantities α_i in general when the numerical error ε in evaluations of f is reasonably small. Suppose that, for some scalar $\alpha_f > 1$, an absolute numerical error of $a_f\varepsilon$ is permitted in evaluations of the affine underestimator provided by [4, Theorem 1]. Then, a similar argument to the proof of Corollary 7.4.1 shows that whenever $\alpha_i \in [\frac{3n}{\alpha_f - 1}, 1]$ for each $i \in I$, this error bound will be satisfied.

7.5 Moving the sampled set

This section adapts the above results to accommodate sampled sets W for which $\mathbf{w}^{(0)}$ is not the midpoint of X, and so W is not centered within X. One benefit of the results of this section is that these allow constructing piecewise-affine underestimators via black-box sampling, by applying the sampling-based affine relaxation approach to multiple $\mathbf{w}^{(0)}$s on a box domain. As shown in [44], piecewise-affine underestimators, constructed as point-wise maximum of multiple affine underestimators, are efficient to evaluate and effective for lower bounding in global optimization. Throughout this section, consider the following variant of Definition 7.2.2.

Definition 7.5.1. Consider the problem setup in Definition 7.2.1, and define the following additional sets and quantities:

- an index set $I := \{i \in \{1,\ldots,n\} : x_i^{\mathrm{L}} < x_i^{\mathrm{U}}\}$,

- the midpoint $\mathbf{x}^{\mathrm{mid}} := \frac{1}{2}(\mathbf{x}^{\mathrm{L}} + \mathbf{x}^{\mathrm{U}})$ of X,

- a scaled displacement $\lambda_i \in (-1,1)$ for each $i \in I$,

- a central sampled point $\mathbf{w}^{(0)} \in X$ for which, for each $i \in I$,

$$w_i^{(0)} := x_i^{\mathrm{mid}} + \tfrac{1}{2}\lambda_i(x_i^{\mathrm{U}} - x_i^{\mathrm{L}}),$$

- for each $i \in I$, a step length $\alpha_i \in (0, 1-|\lambda_i|]$ and vectors

$$\mathbf{w}^{(\pm i)} := \mathbf{w}^{(0)} \pm \frac{\alpha_i}{2}(x_i^{\mathrm{U}} - x_i^{\mathrm{L}})\,\mathbf{e}^{(i)} \in X,$$

- function values $y_0 := f(\mathbf{w}^{(0)})$ and $y_{\pm i} := f(\mathbf{w}^{(\pm i)})$ for each $i \in I$,

- a vector $\mathbf{b} \in \mathbb{R}^n$ for which, for each $i \in \{1,\ldots,n\}$:

$$b_i := \begin{cases} \dfrac{y_{+i} - y_{-i}}{\alpha_i(x_i^{\mathrm{U}} - x_i^{\mathrm{L}})} = \dfrac{y_{+i} - y_{-i}}{\|\mathbf{w}^{(+i)} - \mathbf{w}^{(-i)}\|} & \text{if } i \in I \\ 0 & \text{if } i \notin I, \end{cases}$$

- a scalar $c \in \mathbb{R}$ for which

$$c := y_0 - \sum_{i \in I}\left(\frac{(1+|\lambda_i|)(y_{+i} + y_{-i} - 2y_0)}{2\alpha_i}\right),$$

- and a scalar $f^{\mathrm{L}} \in \mathbb{R}$ for which

$$f^{\mathrm{L}} := y_0 - \sum_{i \in I}\left(\frac{(1+|\lambda_i|)(\max(y_{+i}, y_{-i}) - y_0)}{\alpha_i}\right).$$

Under this modified setup, a guaranteed affine underestimator and lower bound may still be generated as before, and these results may be improved as before in the one-dimensional case.

Corollary 7.5.2. Consider the problem setup in Definition 7.2.1 and the auxiliary quantities in Definition 7.5.1. For each $\mathbf{x} \in X$,

$$f(\mathbf{x}) \geq c + \langle \mathbf{b}, \mathbf{x} - \mathbf{w}^{(0)} \rangle \geq f^{\mathrm{L}}.$$

Proof. This result follows from an analogous argument to the proofs of [4, Theorem 1 and Corollary 1]. □

Corollary 7.5.3. Consider the problem setup in Definition 7.2.1 and the auxiliary quantities in Definition 7.5.1. Suppose that $n = 1$, and define another quantity

$$\hat{c} := 2y_0 - \tfrac{1}{2}(y_{+1} + y_{-1}).$$

Then $c \leq \hat{c} \leq y_0$. Moreover, for each $x \in X$, $f(x) \geq \hat{c} + b_1(x - w^{(0)})$.

Define another quantity

$$\begin{aligned} \hat{f}^{\mathrm{L}} := \min\{&2y_0 - y_{+1}, \quad 2y_0 - y_{-1}, \\ &\frac{2(y_{-1} - y_0)}{\alpha_1(x_1^{\mathrm{U}} - x_1^{\mathrm{L}})}(w_1^{(0)} - x_1^{\mathrm{L}}) + y_0, \quad \frac{2(y_{+1} - y_0)}{\alpha_1(x_1^{\mathrm{U}} - x_1^{\mathrm{L}})}(x_1^{\mathrm{U}} - w_1^{(0)}) + y_0\Big\}. \end{aligned}$$

Then $f^{\mathrm{L}} \leq \hat{f}^{\mathrm{L}} \leq y_0$. Moreover, for each $\mathbf{x} \in X$, $f(\mathbf{x}) \geq \hat{f}^{\mathrm{L}}$.

Proof. This result follows from an analogous argument to the proofs of [4, Theorem 2 and Corollary 2]. □

An analogous argument to the proof of Theorem 7.3.1 shows that the affine underestimators and lower bounds provided by Corollary 7.5.2 become tighter as any α_i is decreased.

Theorem 7.3.4 still holds with Definition 7.5.1 in place of Definition 7.2.2, with the additional requirement that there exists a scalar $0 < \gamma < 1$ for which

$$|\lambda_i^X| \leq \gamma, \qquad \forall X \in \mathbb{I}Q, \quad \forall i \in I^X.$$

This requirement is necessary to apply the result [46, Theorem 4.2] where required in the theorem's proof. Roughly, this avoids problems due to convex functions potentially behaving oddly near their domains' boundaries. Hence, with this requirement enforced, second-order pointwise convergence is still inherited by the sampled affine underestimators even when $\mathbf{w}^{(0),X}$ is not the midpoint of X.

Lastly, uncertainty in the function evaluations in Definition 7.5.1 may be handled as follows.

Corollary 7.5.4. Consider the setup of Corollary 7.5.2, and suppose that there exist values $\varepsilon > 0$, $\tilde{y}_0 \in \mathbb{R}$, and $\tilde{y}_{\pm i} \in \mathbb{R}$ for each $i \in I$, for which:

$$|\tilde{y}_0 - y_0| \leq \varepsilon, \quad |\tilde{y}_{+i} - y_{+i}| \leq \varepsilon, \quad \text{and} \quad |\tilde{y}_{-i} - y_{-i}| \leq \varepsilon \qquad \forall i \in I.$$

Define a vector $\tilde{\mathbf{b}} \in \mathbb{R}^n$ so that for each $i \in \{1, \ldots, n\}$,

$$\tilde{b}_i := \begin{cases} \dfrac{\tilde{y}_{+i} - \tilde{y}_{-i}}{\alpha_i (x_i^{\mathrm{U}} - x_i^{\mathrm{L}})} & \text{if } i \in I \\ 0 & \text{if } i \notin I, \end{cases}$$

and a scalar $\tilde{c} \in \mathbb{R}$ for which

$$\tilde{c} := \tilde{y}_0 - \varepsilon - \sum_{i \in I} \left(\frac{(1 + |\lambda_i|)(\tilde{y}_{+i} + \tilde{y}_{-i} - 2\tilde{y}_0 + 4\varepsilon)}{2\alpha_i} \right).$$

Then $f(\mathbf{x}) \geq \tilde{c} + \langle \tilde{\mathbf{b}}, \mathbf{x} - \mathbf{w}^{(0)} \rangle$ for each $\mathbf{x} \in X$. Moreover, for each $\mathbf{x} \in X$,

$$f(\mathbf{x}) \geq \tilde{y}_0 - \varepsilon - \sum_{i \in I} \left(\frac{(1 + |\lambda_i|)(\max(\tilde{y}_{+i}, \tilde{y}_{-i}) - \tilde{y}_0 + 2\varepsilon)}{\alpha_i} \right).$$

Proof. This result follows from an analogous argument to the proofs of Corollaries 7.4.1 and 7.4.2. □

7.6 Conclusions and future work

This work shows that, given a convex function of n variables on a box, a correct closed-form affine underestimator of this function may be constructed by sampling the function $(2n+1)$ times in a star-shaped stencil W. Such affine underestimator is particularly helpful for using newly-developed convex relaxations whose subgradients are not yet available in global optimization. Subsequent results show that we can also compute an analogous lower bound, and improve these results further when $n = 1$. Variants of these methods permit explicit consideration of noise in the black-box function evaluations. The new affine underestimators also inherit second-order pointwise convergence from an underlying scheme of convex relaxations. Refer to [4, Section 4] for implementations and case studies which employ the new affine underestimators in deterministic algorithms of global optimization.

The results in this chapter depend heavily on convexity of the sampled system; convexity allows us to infer the global behavior of the considered function from samples that are restricted to certain search directions. It may be possible to use these new underestimators in a derivative-free method for convex optimization. However, we do not expect this approach to generalize in a useful way to black-box nonconvex process models, beyond the finite differencing results of [130].

Chapter 8

Conclusions and Future Work

8.1 Conclusions

This thesis has proposed novel approaches for computing convex relaxations with the corresponding subgradients for the solution of the parametric ODE system (2.3.1), to ultimately improve computational efficiency of deterministic global dynamic optimization. This work is based on a state-of-the-art ODE relaxation framework (2.4.1) by Scott and Barton [2], and achieve the various goals of this thesis by resolving the limitations of this framework summarized in Section 1.3.

Firstly, in the Scott–Barton framework, it was previously unknown whether the tightness of the original right-hand side's relaxations translates into tightness of ODE relaxations. Hence, Chapter 3 proposes new ODE bounding and comparison results, which have significantly less stringent conditions than the previously established results. These new results are then applied to show for the first time that in the Scott–Barton framework, tighter relaxations of the original right-hand side will necessarily lead to ODE relaxations that are at least as tight.

Secondly, since only generalized McCormick relaxations were previously allowed to be used in the Scott–Barton framework, Chapter 4 proposes a new versatile optimization-based (OB) ODE relaxation formulation by using Scott–Barton relaxation theory in a new way. This new formulation allows using any convex and concave relaxations of the original right-hand side, and tighter such relaxations necessarily lead to ODE relaxations that are at least as tight, as indicated by the tightness results developed in Chapter 3. Notably, if McCormick relaxations are applied in the new formulation, then the resulting ODE relaxations are guaranteed to be at least as tight as the SBM relaxations [2]. Such tightness is shown to lead to fewer iterations required in branch-and-bound in a global dynamic optimization case study. An efficient evaluation method for the OB relaxations is described in Chapter 5, provided that the employed right-hand side relaxations have known monotonicity. Chapter 5 also proposes a new AVM-based ODE relaxation formulation based on the OB formulation. The AVM-based formulation effectively handles a factorable original right-hand side function using the Auxiliary Variable Method (AVM) [8, 50], which is shown to yield ODE relaxations that are at least as tight as both the SBM relaxations and the OBM relaxations.

Thirdly, since there was previously no subgradient evaluation approach for ODE relaxations obtained using Scott–Barton framework, Chapter 6 proposes new subgradient evaluation theory, and proposes new forward sensitivity methods for evaluating subgradients of the SBM relaxations and the OB relaxations in the Scott–Barton framework. These methods assume that the ODE relaxations are strictly within the predefined state bounds during ODE solving, which is guaranteed to be satisfied for a sufficiently small domain of the uncertain parameters. Moreover, it is shown for the first time that the dynamic subgradients may be computed efficiently using a modified adjoint ODE sensitivity system, which would improve computational efficiency of the lower-bounding procedure in global dynamic optimization.

Lastly, Chapter 7 proposes a new approach for bounding convex functions from below via black-box sampling. This new approach allows computing bounding information in global optimization using convex relaxations with unknown subgradients. The new bounds are shown to have second-order pointwise convergence [36] to an original nonconvex function as domain shrinks, which may help mitigate the cluster effect [37, 38] in global optimization. Numerical error or noise in the sampling procedure can be easily handled in the new approach.

8.2 Future work

As introduced in Section 4.4.1, the current proof-of-concept implementation of OB relaxations numerically solves convex NLPs for each right-hand side evaluation, which may lead to prohibitively expensive computational effort. In branch-and-bound deterministic global optimization algorithms, both tightness and efficiency of convex relaxations are extremely important. If a convex relaxation is promisingly tight but very difficult to evaluate, then this may still lead to expensive overall computational effort for an overarching global optimization method. Thus, a first avenue for future work would be developing more efficient implementations for the OB relaxations proposed in Chapter 4, using the techniques outlined in Section 4.3.7. I believe that these techniques would indeed drastically improve computational efficiency for evaluating the OB relaxations. From numerical experiments, the active constraints of the right-hand side optimization problems typically only change several times when computing the OB relaxations. This implies that if these active constraints are effectively tracked during ODE solving, then at most time steps, the right-hand side of OB relaxation system may be efficiently evaluated in closed form. Similar behaviour is also observed for the AVM-based ODE relaxations in Chapter 5.

A second avenue for future work would be developing improved state bounds $(\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}})$ in Definition 2.4.2, which are employed in the Scott–Barton framework (2.4.1). As revealed by the tightness results developed in Chapter 3, the tightness of these bounds necessarily translates into the tightness of ODE relaxations. Moreover, if the employed state bounding method is not applicable to the original ODE system (2.3.1), then all approaches in the Scott–Barton framework, including the SBM relaxation approach, OB relaxation approach, and AVM-based relaxation approach, cannot yield valid relaxations. For example, Harrison's state bounding method [69] is prevalent over the past decades for its efficiency and simplicity of implementation, and is employed in many dynamic reachable-set generation researches [2, 3, 13, 75, 76, 78, 93, 104, 159]. However, for oscillating systems such as the Lotka–Volterra system and the Lorenz attractor system, Harrison's state bounds explode very easily. Thus, no ODE relaxation approaches based on Harrison's bounds would work for these systems. Note that there are several recent advances [70,72,74] for state bounding. It is also encouraged to use these bounds to construct ODE relaxations.

A third avenue for future work would be developing an implementation for the dynamic adjoint subgradient evaluation proposed in Chapter 6. Unfortunately, due to time limitation, the work in Chapter 6 only provides underlying mathematics for adjoint subgradient evaluation. Considering the success of adjoint sensitivity analysis for smooth dynamic systems, it is reasonable to expect that such adjoint subgradients could significantly improve computational efficiency for lower bounding in global dynamic optimization. Besides, it will be promising to develop new subgradient evaluation methods for the Scott–Barton framework, without assuming that the state relaxations are strictly within the state bounds; this seems to be difficult due to the switching conditions in (2.4.1). Currently, my colleague Huiyi Cao is working on eliminating the switching conditions using a safe-landing mechanism.

The ultimate long-term goal for this line of research is to use deterministic global dynamic optimization techniques to solve problems in engineering applications. In the current stage, by using our ongoing Julia global dynamic optimization implementation with the new ODE relaxation and subgradient evaluation techniques embedded, we are already able to solve several benchmark problems of dynamic parameter estimation and global optimal control in [152]. Here, I would suggest for researchers in this field to always publish a usable version of their methods along with their publications. This will save much time for researchers when reproducing others' methods for comparison. Since significant state bounding and relaxation research has been conducted over the past decades, I believe that facilitating easy comparison between competing methods is important for the community to foster future advances in this field.

Bibliography

[1] I. Papamichail and C. S. Adjiman, "A rigorous global optimization algorithm for problems with ordinary differential equations," *Journal of Global Optimization*, vol. 24, no. 1, pp. 1–33, 2002.

[2] J. K. Scott and P. I. Barton, "Improved relaxations for the parametric solutions of ODEs using differential inequalities," *Journal of Global Optimization*, vol. 57, no. 1, pp. 143–176, 2013.

[3] Y. Song and K. A. Khan, "Optimization-based convex relaxations for nonconvex parametric systems of ordinary differential equations," *Mathematical Programming*, 2021. In press, DOI: https://doi.org/10.1007/s10107-021-01654-x.

[4] Y. Song, H. Cao, C. Mehta, and K. A. Khan, "Bounding convex relaxations of process models from below by tractable black-box sampling," *Computers & Chemical Engineering*, 2021. In press, DOI: https://doi.org/10.1016/j.compchemeng.2021.107413.

[5] G. P. McCormick, "Computability of global solutions to factorable nonconvex programs: Part I-Convex underestimating problems," *Mathematical Programming*, vol. 10, no. 1, pp. 147–175, 1976.

[6] J. K. Scott, M. D. Stuber, and P. I. Barton, "Generalized McCormick relaxations," *Journal of Global Optimization*, vol. 51, no. 4, pp. 569–606, 2011.

[7] A. Tsoukalas and A. Mitsos, "Multivariate McCormick relaxations," *Journal of Global Optimization*, vol. 59, no. 2-3, pp. 633–662, 2014.

[8] E. M. Smith and C. C. Pantelides, "A symbolic reformulation/spatial branch-and-bound algorithm for the global optimisation of nonconvex MINLPs," *Computers & Chemical Engineering*, vol. 23, no. 4-5, pp. 457–478, 1999.

[9] C. S. Adjiman, S. Dallwig, C. A. Floudas, and A. Neumaier, "A global optimization method, αBB, for general twice-differentiable constrained NLPs-I. Theoretical Advances," *Computers & Chemical Engineering*, vol. 22, no. 9, pp. 1137–1158, 1998.

[10] B. Srinivasan, S. Palanki, and D. Bonvin, "Dynamic optimization of batch processes: I. characterization of the nominal solution," *Computers & Chemical Engineering*, vol. 27, no. 1, pp. 1–26, 2003.

[11] E. F. Carrasco and J. R. Banga, "Dynamic optimization of batch reactors using adaptive stochastic algorithms," *Industrial & Engineering Chemistry Research*, vol. 36, no. 6, pp. 2252–2261, 1997.

[12] B. Houska and B. Chachuat, "Branch-and-lift algorithm for deterministic global optimization in nonlinear optimal control," *Journal of Optimization Theory and Applications*, vol. 162, no. 1, pp. 208–248, 2014.

[13] H. Diedam and S. Sager, "Global optimal control with the direct multiple shooting method," *Optimal Control Applications and Methods*, vol. 39, no. 2, pp. 449–470, 2018.

[14] R. Luus, J. Dittrich, and F. J. Keil, “Multiplicity of solutions in the optimization of a bifunctional catalyst blend in a tubular reactor,” *The Canadian Journal of Chemical Engineering*, vol. 70, no. 4, pp. 780–785, 1992.

[15] F. Castiglione and B. Piccoli, “Cancer immunotherapy, mathematical modeling and optimal control,” *Journal of theoretical Biology*, vol. 247, no. 4, pp. 723–732, 2007.

[16] R. B. Martin, “Optimal control drug scheduling of cancer chemotherapy,” *Automatica*, vol. 28, no. 6, pp. 1113–1123, 1992.

[17] M. Čižniar, M. Podmajerskỳ, T. Hirmajer, M. Fikar, and A. M. Latifi, “Global optimization for parameter estimation of differential-algebraic systems,” *Chemical Papers*, vol. 63, no. 3, pp. 274–283, 2009.

[18] Y. Lin and M. A. Stadtherr, “Deterministic global optimization for parameter estimation of dynamic systems,” *Industrial & Engineering Chemistry Research*, vol. 45, no. 25, pp. 8438–8448, 2006.

[19] A. B. Singer, J. W. Taylor, P. I. Barton, and W. H. Green, “Global dynamic optimization for parameter estimation in chemical kinetics,” *The Journal of Physical Chemistry A*, vol. 110, no. 3, pp. 971–976, 2006.

[20] H. Huang, C. S. Adjiman, and N. Shah, “Quantitative framework for reliable safety analysis,” *AIChE Journal*, vol. 48, no. 1, pp. 78–96, 2002.

[21] J. H. Holland *et al.*, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT press, Cambridge, 1992.

[22] R. Storn and K. Price, “Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces,” *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[23] R. Eberhart and J. Kennedy, “Particle swarm optimization,” in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, Citeseer, 1995.

[24] J. R. Banga and W. D. Seider, “Global optimization of chemical processes using stochastic algorithms,” in *State of the Art in Global Optimization*, pp. 563–583, Springer, 1996.

[25] J. R. Banga, C. G. Moles, and A. A. Alonso, “Global optimization of bioprocesses using stochastic and hybrid methods,” in *Frontiers in Global Optimization*, pp. 45–70, Springer, 2004.

[26] J. A. Egea, E. Vazquez, J. R. Banga, and R. Martí, “Improved scatter search for the global optimization of computationally expensive dynamic models,” *Journal of Global Optimization*, vol. 43, no. 2-3, pp. 175–190, 2009.

[27] J. E. Falk and R. M. Soland, “An algorithm for separable nonconvex programming problems,” *Management Science*, vol. 15, no. 9, pp. 550–569, 1969.

[28] A. Khajavirad and N. V. Sahinidis, “A hybrid LP/NLP paradigm for global optimization relaxations,” *Mathematical Programming Computation*, vol. 10, no. 3, pp. 383–421, 2018.

[29] H. S. Ryoo and N. V. Sahinidis, “A branch-and-reduce approach to global optimization,” *Journal of Global Optimization*, vol. 8, no. 2, pp. 107–138, 1996.

[30] M. Tawarmalani and N. V. Sahinidis, *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications*. Springer, Dordrecht, 2002.

[31] M. Tawarmalani and N. V. Sahinidis, "A polyhedral branch-and-cut approach to global optimization," *Mathematical programming*, vol. 103, no. 2, pp. 225–249, 2005.

[32] R. Misener and C. A. Floudas, "ANTIGONE: algorithms for continuous/integer global optimization of nonlinear equations," *Journal of Global Optimization*, vol. 59, no. 2-3, pp. 503–526, 2014.

[33] K. A. Khan, H. A. Watson, and P. I. Barton, "Differentiable McCormick relaxations," *Journal of Global Optimization*, vol. 67, no. 4, pp. 687–729, 2017.

[34] J. Najman and A. Mitsos, "Tighter McCormick relaxations through subgradient propagation," *Journal of Global Optimization*, vol. 75, no. 3, pp. 565–593, 2019.

[35] A. Mitsos, B. Chachuat, and P. I. Barton, "McCormick-based relaxations of algorithms," *SIAM Journal on Optimization*, vol. 20, no. 2, pp. 573–601, 2009.

[36] A. Bompadre and A. Mitsos, "Convergence rate of McCormick relaxations," *Journal of Global Optimization*, vol. 52, no. 1, pp. 1–28, 2012.

[37] K. Du and R. B. Kearfott, "The cluster problem in multivariate global optimization," *Journal of Global Optimization*, vol. 5, no. 3, pp. 253–265, 1994.

[38] A. Wechsung, S. D. Schaber, and P. I. Barton, "The cluster problem revisited," *Journal of Global Optimization*, vol. 58, no. 3, pp. 429–438, 2014.

[39] J. K. Scott, *Reachability Analysis and Deterministic Global Optimization of Differential-Algebraic Systems*. PhD thesis, Massachusetts Institute of Technology, 2012.

[40] Y. Nesterov, *Lectures on Convex Optimization*. Springer, Cham, 2018.

[41] K. C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*. Springer, Berlin, 1985.

[42] C. Lemaréchal, J.-J. Strodiot, and A. Bihain, "On a bundle algorithm for nonsmooth optimization," in *Nonlinear Programming 4* (O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds.), pp. 245–282, Academic Press, New York, 1981.

[43] L. Lukšan and J. Vlček, "A bundle-Newton method for nonsmooth unconstrained minimization," *Mathematical Programming*, vol. 83, no. 1, pp. 373–391, 1998.

[44] H. Cao, Y. Song, and K. A. Khan, "Convergence of subtangent-based relaxations of nonlinear programs," *Processes*, vol. 7, no. 4, p. 221, 2019.

[45] J. Najman, D. Bongartz, and A. Mitsos, "Linearization of mccormick relaxations and hybridization with the auxiliary variable method," *Journal of Global Optimization*, pp. 1–26, 2021.

[46] K. A. Khan, "Subtangent-based approaches for dynamic set propagation," in *Proceedings of the 57th IEEE Conference on Decision and Control*, December 2018.

[47] J. K. Scott and P. I. Barton, "Convex enclosures for the reachable sets of nonlinear parametric ordinary differential equations," in *49th IEEE Conference on Decision and Control (CDC)*, pp. 5695–5700, IEEE, 2010.

[48] R. E. Moore, *Methods and Applications of Interval Analysis*. SIAM, Philadelphia, 1979.

[49] C. S. Adjiman, I. P. Androulakis, and C. A. Floudas, "A global optimization method, αBB, for general twice-differentiable constrained NLPs—II. Implementation and computational results," *Computers & Chemical Engineering*, vol. 22, no. 9, pp. 1159–1179, 1998.

[50] E. M. Smith and C. C. Pantelides, "Global optimisation of nonconvex MINLPs," *Computers & Chemical Engineering*, vol. 21, pp. S791–S796, 1997.

[51] C. D. Maranas and C. A. Floudas, "Finding all solutions of nonlinearly constrained systems of equations," *Journal of Global Optimization*, vol. 7, no. 2, pp. 143–182, 1995.

[52] L. Liberti and C. C. Pantelides, "Convex envelopes of monomials of odd degree," *Journal of Global Optimization*, vol. 25, no. 2, pp. 157–168, 2003.

[53] A. Khajavirad and N. V. Sahinidis, "Convex envelopes of products of convex and component-wise concave functions," *Journal of Global Optimization*, vol. 52, no. 3, pp. 391–409, 2012.

[54] A. Khajavirad and N. V. Sahinidis, "Convex envelopes generated from finitely many compact convex sets," *Mathematical Programming*, vol. 137, no. 1, pp. 371–408, 2013.

[55] M. Tawarmalani and N. V. Sahinidis, "Semidefinite relaxations of fractional programs via novel convexification techniques," *Journal of Global Optimization*, vol. 20, no. 2, pp. 133–154, 2001.

[56] M. Tawarmalani and N. V. Sahinidis, "Convex extensions and envelopes of lower semi-continuous functions," *Mathematical Programming*, vol. 93, no. 2, pp. 247–263, 2002.

[57] W. R. Esposito and C. A. Floudas, "Global optimization for the parameter estimation of differential-algebraic systems," *Industrial & Engineering Chemistry Research*, vol. 39, no. 5, pp. 1291–1310, 2000.

[58] M. E. Wilhelm, A. V. Le, and M. D. Stuber, "Global optimization of stiff dynamical systems," *AIChE Journal*, vol. 65, no. 12, p. e16836, 2019.

[59] X. Yang and J. K. Scott, "Efficient reachability bounds for discrete-time nonlinear systems by extending the continuous-time theory of differential inequalities," in *2018 Annual American Control Conference (ACC)*, pp. 6242–6247, IEEE, 2018.

[60] N. V. Sahinidis, "BARON: A general purpose global optimization software package," *Journal of Global Optimization*, vol. 8, no. 2, pp. 201–205, 1996.

[61] M. D. Stuber, J. K. Scott, and P. I. Barton, "Convex and concave relaxations of implicit functions," *Optimization Methods and Software*, vol. 30, no. 3, pp. 424–460, 2015.

[62] Y. Lin and M. A. Stadtherr, "Validated solutions of initial value problems for parametric ODEs," *Applied Numerical Mathematics*, vol. 57, no. 10, pp. 1145–1162, 2007.

[63] N. S. Nedialkov, K. R. Jackson, and G. F. Corliss, "Validated solutions of initial value problems for ordinary differential equations," *Applied Mathematics and Computation*, vol. 105, no. 1, pp. 21–68, 1999.

[64] A. M. Sahlodin and B. Chachuat, “Discretize-then-relax approach for state relaxactions in global dynamic optimization,” in *Computer Aided Chemical Engineering*, vol. 28, pp. 427–432, Elsevier, 2010.

[65] A. M. Sahlodin and B. Chachuat, “Discretize-then-relax approach for convex/concave relaxations of the solutions of parametric ODEs,” *Applied Numerical Mathematics*, vol. 61, no. 7, pp. 803–820, 2011.

[66] A. M. Sahlodin and B. Chachuat, “Convex/concave relaxations of parametric ODEs using Taylor models,” *Computers & Chemical Engineering*, vol. 35, no. 5, pp. 844–857, 2011.

[67] C. Pérez-Galván and I. D. L. Bogle, “Global optimisation for dynamic systems using interval analysis,” *Computers & Chemical Engineering*, vol. 107, pp. 343–356, 2017.

[68] A. Bompadre, A. Mitsos, and B. Chachuat, “Convergence analysis of Taylor models and McCormick-Taylor models,” *Journal of Global Optimization*, vol. 57, no. 1, pp. 75–114, 2013.

[69] G. Harrison, “Dynamic models with uncertain parameters,” in *Proceedings of the 1st International Conference on Mathematical Modeling* (X. Avula, ed.), vol. 1, pp. 295–304, 1977.

[70] S. M. Harwood and P. I. Barton, “Efficient polyhedral enclosures for the reachable set of nonlinear control systems,” *Mathematics of Control, Signals, and Systems*, vol. 28, no. 1, p. 8, 2016.

[71] S. M. Harwood and P. I. Barton, "Affine relaxations for the solutions of constrained parametric ordinary differential equations," *Optimal Control Applications and Methods*, vol. 39, no. 2, pp. 427–448, 2018.

[72] S. M. Harwood, J. K. Scott, and P. I. Barton, "Bounds on reachable sets using ordinary differential equations with linear programs embedded," *IMA Journal of Mathematical Control and Information*, vol. 33, no. 2, pp. 519–541, 2016.

[73] J. K. Scott and P. I. Barton, "Tight, efficient bounds on the solutions of chemical kinetics models," *Computers & Chemical Engineering*, vol. 34, no. 5, pp. 717–731, 2010.

[74] J. K. Scott and P. I. Barton, "Bounds on the reachable sets of nonlinear control systems," *Automatica*, vol. 49, no. 1, pp. 93–100, 2013.

[75] J. K. Scott, B. Chachuat, and P. I. Barton, "Nonlinear convex and concave relaxations for the solutions of parametric ODEs," *Optimal Control Applications and Methods*, vol. 34, no. 2, pp. 145–163, 2013.

[76] A. B. Singer and P. I. Barton, "Bounding the solutions of parameter dependent nonlinear ordinary differential equations," *SIAM Journal on Scientific Computing*, vol. 27, no. 6, pp. 2167–2182, 2006.

[77] W. Walter, *Differential and Integral Inequalities*. Springer, Berlin, 1970.

[78] I. Bajaj and M. F. Hasan, "Global dynamic optimization using edge-concave underestimator," *Journal of Global Optimization*, vol. 77, no. 3, pp. 487–512, 2020.

[79] S. D. Schaber, J. K. Scott, and P. I. Barton, "Convergence-order analysis for differential-inequalities-based bounds and relaxations of the solutions of ODEs," *Journal of Global Optimization*, vol. 73, no. 1, pp. 113–151, 2019.

[80] B. Chachuat and M. Villanueva, "Bounding the solutions of parametric ODEs: When Taylor models meet differential inequalities," in *Computer Aided Chemical Engineering*, vol. 30, pp. 1307–1311, Elsevier, 2012.

[81] M. E. Villanueva, B. Houska, and B. Chachuat, "Unified framework for the propagation of continuous-time enclosures for parametric nonlinear ODEs," *Journal of Global Optimization*, vol. 62, no. 3, pp. 575–613, 2015.

[82] A. Kurzhanski and P. Varaiya, "Reachability analysis for uncertain systems-the ellipsoidal technique," *Dynamics of Continuous Discrete and Impulsive Systems Series B*, vol. 9, pp. 347–368, 2002.

[83] M. Althoff, O. Stursberg, and M. Buss, "Reachability analysis of nonlinear systems with uncertain parameters using conservative linearization," in *2008 47th IEEE Conference on Decision and Control*, pp. 4042–4048, IEEE, 2008.

[84] A. Chutinan and B. H. Krogh, "Verification of polyhedral-invariant hybrid automata using polygonal flow pipe approximations," in *International Workshop on Hybrid Systems: Computation and Control*, pp. 76–90, Springer, 1999.

[85] T. He and M. Tawarmalani, "A new framework to relax composite functions in nonlinear programs," *Mathematical Programming*, pp. 1–40, 2020.

[86] J.-P. Aubin, *Viability Theory*. Springer, New York, 2009.

[87] J. Szarski, *Differential Inequalities*. Polish Scientific Publishers, Warsaw, 1965.

[88] V. Lakshmikantham and S. Leela, *Differential and Integral Inequalities: Theory and Applications: Volume I: Ordinary Differential Equations*. Academic Press, New York, 1969.

[89] M. Müller, "Über das fundamentaltheorem in der theorie der gewöhnlichen differentialgleichungen," *Mathematische Zeitschrift*, vol. 26, no. 1, pp. 619–645, 1927.

[90] Y. Cao, S. Li, L. Petzold, and R. Serban, "Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution," *SIAM Journal on Scientific Computing*, vol. 24, no. 3, pp. 1076–1089, 2003.

[91] Y. Song and K. A. Khan, "Comparing solutions of related ordinary differential equations using new differential inequalities." In review.

[92] Y. Song and K. A. Khan, "Computing subgradients of convex relaxations for the solutions of parametric systems of ordinary differential equations." In review.

[93] Y. Song and K. A. Khan, "Extending optimization-based convex relaxations for dynamic process models." In review.

[94] M. E. Wilhelm and M. D. Stuber, "EAGO.jl: easy advanced global optimization in Julia," *Optimization Methods and Software*, pp. 1–26, 2020.

[95] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017.

[96] B. Chachuat, "MC++: A toolkit for bounding factorable functions." Available at: https://github.com/coin-or/MCpp.

[97] K. A. Khan and P. I. Barton, "Generalized derivatives for solutions of parametric ordinary differential equations with non-differentiable right-hand sides," *Journal of Optimization Theory and Applications*, vol. 163, no. 2, pp. 355–386, 2014.

[98] A. Wechsung and P. I. Barton, "Global optimization of bounded factorable functions with discontinuities," *Journal of Global Optimization*, vol. 58, no. 1, pp. 1–30, 2014.

[99] P. Hartman, *Ordinary Differential Equations,* 2nd edn. SIAM, Philadelphia, 2002.

[100] H. K. Khalil and J. W. Grizzle, *Nonlinear Systems*, vol. 3. Prentice Hall Press, Upper Saddle River, 2002.

[101] K. A. Khan and P. I. Barton, "Switching behavior of solutions of ordinary differential equations with abs-factorable right-hand sides," *Systems & Control Letters*, vol. 84, pp. 27–34, 2015.

[102] A. Filippov, *Differential Equations with Discontinuous Righthand Sides*. Kluwer, Dordrecht, 1988.

[103] W. F. Ames and B. Pachpatte, *Inequalities for Differential and Integral Equations*. Academic Press, London, 1998.

[104] M. Kieffer, E. Walter, and I. Simeonov, "Guaranteed nonlinear parameter estimation for continuous-time dynamical models," *IFAC Proceedings Volumes*, vol. 39, no. 1, pp. 843–848, 2006.

[105] Y. Lin and M. A. Stadtherr, "Fault detection in nonlinear continuous-time systems with uncertain parameters," *AIChE Journal*, vol. 54, no. 9, pp. 2335–2345, 2008.

[106] B. Houska, J. C. Li, and B. Chachuat, "Towards rigorous robust optimal control via generalized high-order moment expansion," *Optimal Control Applications and Methods*, vol. 39, no. 2, pp. 489–502, 2018.

[107] T. Raıssi, N. Ramdani, and Y. Candau, "Set membership state and parameter estimation for systems described by nonlinear differential equations," *Automatica*, vol. 40, no. 10, pp. 1771–1777, 2004.

[108] A. B. Singer and P. I. Barton, "Global optimization with nonlinear ordinary differential equations," *Journal of Global Optimization*, vol. 34, no. 2, pp. 159–190, 2006.

[109] I. A. Hiskens, "Stability of hybrid system limit cycles: Application to the compass gait biped robot," in *Proceedings of the 40th IEEE Conference on Decision and Control*, pp. 774–779, 2001.

[110] R. I. Leine and H. Nijmeijer, *Dynamics and Bifurcations of Non-smooth Mechanical Systems*. Springer, Berlin, 2013.

[111] P. Mhaskar, N. H. El-Farra, and P. D. Christofides, "Predictive control of switched nonlinear systems with scheduled mode transitions," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1670–1680, 2005.

[112] A. B. Singer and P. I. Barton, "Global solution of optimization problems with parameter-embedded linear dynamic systems," *Journal of Optimization Theory and Applications*, vol. 121, no. 3, pp. 613–646, 2004.

[113] MATLAB, *9.6.0.1072779 (R2019a)*. Natick, Massachusetts: The MathWorks Inc., 2019.

[114] L. Thibault, "On subdifferentials of optimal value functions," *SIAM Journal on Control and Optimization*, vol. 29, no. 5, pp. 1019–1036, 1991.

[115] C. Berge, *Topological Spaces: Including a Treatment of Multi-valued functions, Vector spaces, and Convexity*. Oliver and Boyd, Edinburgh, 1963.

[116] F. H. Clarke, "Generalized gradients and applications," *Transactions of the American Mathematical Society*, vol. 205, pp. 247–262, 1975.

[117] V. Acary and F. Pérignon, "Siconos: A software platform for modeling, simulation, analysis and control of nonsmooth dynamical systems," *Simulation Notes Europe*, vol. 17, no. 34, pp. 19–26, 2007.

[118] J. A. Gomez, K. Höffner, and P. I. Barton, "DFBAlab: a fast and reliable MATLAB code for dynamic flux balance analysis," *BMC Bioinformatics*, vol. 15, no. 1, p. 409, 2014.

[119] M. Wilhelm and M. D. Stuber, "Easy Advanced Global Optimization (EAGO): An open-source platform for robust and global optimization in Julia," in *Proceedings of 2017 AIChE Annual Meeting*, AIChE, October 2017.

[120] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, 2006.

[121] I. Dunning, J. Huchette, and M. Lubin, "JuMP: A modeling language for mathematical optimization," *SIAM Review*, vol. 59, no. 2, pp. 295–320, 2017.

[122] O. Bernard, Z. Hadj-Sadok, D. Dochain, A. Genovesi, and J.-P. Steyer, "Dynamical model development and parameter identification for an anaerobic wastewater

treatment process," *Biotechnology and Bioengineering*, vol. 75, no. 4, pp. 424–438, 2001.

[123] C. Rackauckas and Q. Nie, "Differentialequations. jl–a performant and feature-rich ecosystem for solving differential equations in Julia," *Journal of Open Research Software*, vol. 5, no. 1, 2017.

[124] H. A. Watson, M. Vikse, T. Gundersen, and P. I. Barton, "Optimization of single mixed-refrigerant natural gas liquefaction processes described by nondifferentiable models," *Energy*, vol. 150, pp. 860–876, 2018.

[125] T. Park and P. I. Barton, "State event location in differential-algebraic models," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 6, no. 2, pp. 137–165, 1996.

[126] F. A. Valentine, "A Lipschitz condition preserving extension for a vector function," *American Journal of Mathematics*, vol. 67, no. 1, pp. 83–93, 1945.

[127] S. Scholtes, *Introduction to Piecewise Differentiable Equations*. Springer, Berlin, 2012.

[128] D. Klatte and B. Kummer, "Stability properties of infima and optimal solutions of parametric optimization problems," in *Nondifferentiable Optimization: Motivations and Applications* (V. F. Demyanov and D. Pallaschke, eds.), pp. 215–229, Springer, 1985.

[129] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I: Fundamentals*, vol. 305. Springer, Berlin, 2013.

[130] Y. Yuan and K. A. Khan, "Constructing a subgradient from directional derivatives for functions of two variables," *Journal of Nonsmooth Analysis and Optimization*, vol. 1, 2020.

[131] S. Li and L. Petzold, "Software and algorithms for sensitivity analysis of large-scale differential algebraic systems," *Journal of Computational and Applied Mathematics*, vol. 125, no. 1-2, pp. 131–145, 2000.

[132] T. Maly and L. R. Petzold, "Numerical methods and software for sensitivity analysis of differential-algebraic systems," *Applied Numerical Mathematics*, vol. 20, no. 1-2, pp. 57–79, 1996.

[133] S. Galán, W. F. Feehery, and P. I. Barton, "Parametric sensitivity functions for hybrid discrete/continuous systems," *Applied Numerical Mathematics*, vol. 31, no. 1, pp. 17–47, 1999.

[134] R. Hannemann-Tamas, D. A. Munoz, and W. Marquardt, "Adjoint sensitivity analysis for nonsmooth differential-algebraic equation systems," *SIAM Journal on Scientific Computing*, vol. 37, no. 5, pp. A2380–A2402, 2015.

[135] I. Ruban, "Sensitivity coefficients for discontinuous dynamic systems," *Journal of Computer and System Sciences International*, vol. 36, pp. 536–542, 1997.

[136] F. H. Clarke, *Optimization and Nonsmooth Analysis*, vol. 5. SIAM, Philadelphia, 1990.

[137] J.-S. Pang and D. E. Stewart, "Solution dependence on initial conditions in differential variational inequalities," *Mathematical Programming*, vol. 116, no. 1-2, pp. 429–460, 2009.

[138] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York, 2007.

[139] L. Qi and J. Sun, "A nonsmooth version of Newton's method," *Mathematical Programming*, vol. 58, no. 1, pp. 353–367, 1993.

[140] M. Yunt, *Nonsmooth Dynamic Optimization of Systems with Varying Structure*. PhD thesis, Massachusetts Institute of Technology, 2011.

[141] K. A. Khan and P. I. Barton, "Generalized derivatives for hybrid systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 3193–3208, 2017.

[142] P. G. Stechlinski and P. I. Barton, "Generalized derivatives of differential–algebraic equations," *Journal of Optimization Theory and Applications*, vol. 171, no. 1, pp. 1–26, 2016.

[143] Y. Nesterov, "Lexicographic differentiation of nonsmooth functions," *Mathematical Programming*, vol. 104, no. 2-3, pp. 669–700, 2005.

[144] M. Beckers, V. Mosenkis, and U. Naumann, "Adjoint mode computation of subgradients for McCormick relaxations," in *Recent Advances in Algorithmic Differentiation* (S. Forth, P. Hovland, E. Phipps, J. Utke, and A. Walther, eds.), pp. 103–113, Springer, Berlin, 2012.

[145] Y. Yuan, *Sensitivity Analysis of Convex Relaxations for Nonsmooth Global Optimization*. Master thesis, McMaster University, 2020.

[146] R. T. Rockafellar, *Convex Analysis*. Princeton University press, Princeton, 1970.

[147] V. I. Bogachev, *Measure Theory*, vol. 1. Springer, Berlin, 2007.

[148] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317. Springer, Berlin, 2009.

[149] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, vol. 105. SIAM, Philadelphia, 2008.

[150] U. Naumann, *The Art of Differentiating Computer Programs: An Introduction to Algorithmic Differentiation*. SIAM, Philadelphia, 2012.

[151] W. Hogan, "Directional derivatives for extremal-value functions with applications to the completely convex case," *Operations Research*, vol. 21, no. 1, pp. 188–209, 1973.

[152] C. A. Floudas, P. M. Pardalos, C. Adjiman, W. R. Esposito, Z. H. Gümüs, S. T. Harding, J. L. Klepeis, C. A. Meyer, and C. A. Schweiger, *Handbook of Test Problems in Local and Global Optimization*. Springer, Dordrecht, 1999.

[153] M. A. Duran and I. E. Grossmann, "Simultaneous optimization and heat integration of chemical processes," *AIChE Journal*, vol. 32, no. 1, pp. 123–138, 1986.

[154] H. A. Watson, K. A. Khan, and P. I. Barton, "Multistream heat exchanger modeling and design," *AIChE Journal*, vol. 61, no. 10, pp. 3390–3403, 2015.

[155] A. Demissie, W. Zhu, and C. T. Belachew, "A multi-objective optimization model for gas pipeline operations," *Computers & Chemical Engineering*, vol. 100, pp. 94–103, 2017.

[156] L. M. Rios and N. V. Sahinidis, "Derivative-free optimization: a review of algorithms and comparison of software implementations," *Journal of Global Optimization*, vol. 56, no. 3, pp. 1247–1293, 2013.

[157] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*. Springer, Berlin, 2 ed., 1993.

[158] M. F. Hasan, "An edge-concave underestimator for the global optimization of twice-differentiable nonconvex problems," *Journal of Global Optimization*, vol. 71, no. 4, pp. 735–752, 2018.

[159] L. Jaulin, "Nonlinear bounded-error state estimation of continuous-time systems," *Automatica*, vol. 38, no. 6, pp. 1079–1082, 2002.

[160] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to Derivative-Free Optimization*. SIAM, Philadelphia, 2009.

[161] J. Larson, S. Leyffer, P. Palkar, and S. M. Wild, "A method for convex black-box integer global optimization," *Journal of Global Optimization*, vol. 80, pp. 439–477, 2021.

[162] J. Nocedal and S. Wright, *Numerical Optimization*. Springer, New York, 2006.