BOUNDING REACHABLE SETS FOR GLOBAL DYNAMIC OPTIMIZATION

BOUNDING REACHABLE SETS FOR GLOBAL DYNAMIC OPTIMIZATION

BY

HUIYI CAO, B.Sc., M.Sc.

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

© Copyright by Huiyi Cao, December 2021

All Rights Reserved

Doctor of Philosophy (2021) McMaster University

(Chemical Engineering) Hamilton, Ontario, Canada

TITLE:	Bounding Reachable Sets for Global Dynamic Optimization
AUTHOR:	Huiyi Cao B.Sc., M.Sc.
SUPERVISOR:	Dr. Kamil A. Khan
NUMBER OF PAGES:	xvii, 309

Abstract

Many chemical engineering applications, such as safety verification and parameter estimation, require global optimization of dynamic models. Global optimization algorithms typically require obtaining global bounding information of the dynamic system, to aid in locating and verifying the global optimum. The typical approach for providing these bounds is to generate convex relaxations of the dynamic system and minimize them using a local optimization solver. Tighter convex relaxations typically lead to tighter lower bounds, so that the number of iterations in global optimization algorithms can be reduced. To carry out this local optimization efficiently, subgradient-based solvers require gradients or subgradients to be furnished. Smooth convex relaxations would aid local optimization even more. To address these issues and improve the computational performance of global dynamic optimization, this thesis proposes several novel formulations for constructing tight convex relaxations of dynamic systems. In some cases, these relaxations are smooth.

Firstly, a new strategy is developed to generate convex relaxations of implicit functions. These convex relaxations are described by parametric programs whose

constraints are convex relaxations of the residual function. Compared with established methods for relaxing implicit functions, this new approach does not assume uniqueness of the implicit function and does not require the original residual function to be factorable. This new strategy was demonstrated to construct tighter convex relaxations in multiple numerical examples. Moreover, this new convex relaxation strategy extends to inverse functions, feasible-set mappings in constraint satisfaction problems, as well as parametric ordinary differential equations (ODEs). Using a proof-of-concept implementation in Julia, numerical examples are presented to illustrate the convex relaxations produced for various implicit functions and optimal-value functions. In certain cases, these convex relaxations are tighter than those generated with existing methods.

Secondly, a novel optimization-based framework is introduced for computing time-varying interval bounds for ODEs. Such interval bounds are useful for constructing convex relaxations of ODEs, and tighter interval bounds typically translate into tighter convex relaxations. This framework includes several established bounding approaches, but also includes many new approaches. Some of these new methods can generate tighter interval bounds than established methods, which are potentially helpful for constructing tighter convex relaxations of ODEs. Several of these approaches have been implemented in Julia.

Thirdly, a new approach is developed to improve a state-of-the-art ODE relaxation method and generate tighter and smooth convex relaxations. Unlike state-of-the-art methods, the auxiliary ODEs used in these new methods for computing convex relaxations have continuous right-hand side functions. Such continuity not only makes the new methods easier to implement, but also permits the

evaluation of the subgradients of convex relaxations. Under some additional assumptions, differentiable convex relaxations can be constructed. Besides that, it is demonstrated that the new convex relaxations are at least as tight as state-of-the-art methods, which benefits global dynamic optimization. This approach has been implemented in Julia, and numerical examples are presented.

Lastly, a new approach is proposed for generating a guaranteed lower bound for the optimal solution value of a nonconvex optimal control problem (OCP). This lower bound is obtained by constructing a relaxed convex OCP that satisfies the sufficient optimality conditions of Pontryagin's Minimum Principle. Such lower bounding information is useful for optimizing the original nonconvex OCP to a global minimum using deterministic global optimization algorithms. Compared with established methods for underestimating nonconvex OCPs, this new approach constructs tighter lower bounds. Moreover, since it does not involve any numerical approximation of the control and state trajectories, it provides lower bounds that are reliable and consistent. This approach has been implemented for control-affine systems, and numerical examples are presented.

Acknowledgements

I would like to first express my sincere gratitude to my supervisor, Dr. Kamil Khan. Thank you for your training and guidance through these years, for being there whenever I have questions, for listening to my various ideas and letting me pursue them, and for reviewing my rough drafts with patience. I would also like to thank my thesis committee members, Dr. Thomas Adams and Dr. Ned Nedialkov, who have contributed lots of helpful feedback and suggestions.

I am also thankful to McMaster Advanced Control Consortium (MACC) for providing support and a great atmosphere for my research. I am particularly grateful to my lab-mate and friend Yingkai Song. We had so many lively discussions over the years, which gave me a deeper understanding of our research.

My interest in process system engineering developed when I was a M.Sc. student at Rutgers University. I am greatly thankful to my M.Sc. supervisor, Dr. Rohit Ramachandran, for his advice and help with both my studies and my professional development.

I am so fortunate to have wonderful friends, Chenchen, Jia, and Zilong, who give me lots of encouragement. I am particularly grateful to Zilong, who has always been like a big brother and gives me advice and help whenever I need it. I also thank my parents for their unconditional love and support. Lastly, to Yexuan,

thank you for your company, patience, and everything else. Life is a long journey, but luckily, I have you by my side.

The work presented in this thesis was funded by the McMaster Advanced Control Consortium (MACC).

Contents

Abstract **iii**

Acknowledgements **vi**

1 Introduction **1**

1.1 Background 1

1.2 Motivation and goals 20

1.3 Contributions and thesis structure 24

2 Convex Relaxations of Implicit Functions **29**

2.1 Introduction 29

2.2 Background 33

2.3 Convex Relaxations of Implicit Functions 37

2.4 Convex Relaxations of Constraint Satisfaction Problems 44

2.5 Tightening Interval Bounds 51

2.6 Relaxations of Numerical ODE Solutions 53

2.7 Numerical Examples 57

2.8 Conclusion 65

3 A Smoothing Method for Generating Tighter Reachable Set Enclosures for Parametric Ordinary Differential Equations **67**

3.1 Introduction 67

3.2 Problem Formulation 72

3.3 Background 78

3.4 New State Relaxation Formulation 80

3.5 Theoretical Development 85

3.6 Implementation Considerations 112

3.7 Examples 120

3.8 Conclusion 126

4 Bounding Nonconvex Optimal Control Problems using Pontryagin's Minimum Principle **129**

4.1 Introduction 129

4.2 Problem Formulation 134

4.3 Background 136

4.4 New optimal control relaxation 144

4.5 Theoretical Development 151

4.6 Solving the Underestimating Problem 160

4.7 Numerical Examples 164

4.8 Conclusion 171

5 An Optimization-Based Framework for Enclosing Reachable Sets with Differential Inequalities **173**

5.1 Introduction 173

5.2 Preliminaries . . . 178
5.3 Problem Statement . . . 186
5.4 New Framework for Enclosing Reachable Sets . . . 187
5.5 Use Cases . . . 194
5.6 Constructing Operators Π_i^L and Π_i^U with an *a priori* Enclosure . . . 223
5.7 Numerical Examples . . . 227
5.8 Conclusions . . . 233

6 Enclosing Reachable Sets for Nonlinear Control Systems using Complementarity-Based Intervals 236

6.1 Introduction . . . 236
6.2 Problem Statement . . . 238
6.3 Background . . . 239
6.4 New Formulation . . . 241
6.5 Complementarity Reformulation . . . 244
6.6 Constructing convex inclusion functions . . . 246
6.7 Numerical Examples . . . 250
6.8 Conclusion . . . 254

7 A Differential Inequality-Based Framework for Computing Convex Enclosures of Reachable Sets 256

7.1 Introduction . . . 256
7.2 Preliminaries . . . 260
7.3 Problem Formulation . . . 261
7.4 New Framework for State Relaxation . . . 270

7.5 Constructing State Bound RHS 278
7.6 Constructing State Relaxation RHS 280
7.7 Ensuring Inclusion-amplifying Dynamics 287
7.8 Numerical Examples 295
7.9 Conclusion 302

8 Concluding Remarks **304**
8.1 Conclusions 304
8.2 Outlook 306

List of Figures

1.1 Existing approaches for local dynamic optimization 3

1.2 Existing approaches for solving optimal control problems 5

1.3 Classification of approaches for global dynamic optimization 15

1.4 Branch-and-bound algorithm 17

2.1 The implicit functions $x^{\dagger}$ and $x^{\ddagger}$ in Example 2.1 (solid), along with their interval bounds (dashed) reported in [136] and convex and concave relaxations (dotted) constructed with the new method on P, plotted as functions of p. 59

2.2 The implicit functions $x^{\dagger}$ and $x^{\ddagger}$ in Example 2.1 (solid), along with their original interval bounds $X^{\dagger,0}$ and $X^{\ddagger,0}$ (dashed) and improved interval bounds $X^{\dagger,1}$ and $X^{\ddagger,1}$ (dotted) on P, plotted as functions of p. 59

2.3 The implicit functions $x^{\dagger}$ and $x^{\ddagger}$ in Example 2.1 (solid), along with their relaxation constructed on $X^{\dagger,0}$ and $X^{\ddagger,0}$ (dashed) and improved relaxations constructed on $X^{\dagger,1}$ and $X^{\ddagger,1}$ (dotted) on P, plotted as functions of p. 60

2.4 The implicit functions $x^\dagger$ and $x^\ddagger$ in Example 2.1 (solid), along with their interval bounds (dot-dashed) and convex and concave relaxations (dashed) where the relaxations of the original residual function f are constructed with αBB relaxations, plotted as functions of p. 61

2.5 The implicit function of pressure with respect to volume in Example 2.2 (solid), along with its convex and concave relaxations (dashed) 62

2.6 The numerical solution of (2.27) via the implicit Euler method (solid) at $t = 1$, along with its convex and concave relaxations (dashed), plotted as a function of p 63

2.7 (a) Interval bounds $Z^{m,0}$ (dashed) and tighter interval bounds $Z^{m,1}$ (dotted), $m \in \{1, \ldots, 20\}$, in Example 2.3. Solid lines are trajectories of $z(\cdot, p)$ in (2.27) with different p. (b) The parametric solution of (2.21) (solid), along with its convex and concave relaxations constructed on conservative interval bounds (dashed) and improved interval bounds (dotted), plotted as a function of p at $t = 1$ 64

3.1 The parametric solution (3.40) of ODE (3.39) in Example 3.1, along its lower bound $x^L(t) = 0$ and convex relaxations described in (3.42) and (3.43), plotted as a function of p at $t = 1$ 123

3.2 The parametric solution of x_2 in Example 3.2, along with its state bounds and state relaxations, plotted as functions of p at $t = 1.2$. (a) comparison between Scott-Barton method and new method with $\boldsymbol{u}, \boldsymbol{o}$ constructed by GMR; (b) smooth relaxations generated with new method with $\boldsymbol{u}, \boldsymbol{o}$ are constructed by DMR 125

3.3 The parametric solution x_2 in Example 3.2, along with its state bounds, state relaxations, and subtangents of state relaxations, plotted as functions of p at $t = 1.2$. $k = 1$ (or 20) means that $\underline{\boldsymbol{k}}$ and $\overline{\boldsymbol{k}}$ are both set to $\mathbf{1}$ (or $\mathbf{20}$). 126

4.1 The system trajectories and the relaxed trajectory in Example 4.1: trajectories $t \mapsto \phi(x(t,u))$ (dotted) where u is a suboptimal control and trajectory $t \mapsto \phi^{\rm cv}(x(t,u^*))$ (solid) where u^* is a globally optimal control 166

4.2 The system trajectories and the relaxed trajectory in Example 4.2: trajectories $t \mapsto \phi(\boldsymbol{x}(t,u))$ (dotted) where u is a suboptimal control and trajectory $t \mapsto \phi^{\rm cv}(\boldsymbol{x}(t,u^*))$ (solid) where u^* is a globally optimal control 167

4.3 The system trajectories and the relaxed trajectories in Example 4.3: our new trajectory $t \mapsto \phi^{\rm cv}(x(t,u^*))$ (solid), the Scott-Barton trajectory $t \mapsto \hat{\phi}^{\rm cv}(x(t,\hat{u}^*))$ (dashed), and the sequence $[\tilde{\phi}^{\rm cv}(\tilde{\boldsymbol{\xi}}^i)]_{i\in\{1,\ldots,n\}}$ with $n = 10$ (star) and $n = 30$ (triangle), and trajectories $t \mapsto \phi(x(t,u))$ (dotted) for various suboptimal control inputs u 170

5.1 An illustration of the hierarchical relationship of McCormick-type relaxations 176

5.2 The state bound trajectories generated with different methods for (a) x_1 and (b) x_2 (dotted) in Example 5.1. Solid lines are real trajectories. 228

5.3 The state bound trajectories (dotted) generated with ◯-N-I (square), px-N-EC (star), and px-N-α (diamond) for x_1 in Example 5.2. Solid lines are real trajectories. 229

5.4 The state bound trajectories of S_2 in Example 5.3 (dotted) generated with no refinement (square) and refinement operators I_A (star) and I_G^B (diamond). Solid lines are real trajectories. 233
6.1 State bounds of x_1 in Example 6.1 computed by relaxing RHS functions with NIE (dotted) and αBB relaxation in (6.4) (dashed). Solid (overlapping) lines are real trajectories. 252
6.2 State bounds of x_1 in Example 6.2 computed by relaxing RHS functions with NIE (dotted) and αBB relaxation in (6.4) (dashed). Solid (overlapping) lines are real trajectories. 253
6.3 State bounds of X in Example 6.3 computed by relaxing RHS functions with NIE (dotted) and αBB relaxation in (6.4) (dashed). Solid lines are real trajectories. 254
7.1 The parametric solution of x_2 in (7.19) (black solid), along with the convex and concave relaxations generated with the Scott-Barton method (green dotted) and our new methods (violet dashed and red dashed) with $p_1 = 0$ at $t_f = 0.15s$ 298
7.2 The parametric solution of x_2 in (7.19) (black solid), along with the convex and concave relaxations generated with Scott-Barton method [120] (green dotted) and Song-Khan method [131] (blue dotted) and our new methods (pink dashed and orange dashed) with $p_1 = 0$ at $t_f = 0.15s$ 299
7.3 The parametric solution of x_2 in (7.19) (black solid), along with the differentiable convex/concave relaxations (yellow dashed) and their tangents (purple dash-dot), with $p_1 = 0$ at $t_f = 0.05s$ 299

7.4 The parametric solution of x_2 in (7.20) (black solid), along with the convex and concave relaxations generated with the Scott-Barton method [120] (green dotted) and our new methods (violet dashed and red dashed) with $K_S = 7.1$ at $t_f = 5s$ 302

List of Tables

5.1 Preview of use cases introduced in Section 5.5 195

5.2 Summary of available methods in Section 5.5 220

5.3 Computing time of Example 5.2 with different methods 230

5.4 Parameters and initial conditions for Example 5.3 232

6.1 Microbial growth process parameters 253

7.1 Summary of available methods 281

7.2 Computing times of methods shown in Figures 7.1 and 7.2 300

7.3 Microbial growth process parameters 301

7.4 Computing times of methods shown in Figures 7.4 of Example 7.2 . . 301

Chapter 1

Introduction

In this thesis, novel approaches will be presented for efficiently and automatically generate guaranteed interval bounds and convex relaxations for nonlinear process models with uncertainty, including dynamic process models. Such global bounding information is useful in the deterministic global optimization of these systems. Numerical implementations are also presented for computing these bounds automatically. The remainder of this chapter elaborates on the background, goals, and contributions of this thesis.

1.1 Background

This section summarizes established concepts and approaches that will be considered throughout the remainder of this thesis.

1.1.1 Local dynamic optimization

In chemical engineering, systems of differential equations are widely used to model the dynamic behavior of process systems, including chemical reactions [140], cell culture processes [77], and distillation columns [34]. These differential equation models enable quantitative analysis via predictive simulation. Nevertheless, many real-world engineering problems, such as parameter estimation and optimal design, may require additional computation beyond simulation. In these cases, we typically combine numerical models with optimization algorithms, for example to determine the best model parameters [117]. The problem of optimizing parameters for dynamic models is referred to as *dynamic optimization*, e.g., [153]. The mathematical formulation of a generic dynamic optimization problem with an ODE system embedded is as follows.

$$\min_{\boldsymbol{p}\in P} \quad J(\boldsymbol{p}) := \phi(\boldsymbol{x}(t_f, \boldsymbol{p}), \boldsymbol{p}), \tag{1.1}$$

where $P \subset \mathbb{R}^{n_p}$ is a box and $\boldsymbol{x}$ solves the following ODE:

$$\begin{aligned} &\dot{\boldsymbol{x}}(t, \boldsymbol{p}) = \boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{x}(t, \boldsymbol{p})), \quad t \in (t_0, t_f], \\ &\boldsymbol{x}(t_0, \boldsymbol{p}) = \boldsymbol{x}_0(\boldsymbol{p}). \end{aligned} \tag{1.2}$$

In this thesis, we do not consider dynamic optimization problems with path constraints. Broadly, this thesis focuses on developing methods to help solve (1.1) to global optimality. Dynamic optimization problems in chemical engineering may also have index-1 differential-algebraic equations (DAEs) embedded; these behave similarly to (1.2) in many respects, and we do not pursue them further.

There are three primary approaches to solve the dynamic optimization problem (1.1) to a local optimum [21], including: (i) the simultaneous approach, (ii) the sequential approach, and (iii) the hybrid approach; these are depicted in Figure 1.1, and described in the following paragraphs.

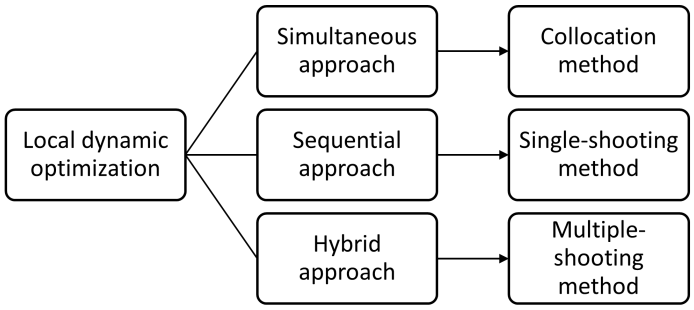


Figure 1.1: Existing approaches for local dynamic optimization

The simultaneous approach performs a discretization over the state trajectories of the ODE (1.2), and collocation techniques are commonly used here [141]. With the dynamic system approximated by a large system of nonlinear equations, the original dynamic optimization problem is transformed into a nonlinear program (NLP), and standard NLP algorithms are immediately applicable here. However, to perform an accurate approximation of the original system, the generated NLP is typically very large and might be difficult to solve.

The sequential approach involves successively solving the parametric ODE (1.2) numerically and searching for the optimal parameters of the NLP (1.1) according to the iterations of an NLP solver. Compared with the simultaneous approach, one advantage of the sequential approach is that it keeps the size of the resulting NLP relatively small. Moreover, thanks to advanced ODE solvers, numerically solving the embedded ODE (1.2) is usually efficient and accurate.

Nevertheless, there are some drawbacks of the sequential approach. One major

concern is the issue of instability. If the embedded dynamic system is not numerically stable to integrate for some parameters or initial conditions, then the whole sequential approach might fail. Since the simultaneous approach does not suffer from this problem, a hybrid of these two approaches, also known as the multiple-shooting method, was developed to reduce the potential drawback of instability [23]. In this hybrid approach, the time horizon of the original dynamic system is divided into a coarse grid. Inside each interval, a new system of ODEs with a latent initial condition is formulated. As a result, the original problem is reformulated into an NLP with several ODEs embedded. The additional latent initial values are treated as decision variables to introduce more flexibility and adaptability to the system.

1.1.2 Optimal control

An optimal control problem (OCP) involves finding a control profile for a dynamic control system that optimizes a particular objective function. Optimal control appears in many engineering applications, such as the determination of optimal control inputs of batch chemical rectors [87], the nonlinear model predictive control of continuous systems [86], and the safe landing of an autonomous spacecraft on a planet surface [2]. A typical mathematical OCP formulation is as follows:

$$\min_{\boldsymbol{u}\in\mathcal{U}} \quad J(\boldsymbol{u}) := \phi(\boldsymbol{x}(t_f,\boldsymbol{u}),\boldsymbol{u}(t_f)), \tag{1.3}$$

where $\mathcal{U}$ is the set of all admissible controls $\boldsymbol{u} : [t_0, t_f] \rightarrow W$ and $\boldsymbol{x}$ solves the following ODE:

$$\begin{aligned} \dot{\boldsymbol{x}}(t, \boldsymbol{u}) &= \boldsymbol{f}(t, \boldsymbol{u}(t), \boldsymbol{x}(t, \boldsymbol{u})), \quad t \in (t_0, t_f], \\ \boldsymbol{x}(t_0, \boldsymbol{u}) &= \boldsymbol{x}_0. \end{aligned} \tag{1.4}$$

Unlike the dynamic optimization problem (1.1) introduced in the previous section, one major challenge of solving the OCP (1.3) is that it is infinite dimensional. Established approaches for solving this problem are outlined in Figure 1.2 and are briefly summarized below.

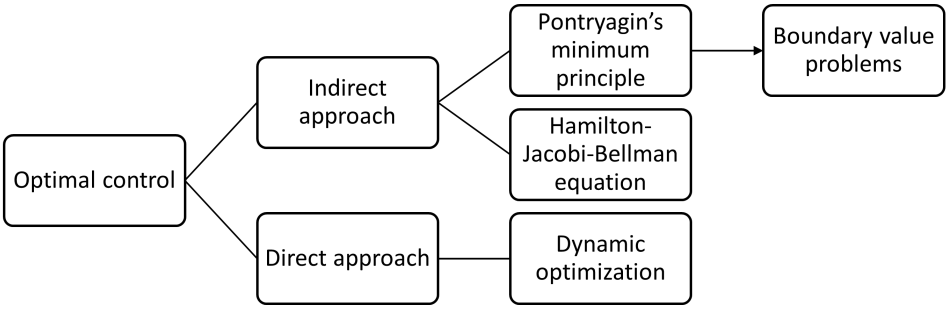


Figure 1.2: Existing approaches for solving optimal control problems

Indirect approaches for solving (1.3) address the dynamic optimization problem in its original infinite dimensional space. The Hamilton-Jacobi-Bellman (HJB) equation, developed based on the dynamic programming theory by Bellman [17], provides both necessary and sufficient optimality conditions for solving OCPs. However, the HJB equation is a partial differential equation and solving it may be quite cumbersome [95]. Pontryagin's Minimum Principle (PMP) describes a necessary condition of optimality for solving OCPs. Obtaining a local solution with

PMP usually involves solving a two-point boundary value problem either analytically or numerically. Furthermore, if the mapping from the control to the objective function is convex, then PMP is also a sufficient condition for a control input to be globally optimal [26].

In contrast, the direct approach for solving (1.3) reduces the original infinite dimensional problem to a finite dimensional problem via control parameterization. The original control trajectory $\boldsymbol{u}$ is discretized and approximated with simple functions. These functions are usually piecewise-constant functions or piecewise-affine functions that can be described by a finite vector of parameters $\boldsymbol{p}$. The original OCP (1.3) is therefore approximated as the dynamic optimization problem in (1.1). The standard dynamic optimization approaches introduced in the previous section can then be applied.

1.1.3 Convex relaxations and reachability analysis

Given a convex set $P \subset \mathbb{R}^{n_p}$, a function $h^{\text{cv}} : P \to \mathbb{R}$ is a *convex relaxation* of a nonconvex function $h : \mathbb{R}^{n_p} \to \mathbb{R}$ on P if h^{cv} is convex on P and $h^{\text{cv}}(\boldsymbol{p}) \leq h(\boldsymbol{p})$ for all $\boldsymbol{p} \in P$. Convex relaxations provide useful global intuition for nonconvex process models that is used by methods for deterministic global optimization. Since these relaxations are convex, they can be minimized using local optimization solvers. They also underestimate the original problem by construction, and so their optimal values are valid lower bounds for the original nonconvex function. Convex relaxations are used in several state-of-the-art deterministic global optimization solvers, such as BARON [139] and ANTIGONE [90]. Several approaches have been established to generate useful convex relaxations automatically; these are summarized

as follows.

If the original nonconvex function h is a finite composition of known intrinsic operations from a library, such as the operations that can be represented on a typical scientific calculator, then this function h is said to be *factorable*. Interval arithmetic [94] is an established approach for computing parameter-invariant interval bounds for factorable functions. McCormick and Mitsos et al. [89, 92] proposed a method for efficiently and automatically constructing convex relaxations for factorable functions by propagating relaxations for intrinsic operations. These relaxations are known as the McCormick relaxations (MR), and are always at least as tight as the bounds provided by interval analysis. A method for propagating the subgradients for MR was also developed by Mitsos, et al. [92]. Note that affine relaxations and piecewise-affine relaxations can be constructed by linearizing the nonlinear MR at fixed points using subgradients. Compared with nonlinear relaxations, these affine or piecewise-affine relaxations are computationally cheaper to optimize [50, 33].

Scott et al. [123] extended MR into the so-called generalized McCormick relaxations (GMR), which can additionally accept different types of inputs in various settings. GMR has the important property of taking previously known convex relaxations as arguments for further calculation, which is useful in computing relaxations for parametric ODEs [120] and implicit functions [136]. Tsoukalas and Mitsos [144] reformulated and generalized McCormick's composition theorem with an embedded optimization problem to generate tighter convex relaxations than standard MR, and to admit virtually any multivariate intrinsic operations. Khan et al.

[72] developed a smooth variant of GMR, termed as differentiable McCormick relaxations (DMR), to eliminate the theoretical and computational obstacles caused by the nonsmoothness of MR and GMR. These approaches are straightforward to implement, and have already been implemented in the open-source software packages MC++ [36] and EAGO [152].

In addition to McCormick relaxation variants, αBB relaxation is another established approach for generating convex relaxations [3]. This approach constructs convex underestimators for nonconvex twice-continuously differentiable functions by adding a sufficiently large convex quadratic term to the original function. Constructing this term typically requires estimating the curvature of the original function by bounding the eigenvalues of the original function's Hessian using interval arithmetic.

As will be discussed in Section 1.1.5 below, methods for global dynamic optimization typically require convex relaxations of the solution $\boldsymbol{x}(t_f, \cdot)$ of the ODE (1.2), which is generally unavailable in closed-form. In this case, the methods summarized above may not be applied directly to construct convex relaxations of $\boldsymbol{x}(t_f, \cdot)$. Instead, this task aligns with constructing convex enclosures for the reachable set of (1.2). The reachable set of a dynamic system is the set of possible final states that the system may attain, given a range of permitted initial conditions, parameters, or inputs [62]. Convex enclosures of the reachable set of (1.2) therefore provide useful global bounding information for finding a global solution of the dynamic optimization problem (1.1) deterministically. Several approaches have been established to construct convex relaxations that enclose the reachable set of (1.2). These methods are summarized in the following paragraphs.

Taylor series methods involve computing a validated solution (i.e. a guaranteed enclosure of the true solution) for ODEs by constructing high-order Taylor expansions of the system states x with respect to time t in discrete time steps, and then bounding the coefficients and remainder terms with interval arithmetic. To overcome the dependency problem of classic Taylor series methods [96], in which repeated terms in algebraic representations of functions can lead to significant overestimation in interval arithmetic, Taylor models were introduced in [88]. These methods bound the the Taylor remainder error by propagating an auxiliary model consisting of a Taylor polynomial and an interval remainder bound. Taylor models were used to enclose the reachable sets for parametric ODEs in [84]. They were further extended by replacing interval arithmetic with McCormick relaxations, yielding tighter enclosures in general [112]. However, Taylor series methods may be limited in computational efficiency because of the complexity of constructing and evaluating high-order Taylor expansions. The number of Taylor coefficients involved grows exponentially with the numbers of states and inputs.

Another major category of methods for enclosing reachable sets involves differential inequalities [148]. Differential inequality-based methods use an auxiliary system of ODEs obtained from the original system (1.2) to describe the corresponding reachable sets. The right-hand side (RHS) functions in the auxiliary relaxation system are modified enclosures of the original ODE RHS function. Under reasonable assumptions, the solutions of this auxiliary system are guaranteed to be component-wise lower and upper bounds for the reachable set. Such auxiliary

systems can be solved via off-the-shelf numerical ODE solvers with adaptive time-stepping, while Taylor series methods require integration procedures with manually configured step-size. Several methods in this different inequality category have been developed in the past decades. A major distinction among these methods is how the original RHS functions are handled. We now briefly review some established methods for constructing the new auxiliary RHS in chronological order. Harrison [59] used interval arithmetic [94] to construct the auxiliary RHS and computed interval bounds for the original states. A flattening technique was applied in this method to reduce the so-called wrapping effect of interval arithmetic. An affine relaxation-based method was introduced by Singer and Barton [128], in which the auxiliary RHS functions are constructed via linearizing the classic MR [89]. Although the solutions of such auxiliary systems are rigorous bounds for the original states, their existence and uniqueness are not guaranteed without additional assumptions. Scott and Barton [120] proposed a method for computing component-wise convex and concave relaxations for the final states of parametric ODEs, which are guaranteed to be at least as tight as the Harrison's interval bounds. This method will be summarized in Section 1.1.4. Harwood et al. [62] proposed a method that embeds linear programs into the auxiliary RHS functions to improve the enclosures. A special relaxation technique is used to ensure the uniqueness of ODE solutions. Moreover, Harwood et al. considered leveraging an *a priori* enclosure to reduce the conservatism in the relaxation of the original RHS functions. This strategy was further explored by Shen and Scott [126, 125, 124]. They made use of known information of the original system, including physical bounds, model redundancy, and path constraints, to further tighten the reachable

sets.

Aside from global optimization, enclosures of the reachable set are also useful when quantifying the influence of uncertainty on the dynamic system. For example, in dynamic models that describe chemical and biochemical reaction kinetics, there typically exist uncertainties in rate parameters. The inputs of these models may also contain uncertainties caused by measurement errors or system disturbances [133]. Constructing enclosures for the reachable set provides an approach to propagate the uncertainties in these dynamic models [117], so that the states of these reactions can be estimated robustly [69]. Accurate estimations on the process states further permit the development of robust process control [4]. Other applications of reachable set enclosures include parameter estimation [106], safety verification [66], and collision avoidance [91].

1.1.4 Scott-Barton relaxations of ODE solutions

One state-of-the-art method for generating convex relaxations for the parametric ODE (1.2) was proposed by Scott and Barton [120]. This method will be referred as the *Scott-Barton method* hereafter. It requires the following functions to be known in advance:

- Interval bounds $\boldsymbol{x}^L, \boldsymbol{x}^U : I \to \mathbb{R}^{n_x}$ of the state variable $\boldsymbol{x}$ such that $\boldsymbol{x}^L(t) \leq \boldsymbol{x}(t, \boldsymbol{p}) \leq \boldsymbol{x}^U(t)$ for all $t \in [t_0, t_f]$ and $\boldsymbol{p} \in P$. $\boldsymbol{x}^L, \boldsymbol{x}^U$ can typically be computed using Harrison's method from Section 1.1.3.

- Convex and concave relaxations $\boldsymbol{x}_0^{cv}, \boldsymbol{x}_0^{cc} : P \to \mathbb{R}^{n_x}$ of the initial condition function $\boldsymbol{x}_0$ over P. $\boldsymbol{x}_0^{cv}, \boldsymbol{x}_0^{cc}$ can be constructed with the convex relaxation methods from Section 1.1.3.

- Modified convex and concave relaxations $\boldsymbol{u}, \boldsymbol{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ of the RHS function $\boldsymbol{f}$, satisfying various conditions described in [120]. Scott and Barton recommend constructing $\boldsymbol{u}, \boldsymbol{o}$ by applying Harrison's flattening technique to GMR.

Then, the convex and concave relaxations $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$ of the ODE (1.2) are computed by solving the following auxiliary system of ODEs with discrete jumps: for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned}
\dot{x}_i^{cv}(t,\boldsymbol{p}) &= \begin{cases} u_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) & \text{if } x_i^{cv}(t,\boldsymbol{p}) > x_i^L(t), \\ \max\left\{\dot{x}_i^L(t), u_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p}))\right\} & \text{if } x_i^{cv}(t,\boldsymbol{p}) = x_i^L(t), \end{cases} \\
x_i^{cv}(t_0,\boldsymbol{p}) &= x_{0,i}^{cv}(\boldsymbol{p}), \\
\dot{x}_i^{cc}(t,\boldsymbol{p}) &= \begin{cases} o_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) & \text{if } x_i^{cc}(t,\boldsymbol{p}) < x_i^U(t), \\ \min\left\{\dot{x}_i^U(t), o_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p}))\right\} & \text{if } x_i^{cc}(t,\boldsymbol{p}) = x_i^U(t), \end{cases} \\
x_i^{cc}(t_0,\boldsymbol{p}) &= x_{0,i}^{cc}(\boldsymbol{p}).
\end{aligned} \tag{1.5}$$

Scott and Barton [120] showed that valid state relaxations of (1.2) are given by the unique Carathéodory solutions of (1.5). Moreover, it was verified in [113] such that, if we construct $\boldsymbol{u}, \boldsymbol{o}$ with flattened GMR [120, Section 4.4], then the Scott-Barton relaxations will have second-order pointwise convergence to $\boldsymbol{x}$ in the sense of [24]. This convergence result is critical for using state relaxations in deterministic global optimization without invoking the "cluster problem" [48, 150] in which a branch-and-bound algorithm (summarized in Section 1.1.5 below) must branch many times before terminating.

Although the solutions of (1.5) provide convex and concave relaxations for

(1.2), the if-statements in the RHS of the ODE system (1.5) will typically create discontinuity in the RHS. To numerically solve (1.5), Scott and Barton [120] proposed to use the event detection feature of CVODES [44] to handle these discontinuities, but this approach increases the difficulty of implementation and limits the use of other off-the-shelf ODE solvers. Without event detection, the numerical error resulting from the integration process will likely be worse than when solving similar ODEs with continuous RHS. Other limitations of the Scott-Barton method include the nonsmoothness of the generated convex relaxations and the difficulty of evaluating gradient information for these relaxations, again due to those discrete jumps.

1.1.5 Global optimization

Since the dynamic optimization problem in (1.1) typically nonconvex, there may be multiple suboptimal local solutions [87]. Classic local optimization algorithms may therefore terminate at a suboptimal solution without identifying a global optimum and verifying it. However, many engineering applications require optimizing nonconvex dynamic optimization problems to global optimality. Some of these are summarized below.

A first application considers analyzing the safety of chemical processes. Safety analysis involves investigating if a process system is guaranteed to operate safely, for any choice among a set of admissible control inputs, system parameters, and and initial conditions. It is required by many safety standards and regulations, including the OSHA in the United States [99], that chemical companies need to carry out systematic analyses to convincingly demonstrate the safety of their processes.

Failure in the execution of these analyses may lead to catastrophic accidents, serious damage to people's life and health, and tremendous economic cost. One established approach for solving this problem is to formulate it as a dynamic optimization problem and search for a worst-case scenario [66]. Evidently, a merely locally optimal solution cannot provide sufficient information, and may misleadingly indicate that an unsafe process is safe. A globally optimal solution of the formulated dynamic optimization problem is therefore necessary in this application.

A second application considers the economic optimization of dynamic processes. Compared with designing and controlling a process using a local suboptimal solution, computing a globally optimal solution may significantly reduce the cost and increase the process profit [80]. This advantage is particularly beneficial for the manufacturing of high-value products, such as pharmaceuticals [149] and bio-products [15].

A third application considers the validation of potential mathematical models for describing a dynamic process. The objective of this task is to determine whether the output of a model is consistent with the measurement of the underlying process, given that the model is fitted with its best-possible parameters [129]. In contrast, if only a local suboptimal solution is obtained in parameter fitting, it is doubtful to discriminate a candidate model. The poor consistency between model outputs and process measurements may be caused by the lack of suitable parameters. Thus, global dynamic optimization is useful in validating possible models for dynamic processes.

The remainder of this section provides a brief review on established approaches

for solving the following generic NLP to numerical global optimality:

$$\begin{aligned} \min_{\boldsymbol{p}\in P} \quad & J(\boldsymbol{p}), \\ \text{subject to} \quad & \boldsymbol{g}(\boldsymbol{p}) \leq \mathbf{0}. \end{aligned} \tag{1.6}$$

These established approaches can be divided into two categories: stochastic approaches and deterministic approaches. Figure 1.3 illustrates these categories and various subcategories.

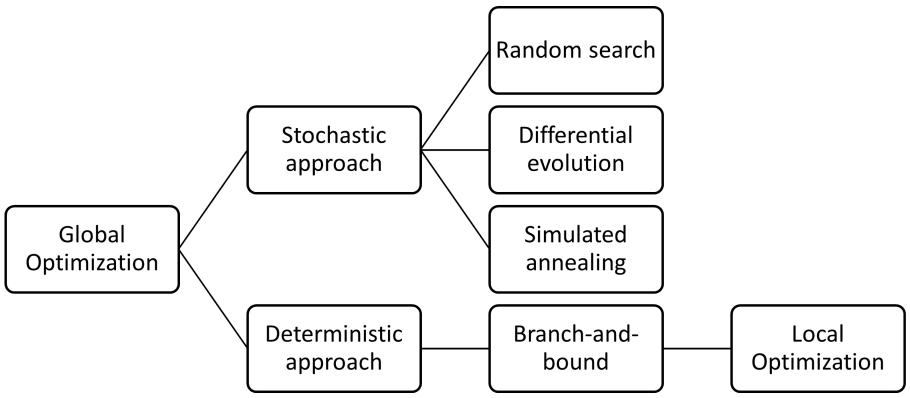


Figure 1.3: Classification of approaches for global dynamic optimization

Popular stochastic global optimization algorithms include random search [15], particle swarm optimization [156], differential evolution [10], and simulated annealing [51]. When used in practice, these stochastic algorithms may treat the considered dynamic process model as a black box. Thus, if we use such a stochastic approach to solve the dynamic optimization problem (1.1), no reformulation of the original dynamic system is required. However, these stochastic algorithms cannot guarantee to locate a global solution within finite time. They also typically cannot verify global optimality at all. Thus, the solution obtained from a stochastic global

optimization algorithm cannot be reliably considered to be global optima. Nevertheless, such a solution may provide a starting point for deterministic methods and improve the convergence of local optimization algorithms [14].

By contrast, deterministic global optimization algorithms can locate and validate a global minimum within finite iterations. They are typically developed based on the spatial branch-and-bound framework [42]. In a typical branch-and-bound method for global optimization on a box domain, we first compute the lower and upper bounds of the optimal objective value of (1.6) on the original domain. If the difference between these two bounds is within a specified tolerance ϵ, then this branch-and-bound algorithm terminates. Otherwise, the original domain is divided into two subdomains, and the upper bounds and lower bounds of the optimal objective value in both subdomains are determined. Figure 1.4 illustrates the branch-and-bound algorithm's progress after the original domain has been divided once. In this figure, if the lower bound in the subdomain R1 is higher than the upper bound in the subdomain R2, then it is impossible for a global minimum to be located in R1. Thus, R1 is eliminated from consideration. This process of comparing, dividing, and possibly eliminating will continue recursively in R2. As this is carried out, we can infer that a smaller and smaller portion of the original search space contains a global optimum. In the limit of infinitely many iterations, the known upper bounds and lower bounds on the unknown globally optimal value will converge. Before this, when their difference reaches the specified tolerance ϵ, the branch-and-bound algorithm terminates and a globally ϵ-optimal solution is found and verified.

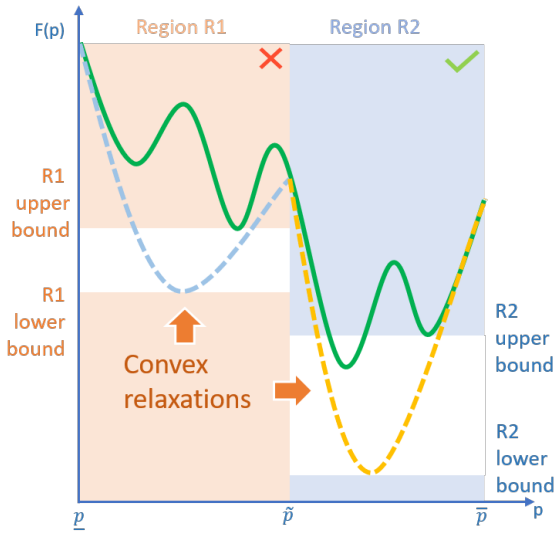


Figure 1.4: Branch-and-bound algorithm

Computing valid upper and lower bounds of the globally optimal value in each subdomain is critical in branch-and-bound algorithms. Providing the upper bounds is relatively straightforward since any feasible point of the problem is greater than or equal to the global minimum. In practice, we usually use a local minimum in a subdomain as the upper bound. However, obtaining a useful valid lower bound is much more difficult. As illustrated in Section 1.1.3, this process typically involves constructing convex relaxations of nonconvex functions and minimizing these relaxations, which is the focus of much of this thesis. Note that this deterministic global optimization approach immediately extends to the dynamic optimization problem (1.1) if we have ODE-based convex relaxations for the objective function J, obtained by combining convex reachable set enclosures with MR to incorporate the cost function ϕ [120]. The simultaneous approach described in Section 1.1.1 may also be applied; this is the most direct way to apply off-the-shelf global optimization solvers like BARON [139] to dynamic global optimization.

Different from the standard lower bounding approaches mentioned above, Scott and Barton [119] proposed a completely different approach for bounding the non-convex OCP in (1.3). They constructed a convex underestimating OCP by relaxing the original cost function and dynamic system with GMR. The optimal solution value of this relaxed OCP is a guaranteed lower bound of the original OCP's optimal solution value, which is useful in branch-and-bound algorithms for deterministic global optimization. They also proposed that this convex underestimating OCP can be solved to its global optimality using a gradient-based numerical method from [13].

1.1.6 Implicit functions and inverse functions

An *implicit function* is a function $\boldsymbol{z} : \mathbb{R}^{n_p} \to \mathbb{R}^{n_z}$ that is defined implicitly so as to satisfy the equation:

$$\boldsymbol{h}(\boldsymbol{z}(\boldsymbol{p}), \boldsymbol{p}) \equiv \boldsymbol{0},$$

where $\boldsymbol{h} : \mathbb{R}^{n_z+n_p} \to \mathbb{R}^{n_p}$ is a known *residual function*. Such implicit functions $\boldsymbol{z}$ appear in many research areas and applications [47], such as the ellipse equation in physics and astronomy, the van der Waals equation of state in chemical engineering, the equality constraints in mathematical programming [151], and the algebraic equations in DAEs. Since a closed-form expression is typically not available for the implicit function $\boldsymbol{z}$, its convex relaxations cannot be constructed using the αBB or McCormick relaxations summarized in Section 1.1.3.

Several existing approaches have been developed to relax it nevertheless. One major category of these approaches is based on applying a fixed-point iteration

solver to the original nonlinear equation system, and then relaxing these closed-form iterations. Scott et al. [123] introduced an approach to construct convex relaxations for implicit functions by applying GMR to finitely many fixed-point iterations [100]. Stuber et al. [136] later discovered that this approach may not provide any refinement over the known interval bounds, which may limit the applicability of this approach. To deal with this issue, they proposed an improved successive fixed-point iteration approach to construct convex relaxations for implicit functions by relaxing iterations based on the mean value theorem [136]. This approach was employed to relax the equality constraints in NLPs as inequality constraints and reduce the dimensionality, which is particularly useful in global optimization. Note that this approach only applies to factorable residue functions, and assumes unique solutions of the corresponding nonlinear equations system. It also requires additional *a priori* knowledge of the Jacobian of the residual function, such as interval bounds and convex relaxations. Khan et al. [72] applied Stuber et al.'s approach to construct differentiable relaxations for implicit functions, using DMR in place of GMR. Wilhelm et al. [153] adapted Stuber et al.'s approach to generate convex relaxations for the numerical solutions of parametric ODEs after discretizing them with implicit ODE solution methods.

A second category of implicit function relaxations is reverse McCormick propagation (RM) proposed by Wechsung et al. [151]. RM is developed based on the standard McCormick relaxations for factorable functions, except that it carefully propagates the convex and concave relaxations backward through the function's

computational graph, like the reverse mode of automatic differentiation [58]. Moreover, RM is also applicable to constraint satisfaction problems (CSPs), which contain both equality and inequality constraints. In this case, convex relaxations are constructed for a point-to-set mapping of a system parameter to the corresponding feasible region. Nevertheless, implementing this relaxation method is a non-trivial coding task and requires advanced coding skills. This implementation task requires operator overloading or source transformation to step through the function's computational graph.

1.2 Motivation and goals

Currently available deterministic global dynamic optimization algorithms can only solve relatively small problem instances with few decision variables and few state variables in a reasonable time [120]. The main barrier is the lack of useful convex relaxations for the ODE solution $\boldsymbol{x}(t_f, \cdot)$ in (1.2) [112]. To improve the computational efficiency of deterministic global optimization algorithms, these convex relaxations need to be tight, so that they provide more useful information about the original dynamic system. Tighter convex relaxations can help the branch-and-bound algorithms described in Section 1.1.5 to converge faster [42]. Branch-and-bound algorithms require minimizing convex relaxations repeatedly using local optimization solvers. Gradient-based local solvers, such as IPOPT and CONOPT, nominally require that the function being minimized is differentiable, and they also require the gradients of this function to be available as a subroutine. Overall, having differentiable objective functions enables faster convergence of local optimization methods [97]. Therefore, developing tight convex relaxations for the

parametric ODE (1.2) are critical to improving the computational performance of solving the dynamic optimization problem (1.1) with deterministic global optimization algorithms. Differentiability of these relaxations would further speed up the computation of lower bounds.

As summarized in Section 1.1.6, Stuber et al. [136] developed a fixed-point iteration approach to construct convex relaxations for implicit functions. Wilhelm et al. [153] applied this approach to generated convex relaxations for the numerical solutions of the parametric ODE (1.2), obtained with implicit integration methods. However, Stuber et al.'s approach for relaxing implicit functions requires the uniqueness of the implicit function and requires the corresponding residual function to be factorable. Their approach is also restricted to few convex relaxation methods, i.e. GMR and DMR. Moreover, to achieve tight convex relaxations, their approach may require many iterations. Thus, a generic and versatile approach for constructing tight convex relaxations for implicit functions can help relax the numerical solutions of the parametric ODE (1.2). These implicit function relaxations are also useful in feasibility analysis [151] and equality-constrained bilevel optimization [135], which commonly appear in many engineering applications, such as process flowsheet optimization [68], chemical reactor optimal design [135], and feedstock optimization [93].

Another category of established approaches for generating ODE relaxations includes the differential inequalities-based methods summarized in Sections 1.1.3-1.1.4. Several state-of-the-art methods in this category [120, 131] solely rely on constructing and numerically solving the ODE system in (1.5) that contains discrete jumps. As illustrated in Section 1.1.4, these discrete jumps increase the difficulty

in numerical implementation, bring nonsmoothness into the generated convex relaxations, and hinder gradient evaluation. To construct smooth ODE relaxations and evaluate their gradients for deterministic global optimization, these discrete jumps ought to be addressed.

Furthermore, those state-of-the-art ODE relaxation methods mentioned in the previous paragraph depend on predefined interval bounds of ODEs, which are typically computed using Harrison's method [59]. This method computes interval bounds for ODEs by solving an auxiliary ODE system whose RHS functions are constructed using interval arithmetic [94]. However, Harrison's method may be vulnerable to the dependency problem and the wrapping effect [94], which may lead to conservative interval bounds of ODEs. To build tighter ODE relaxations, it would be beneficial to develop an approach that generates tighter interval bounds than Harrison's method and replace it in those state-of-the-art ODE relaxation approaches. These tighter interval bounds can also lead to more accurate reachability analysis for many engineering applications, such as the state estimation in biochemical processes [69] and the collision avoidance of aircraft [91].

Lastly, the optimization of OCPs appears in many engineering problems, from determining the optimal control inputs of batch reactors [87] to landing an autonomous spacecraft on a planet surface [2]. To solve the OCP in (1.3) to global optimality using branch-and-bound algorithms, we need to supply global lower bounds of its optimal solution value. Scott and Barton [119] proposed an approach to construct a convex underestimating OCP whose optimal solution value is guaranteed to be a lower bound of the original OCP's optimal solution value. This relaxation approach essentially extended their work for parametric ODEs [122] to

OCPs. Since later they developed a superior relaxation method for parametric ODEs [120], we are interested to see if this superior method can also be extended to OCPs and construct tighter lower bounds than the their prior approach in [119]. Tighter lower bounds help the branch-and-bound algorithms used in deterministic global optimization to converge faster in principle [42]. In addition, their prior approach in [119] has never been implemented. We are also interested in developing a practical implementation to solve the relaxed OCP to global optimality numerically.

To address the issues in previous paragraphs, this thesis focuses on constructing improved convex relaxations for implicit functions, parametric ODEs, and nonconvex OCPs. These improved relaxations can enhance the computational performance of the deterministic global optimization approach as described in Section 1.1.5, for solving the dynamic optimization problem (1.1) and the OCP (1.3) to global optimality. The specific goals of this thesis are listed as follows:

1. Develop general, more versatile approaches for constructing tighter convex relaxations of implicit functions. These approaches can also be used to generate tighter convex relaxations for the numerical solutions of the parametric ODE (1.2) obtained using implicit integration methods.

2. Construct differentiable convex relaxations for the parametric ODE (1.2) by eliminating the discrete jumps in the Scott-Barton method (1.5), to permit the evaluation of gradients and resolve the theoretical obstacles described in Section 1.1.4.

3. Develop a new framework in which families of new bounds and relaxations may be generated for the parametric ODE (1.2), to construct and identify

relaxations that are superior to the current state of the art.

4. Construct tighter lower bounds of the nonconvex OCP (1.3) and develop implementations to compute them, so that they can be used in the deterministic global optimization of OCPs.

The overarching goal of this line of research is to develop computationally efficient deterministic global optimization algorithms and implementations that can solve large-scale dynamic optimization problems with many state variables and many decision variables. These implementations should be easily adapted for solving various real-world engineering problems with dynamic systems embedded, such as parameter estimation problems, open-loop optimal control problems, and safety verification problems. Completing the specific goals mentioned above is a critical step towards the achievement of this ultimate goal.

1.3 Contributions and thesis structure

This section summarizes the novel contributions of this thesis, in the order in which they appear in subsequent chapters.

Chapter 2, reproduced from the manuscript [30] to be submitted before the anticipated thesis defense, presents a new strategy to construct convex relaxations for implicit functions. These relaxations are described as convex parametric programs whose constraints are convex relaxations of the original residual function. It is shown that the optimal objective values of these parametric programs underestimate the original implicit function and are indeed convex. In general, these relaxations can be evaluated at a similar computational cost to evaluating the original

implicit function. Unlike previous approaches, this new approach does not assume uniqueness of the implicit function and does not require the original residual function to be factorable. Any valid convex relaxation techniques can be employed in this approach to relax the residual function. This new convex relaxation strategy is extended to inverse functions, feasible-set mappings in constraint satisfaction problems, as well as parametric ODEs. A proof-of-concept implementation of this strategy is developed in Julia, and numerical examples are presented to illustrate the convex relaxations produced for various implicit functions and optimal-value functions. In certain cases, these convex relaxations are much tighter than those generated with existing methods.

Chapter 3, reproduced from the manuscript [27] to be submitted before the anticipated thesis defense, presents a new method for generating convex relaxations for the parametric ODE (1.2) by improving the Scott-Barton method [120]. These relaxations are described by an auxiliary system of ODEs constructed with a novel smoothing technique. It is shown that these auxiliary ODEs have unique solutions, and these solutions are valid convex relaxations of the solutions of (1.2). Moreover, they are at least as tight as the relaxations constructed with the Scott-Barton method. It is also confirmed that the new method generates differentiable relaxations under additional assumptions, and therefore permits sensitivity analysis for the parametric ODEs via gradient computation. A Julia implementation for automatically constructing and solving the auxiliary relaxations ODEs is presented and applied to several numerical examples.

Chapter 4, reproduced from the manuscript [29] to be submitted before the anticipated thesis defense, considers bounding nonconvex optimal control problems

(OCPs) as described in (1.3). Based on another work by Scott and Barton [119], a new formulation is proposed to construct a guaranteed lower bound for the optimal solution value of a nonconvex OCP by constructing a relaxed OCP. This relaxed problem is subject to Pontryagin's sufficient optimality conditions, while the original OCP is not. It is verified that the optimal solution value of the relaxed OCP is a guaranteed lower bound of the optimal solution value of the original OCP, and is at least as tight as the bound constructed using state-of-the-art methods. A two-point boundary-value problem is developed and implemented to solve the relaxed OCP to global optimality using the Pontryagin's Minimum Principle conditions. This implementation is applied to numerical examples, in which the new relaxations are empirically much tighter than established relaxations.

Chapter 5, reproduced from the manuscript in preparation [28], proposes a new optimization-based framework for computing time-varying interval bounds for the ODE in (1.4). This framework constructs an auxiliary system of ODEs whose RHS functions are embedded optimization problems. It is verified that the solutions of these auxiliary ODEs are valid bounds of the original ODE solution. It is also shown that this framework includes several established bounding approaches [59, 118, 127], but also includes many new approaches. Several of these new approaches are implemented in Julia, and are demonstrated to generate tighter interval bounds than existing methods in many numerical examples. These tighter interval bounds of ODEs are useful for constructing tighter convex relaxations.

Chapter 6, reproduced from the published conference proceeding [31], introduces complementarity reformulations of the optimization-based framework in Chapter 5. The Karush–Kuhn–Tucker (KKT) conditions are used to reformulated

the convex optimization problems embedded in the auxiliary RHS functions. This allows special software designed for mixed nonlinear complementarity systems, such as Siconos [1], to be used in solving these optimization-embedded ODEs much faster than using the original formulation in Chapter 5. As a result, tighter interval bounds than existing methods may be computed without adding significantly to the cost of bound evaluation.

Chapter 7, reproduced from the manuscript in preparation [32], combines the new framework described in Chapter 5 with the Scott-Barton method [120], and develops a novel general framework for constructing convex relaxations for the parametric ODE (1.2). These relaxations are confirmed to be valid convex relaxations, and are shown to be differentiable if the auxiliary RHS functions are differentiable. Unlike the Scott-Barton method, this new approach employs continuous bounding ODEs without discrete jumps, which permits the evaluation of gradients. Moreover, tighter convex relaxations can be constructed with this new approach, making use of the tight interval bounds developed in Chapter 5. Unlike the method in Chapter 3, the convex relaxations in this approach does not require *a priori* knowledge of the original model. A numerical implementation of this approach is developed in Julia and applied to several numerical examples.

For brevity, the following additional contributions of my Ph.D. work are not discussed further in this thesis. First, for the published journal article [33], I implemented a new subtangent-based convex relaxation approach in Julia, and applied this to several numerical examples. The deterministic global optimization algorithm using these convex relaxations is shown to be comparable to BARON

[139], one of the state-of-the-art deterministic global optimization solvers, in multiple benchmark problems. Second, for the published journal article [133], I implemented a new black-box sampling-based to provide a derivative-free technique to compute affine relaxations for NLPs tractably. This algorithm is particularly useful for constructing lower bounds in deterministic global optimization using convex relaxations whose sensitivities are unavailable. Several nonsmooth nonconvex NLPs and dynamic optimization problems are solved to global optimality using this implementation.

Furthermore, the two deterministic global optimization implementations mentioned in the previous paragraph, as well as the implementations of the methods in Chapters 3-7, are integrated into a Julia package `DynamicGlobalOpt.jl` that will be posted publicly to GitHub by December 2021. This package provides routines for computing and plotting various interval bounds and convex relaxations for ODEs automatically. It also contains a numerical solver for deterministic global dynamic optimization. Developed based on the branch-and-bound framework from the EAGO.jl package [152], this solver offers user-friendly interfaces for both general dynamic optimization problems as in (1.1) and parameter estimation problems. Several dynamic optimization benchmark problems [55] and real-world case studies [129] have been solved to global optimality successfully using this Julia package.

Chapter 2

Convex Relaxations of Implicit Functions

This chapter is to be submitted to a journal before my anticipated thesis defense.

2.1 Introduction

Branch-and-bound algorithms for deterministic global optimization require guaranteed lower bounds on the solution of a nonconvex nonlinear program (NLP) on particular subsets of the search space. This bounding information is typically obtained by generating and minimizing a convex relaxation of the original NLP to its global optimum with a convex solver [92]. For a function described explicitly by a closed-form expression, several established relaxation techniques can effectively generate associated convex relaxations. In particular, if the nonconvex function is twice-continuously differentiable, we may construct its convex relaxations using αBB relaxations [3], which involve adding a sufficiently large convex quadratic

term to the original function. If the nonconvex function is a finite composition of known intrinsic functions from a library, such as the functions that can be represented on a typical scientific calculator, then the function is said to be *factorable*, and we can construct its convex relaxations using McCormick's relaxation method [89, 92]. This relaxation method generates accurate and computationally cheap convex underestimators [72]. Several open-source implementations of this approach are available, such as the C++ library `MC++` [36] and the Julia package `McCormick.jl` [152]. However, if no closed-form expression for the original nonconvex function is known, then the convex relaxations methods mentioned previously are not directly applicable.

Roughly, this article primarily considers a function $\boldsymbol{x} : \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$ that is defined implicitly so as to satisfy the equation:

$$\boldsymbol{f}(\boldsymbol{x}(\boldsymbol{p}), \boldsymbol{p}) \equiv \boldsymbol{0},$$

where $\boldsymbol{f} : \mathbb{R}^{n_x+n_p} \to \mathbb{R}^{n_p}$ is a known *residual function*. Such *implicit functions* $\boldsymbol{x}$ appear in many research areas and applications [47], such as the ellipse equation in physics and astronomy, the van der Waals equation in chemical engineering, and the equality constraints in mathematical programming [151]. Since a closed-form expression is not available for the implicit function $\boldsymbol{x}$, its convex relaxations cannot be constructed using the αBB or McCormick relaxations. This article seeks improved dedicated convex relaxation techniques for implicit functions.

Several existing approaches have been developed to address this problem. One major category of these approaches is based on applying a fixed-point iteration

solver to the original nonlinear equation system, and then relaxing these closed-form iterations. Scott et al. [123] developed *generalized McCormick relaxations* (GM) based on McCormick's relaxation method, which permit convex and concave relaxations of the inputs to be used as arguments [136]. Using this property, Scott et al. [123] introduced an approach to construct convex relaxations for implicit functions by applying GM to finitely many fixed-point iterations [100]. Stuber et al. [136] later discovered that this approach may not provide any refinement over the known interval bounds, which may limit the applicability of this approach. To deal with this issue, they proposed an improved successive fixed-point iteration approach to construct convex relaxations for implicit functions by relaxing iterations based on the mean value theorem [136]. This approach was employed to relax the equality constraints in NLPs as inequality constraints and reduce the dimensionality, which is particularly useful in global optimization. Note that this approach only applies to factorable residue functions, and assumes unique solutions of the corresponding nonlinear equations system. It also requires additional *a priori* knowledge of the Jacobian of the residual function, such as interval bounds and convex relaxations. Khan et al. [72] applied Stuber et al.'s approach to construct differentiable relaxations for implicit functions, using differentiable McCormick relaxations (DM) [72, 73] in place of GM. Wilhelm et al. [153] adapted Stuber et al.'s approach to generate convex relaxations for the numerical solutions of parametric ordinary differential equations (ODEs) after discretizing them with implicit ODE solution methods.

A second category of implicit function relaxations is reverse McCormick propagation (RM) proposed by Wechsung et al. [151]. RM is developed based on

the standard McCormick relaxations for factorable functions, except that it carefully propagates the convex and concave relaxations backward through the function's computational graph, like the reverse mode of automatic differentiation [58]. Moreover, RM is also applicable to constraint satisfaction problems (CSPs), which contain both equality and inequality constraints. In this case, convex relaxations are constructed for a point-to-set mapping of a system parameter to the corresponding feasible region. Nevertheless, implementing this relaxation method is a nontrivial coding task and requires advanced coding skills. This implementation task requires operator overloading or source transformation to step through the function's computational graph.

In this work, we propose a novel strategy to generate convex and concave relaxations for implicit functions using parametric programming. These relaxations are described by convex optimization problems whose constraints are convex relaxations of the original residual function. This new approach is then extended to construct convex relaxations for inverse functions, point-to-set mappings in CSPs, as well as parametric ODEs. Directional derivatives of these convex relaxations are described by auxiliary convex quadratic programs. Unlike the established approaches described above, our new approach does not assume uniqueness of a solution, and does not require the original residual function to be factorable. It is also easier to implement and automate than previous methods. The parametric programs can be constructed with existing convex relaxation methods and solved with standard local optimization solvers. In our new approach, any convex relaxation techniques may be employed to relax the residual function, such as standard

McCormick relaxations [92], αBB relaxations [3], convex envelopes, and the pointwise best among multiple relaxations, while established methods are limited to one particular relaxation method, such as GM in [136, 123] and RM in [151]. Lastly, the convex and concave relaxations generated with our new approach are comparable to those established methods in tightness. These tight enclosures are potentially useful in applications like global optimization and reachability analysis.

This article is structured as follows. Section 2.2 introduces the mathematical background underlying this work. In Section 2.3, we present the formulation of our new strategy and demonstrate its correctness. We then extend this approach to inverse functions, and combine our approach with the *multivariate McCormick relaxations* of Tsoukalas and Mitsos [144]. Section 2.4 extends this new strategy to CSPs, where convex relaxations are constructed to enclose the point-to-set mappings in CSPs. The directional derivatives of these relaxations are constructed. In Section 2.5, we adapt the new convex relaxation strategy to improve the tightness of interval bounds for implicit functions and CSPs. Section 2.6 applies the new strategy to construct convex relaxations and interval bounds for parametric ODEs. Finally, a proof-of-concept Julia implementation of our results is described. Numerical examples are presented in Section 2.7 to illustrate our new approach.

2.2 Background

This section summarizes the mathematical background underlying this work, and echos the background presented in [31]. The following notation conventions are used in this article. Vectors are denoted with boldface lower-case letters (e.g. $\boldsymbol{x} \in \mathbb{R}^n$). Given vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, inequalities such as $\boldsymbol{x} < \boldsymbol{y}$ or $\boldsymbol{x} \leq \boldsymbol{y}$ are to be

interpreted componentwise. Throughout this article, convexity of a vector-valued function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ refers to convexity of all components f_i, and concavity is analogous. An *interval* in $\mathbb{R}^n$ is a nonempty subset of $\mathbb{R}^n$ of the form $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}\}$, which is also denoted as $[\boldsymbol{a}, \boldsymbol{b}]$. $\mathbb{IR}^n$ denotes the set of all intervals in $\mathbb{R}^n$.

Next, we introduce convex and concave relaxations of functions.

Definition 2.1. *Let $P \subset \mathbb{R}^{n_p}$ be a convex set. Consider a function $\boldsymbol{\phi} : P \to \mathbb{R}^m$.*

1. *$\boldsymbol{\phi}^{\mathrm{cv}} : P \to \mathbb{R}^m$ is a* convex relaxation *of $\boldsymbol{\phi}$ on P if $\boldsymbol{\phi}^{\mathrm{cv}}(\boldsymbol{p}) \leq \boldsymbol{\phi}(\boldsymbol{p})$ for all $\boldsymbol{p} \in P$ and $\boldsymbol{\phi}^{\mathrm{cv}}$ is convex on P.*

2. *$\boldsymbol{\phi}^{\mathrm{cc}} : P \to \mathbb{R}^m$ is a* concave relaxation *of $\boldsymbol{\phi}$ on P if $\boldsymbol{\phi}^{\mathrm{cc}}(\boldsymbol{p}) \geq \boldsymbol{\phi}(\boldsymbol{p})$ for all $\boldsymbol{p} \in P$ and $\boldsymbol{\phi}^{\mathrm{cc}}$ is concave on P.*

Several established approaches generate convex relaxations for closed-form factorable functions automatically. The αBB relaxation method [3] constructs convex relaxations for twice-continuously differentiable functions, and involves adding a sufficiently large positive convex quadratic term to the original function. Another approach is McCormick's relaxation method [89, 92, 123, 144, 72, 73]. Generalized McCormick relaxations (GM) [123], Tsoukalas-Mitsos relaxations (TM) [144], and differentiable McCormick relaxations (DM) [72, 73] have the special property that they describe convex relaxations for a composite function using relaxations of the inner composed functions. We will refer to relaxations with this property as *generalized convex and concave relaxations*, which are formalized according to the following definition adapted from [119].

Definition 2.2. *Let $P \subset \mathbb{R}^{n_p}$ be a convex set. Consider a function $\boldsymbol{\phi} : \mathbb{R}^{n_z} \to \mathbb{R}^{n_q}$. Functions $\boldsymbol{\phi}^{\mathrm{cv}}, \boldsymbol{\phi}^{\mathrm{cc}} : \mathbb{R}^{n_z} \times \mathbb{R}^{n_z} \to \mathbb{R}^{n_q}$ are* generalized convex and concave relaxations *of*

ϕ on $\mathbb{R}^{n_z}$, respectively, if the following holds for each function $\boldsymbol{z} : P \to \mathbb{R}^{n_z}$ and all choices of relaxations $\boldsymbol{z}^{\mathrm{cv}}, \boldsymbol{z}^{\mathrm{cc}} : P \to \mathbb{R}^{n_z}$. Consider $\boldsymbol{\rho} : P \to \mathbb{R}^{n_q}$ such that for each $\boldsymbol{p} \in P$,

$$\boldsymbol{\rho}(\boldsymbol{p}) = \boldsymbol{\phi}(\boldsymbol{z}(\boldsymbol{p})).$$

Then, the functions $\boldsymbol{\rho}^{\mathrm{cv}}, \boldsymbol{\rho}^{\mathrm{cc}} : P \to \mathbb{R}^{n_q}$ such that for each $\boldsymbol{p} \in P$,

$$\begin{aligned} \boldsymbol{\rho}^{\mathrm{cv}}(\boldsymbol{p}) &= \boldsymbol{\phi}^{\mathrm{cv}}(\boldsymbol{z}^{\mathrm{cv}}(\boldsymbol{p}), \boldsymbol{z}^{\mathrm{cc}}(\boldsymbol{p})) \\ \textit{and} \quad \boldsymbol{\rho}^{\mathrm{cc}}(\boldsymbol{p}) &= \boldsymbol{\phi}^{\mathrm{cc}}(\boldsymbol{z}^{\mathrm{cv}}(\boldsymbol{p}), \boldsymbol{z}^{\mathrm{cc}}(\boldsymbol{p})), \end{aligned}$$

are convex and concave relaxations of $\boldsymbol{\rho}$ on P, respectively.

Next, we summarize a sufficient condition for a parametric program to be convex. The following definition and proposition are adapted from [52].

Definition 2.3. *Let $P \subset \mathbb{R}^{n_p}$ be a convex set. A* point-to-set *map $S^R : \mathbb{R}^{n_p} \rightrightarrows \mathbb{R}^{n_x}$ assigns a subset of $\mathbb{R}^{n_x}$ to each element of $\mathbb{R}^{n_p}$. S^R is* convex *on P if, for all $\boldsymbol{p}_1, \boldsymbol{p}_2 \in P$ and $\lambda \in (0,1)$, the Minkowski sum $\lambda S^R(\boldsymbol{p}_1) + (1-\lambda) S^R(\boldsymbol{p}_2)$ is a subset of $S^R(\lambda \boldsymbol{p}_1 + (1-\lambda)\boldsymbol{p}_2)$. Moreover, S^R is a* convex relaxation *of an arbitrary point-to-set map $S : \mathbb{R}^{n_p} \rightrightarrows \mathbb{R}^{n_x}$ on P if, for all $\boldsymbol{p} \in P$, $S(\boldsymbol{p}) \subseteq S^R(\boldsymbol{p})$ and S^R is convex on P.*

Proposition 2.1 (Corollary 2.1 in [52])**.** *Define $X \in \mathbb{R}^{n_x}$ and functions $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}$, $\boldsymbol{g} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_g}$, and $\boldsymbol{h} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_h}$. Let $R : \mathbb{R}^{n_p} \rightrightarrows \mathbb{R}^{n_x}$ be a point-to-set map such that, for each $\boldsymbol{p} \in P$,*

$$R(\boldsymbol{p}) = \{\boldsymbol{x} \in X \mid \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{p}) \leq \boldsymbol{0}, \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{p}) = \boldsymbol{0}\}.$$

For each $\boldsymbol{p} \in P$, consider a general parametric optimization problem

$$\min_{\boldsymbol{x}} \ f(\boldsymbol{x}, \boldsymbol{p}), \qquad \textit{subject to} \ \ \boldsymbol{x} \in R(\boldsymbol{p}).$$

Define an optimal-value function $f^* : \mathbb{R}^{n_p} \to \mathbb{R}$ such that for each $\boldsymbol{p} \in P$,

$$f^*(\boldsymbol{p}) = \begin{cases} \inf_{\boldsymbol{x}} \{ f(\boldsymbol{x}, \boldsymbol{p}) \ \mid \ \boldsymbol{x} \in R(\boldsymbol{p}) \}, & \textit{if } R(\boldsymbol{p}) \neq \emptyset, \\ +\infty, & \textit{if } R(\boldsymbol{p}) = \emptyset. \end{cases}$$

If X and P are convex, f is convex on $X \times P$, $\boldsymbol{g}$ is convex on $X \times P$, and $\boldsymbol{h}$ is affine on $X \times P$, then f^* is convex on P.

Finally, we summarize the *directional derivative*, which provides local radial sensitivity information for a function. The following definition is adapted from [115, Section 3.1].

Definition 2.4. *Let $P \subset \mathbb{R}^{n_p}$ be a convex set. Let $\boldsymbol{\phi} : P \to \mathbb{R}^n$ be a function. If for every $\boldsymbol{d} \in \mathbb{R}^n$ the limit*

$$\boldsymbol{\phi}'(\boldsymbol{z}_0; \boldsymbol{d}) = \lim_{\lambda \downarrow 0} \frac{1}{\lambda} (\boldsymbol{\phi}(\boldsymbol{z}_0 + \lambda \boldsymbol{d}) - \boldsymbol{\phi}(\boldsymbol{z}_0))$$

exists, then $\boldsymbol{\phi}$ is said to be directionally differentiable *at $\boldsymbol{z}_0$ and the function $\boldsymbol{\phi}'(\boldsymbol{z}_0; \cdot)$ is the* directional derivative *of $\boldsymbol{\phi}$ at $\boldsymbol{z}_0$.*

2.3 Convex Relaxations of Implicit Functions

In this section, we present a new formulation for generating convex and concave relaxations for an implicit function using parametric programming. In the remainder of this section, consider a Lipschitz continuous residual function $\boldsymbol{f} : \mathbb{R}^{n_x+n_p} \to \mathbb{R}^{n_p}$, and the following system of equations

$$\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{p}) = \boldsymbol{0}. \tag{2.1}$$

Consider a convex compact set $P \in \mathbb{R}^{n_p}$. If for each $\boldsymbol{p} \in P$, there exists $\boldsymbol{z}$ that satisfy (2.1), then there is an implicit function $\boldsymbol{x} : P \to \mathbb{R}^{n_x}$ (not necessarily unique) such that, for each $\boldsymbol{p} \in P$,

$$\boldsymbol{f}(\boldsymbol{x}(\boldsymbol{p}), \boldsymbol{p}) = \boldsymbol{0}. \tag{2.2}$$

Assumption 2.1. *Assume that there exists at least one Lipschitz continuous function $\boldsymbol{x} : P \to \mathbb{R}^{n_x}$ and a known interval $X \in \mathbb{IR}^{n_x}$ such that* (2.2) *holds and $\boldsymbol{x}(\boldsymbol{p}) \in X$ for every $\boldsymbol{p} \in P$.*

The semi-local implicit function theorem [98] gives sufficient conditions for the uniqueness of $\boldsymbol{x}$ in Assumption 2.1. Roughly, that theorem requires the partial derivative $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}}$ to be nonsingular on $X \times P$ [47]. This result was later extended to nonsmooth functions in [40, Theorem 7.1.1] and subsequent discussions.

2.3.1 Main Result

Under Assumption 2.1, a new description for constructing convex and concave relaxations for any such implicit function $\boldsymbol{x}$ is presented in the theorem below.

Theorem 2.1. *Let $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc} : \mathbb{R}^{n_x+n_p} \to \mathbb{R}^{n_p}$ be convex and concave relaxations of $\boldsymbol{f}$ on $X \times P$, respectively. Define $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc} : P \to \mathbb{R}^{n_x}$ such that, for each $i \in \{1, \ldots, n_x\}$ and $\boldsymbol{p} \in P$,*

$$x_i^{cv}(\boldsymbol{p}) = \min_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \quad \boldsymbol{f}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0} \leq \boldsymbol{f}^{cc}(\boldsymbol{\xi}, \boldsymbol{p}), \tag{2.3a}$$

$$x_i^{cc}(\boldsymbol{p}) = \max_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \quad \boldsymbol{f}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0} \leq \boldsymbol{f}^{cc}(\boldsymbol{\xi}, \boldsymbol{p}). \tag{2.3b}$$

Under Assumption 2.1, $\boldsymbol{x}^{cv}$ and $\boldsymbol{x}^{cc}$ are convex and concave relaxations of $\boldsymbol{x}$ on P, respectively.

Proof. Following Definition 2.1, we first show that $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$ satisfy $\boldsymbol{x}^{cv}(\boldsymbol{p}) \leq \boldsymbol{x}(\boldsymbol{p})$ and $\boldsymbol{x}^{cc}(\boldsymbol{p}) \geq \boldsymbol{x}(\boldsymbol{p})$ for all $\boldsymbol{p} \in P$. Then, we demonstrate their respective convexity and concavity.

To show that $\boldsymbol{x}^{cv}(\boldsymbol{p}) \leq \boldsymbol{x}(\boldsymbol{p})$ for all $\boldsymbol{p} \in P$, define an optimal-value function $\boldsymbol{\omega} : P \to X$ such that, for each $i \in \{1, \ldots, n_x\}$ and $\boldsymbol{p} \in P$,

$$\omega_i(\boldsymbol{p}) = \min_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \quad \boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{p}) = \boldsymbol{0}. \tag{2.4}$$

Choose any $i \in \{1, \ldots, n_x\}$ and $\boldsymbol{p} \in P$. Since $\boldsymbol{x}$ satisfies (2.2), the optimization problem (2.4) is feasible and $\omega_i(\boldsymbol{p}) \leq x_i(\boldsymbol{p})$. Observe that the solution $\boldsymbol{\xi}^*$ of the optimization problem (2.4) is feasible in (2.3a). Therefore, $x_i^{cv}(\boldsymbol{p}) \leq \omega_i(\boldsymbol{p}) \leq x_i(\boldsymbol{p})$ for all $\boldsymbol{p} \in P$.

Next, we verify the convexity of $\boldsymbol{x}^{cv}$ and the concavity of $\boldsymbol{x}^{cc}$. Define $\boldsymbol{\phi} : X \times P \to \mathbb{R}^{2n_y}$ such that $\boldsymbol{\phi}(\boldsymbol{\xi}, \boldsymbol{p}) = (\boldsymbol{f}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}), -\boldsymbol{f}^{cc}(\boldsymbol{\xi}, \boldsymbol{p}))$ for each $\boldsymbol{\xi} \in X$ and $\boldsymbol{p} \in P$. Since $\boldsymbol{f}^{cv}$ and $\boldsymbol{f}^{cc}$ are convex and concave functions, respectively, $\boldsymbol{\phi}$ is convex on $X \times P$. For each $i \in \{1, \ldots, n_x\}$ and $\boldsymbol{p} \in P$, (2.3a) is equivalent to

$$x_i^{cv}(\boldsymbol{p}) = \min_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \quad \boldsymbol{\phi}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0}. \tag{2.5}$$

Under Assumption 2.1, for each $\boldsymbol{p} \in P$, $\boldsymbol{x}(\boldsymbol{p})$ is a feasible point of (2.5), so that the feasible region of (2.5) is non-empty. Since the objective function of (2.5) is linear, X and P are convex, and $\boldsymbol{\phi}$ is convex on $X \times P$, the convexity of x_i^{cv} on P follows from Proposition 2.1. □

The optimization problems in (2.3a) and (2.3b) are convex optimization problems. Thus, the relaxations $\boldsymbol{x}^{cv}$ and $\boldsymbol{x}^{cc}$ can in principle be computed with local NLP solvers, such as IPOPT and CONOPT. Since evaluating $\boldsymbol{x}$ involves solving a nonlinear equation system of similar size, we do not expect that evaluating each $x_i^{cv}(\boldsymbol{p})$ or $x_i^{cc}(\boldsymbol{p})$ would be much more computationally expensive than evaluating $\boldsymbol{x}(\boldsymbol{p})$. Moreover, there are two particular scenarios in which the optimization problems in (2.3a) and (2.3b) are easier to solve. Firstly, if the supplied relaxations $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ are chosen to be affine or piecewise-affine relaxations of $\boldsymbol{f}$ [70, 33, 133], then (2.3a) and (2.3b) become linear programs (LPs), which can be solved efficiently by off-the-shelf solvers such as CPLEX and Gurobi. Secondly, if the original function $\boldsymbol{f}$ is quadratic and $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ are corresponding αBB relaxations [3], then $\boldsymbol{f}^{cv}$ and $\boldsymbol{f}^{cc}$ will also be quadratic. In the case, (2.3a) and (2.3b) become convex parametric quadratically constrained quadratic programs (QCQPs) [102]. The advantage

here is that the optimization problems (2.3a) and (2.3b) may be solved analytically in advance to determine closed-form expressions for $\boldsymbol{x}^{cv}$ and $\boldsymbol{x}^{cc}$, to aid rapid evaluation. This is particularly useful in deterministic global optimization, which typically requires many evaluations of convex relaxations.

2.3.2 Tsoukalas-Mitsos Relaxation

In this subsection, we extend Theorem 2.1 to generate convex and concave relaxations for a composition of an implicit outer function with a known inner function. Suppose that the convex and concave relaxations of the inner function are available. Then, we may construct the relaxations of the composite function by modifying (2.3) using Tsoukalas-Mitsos relaxations [144].

Let $R \subset \mathbb{R}^{n_r}$ and $P \subset \mathbb{R}^{n_p}$ be convex compact sets, and consider continuously differentiable functions $\boldsymbol{r} : P \to R$ and $\widehat{\boldsymbol{f}} : \mathbb{R}^{n_x+n_r} \to \mathbb{R}^{n_p}$. Suppose that an implicit function $\widehat{\boldsymbol{x}} : R \to \mathbb{R}^{n_x}$ is defined so as to satisfy:

$$\widehat{\boldsymbol{f}}(\widehat{\boldsymbol{x}}(\boldsymbol{q}), \boldsymbol{q}) = \boldsymbol{0}, \quad \forall \boldsymbol{q} \in R. \tag{2.6}$$

and suppose that the following assumption holds.

Assumption 2.2. *Assume that there exists at least one differentiable function $\widehat{\boldsymbol{x}} : R \to X$ such that* (2.6) *holds for every $\boldsymbol{q} \in R$.*

Next, consider a composite function $\boldsymbol{x} : P \to \mathbb{R}^{n_x}$ such that, for all $\boldsymbol{p} \in P$,

$$\boldsymbol{x}(\boldsymbol{p}) = \widehat{\boldsymbol{x}}(\boldsymbol{r}(\boldsymbol{p})). \tag{2.7}$$

We will apply the Tsoukalas-Mitsos relaxations to show that, if convex and concave relaxations of $\boldsymbol{r}$ are available, then correct relaxations of $\boldsymbol{x}$ can be described by a formulation similar to (2.3).

Theorem 2.2. *Define* $\widehat{\boldsymbol{f}}^{cv}, \widehat{\boldsymbol{f}}^{cc} : \mathbb{R}^{n_x+n_r} \times \mathbb{R}^{n_x+n_r} \to \mathbb{R}^{n_p}$ *such that* $\widehat{\boldsymbol{f}}^{\mathrm{cv}}$ *and* $\widehat{\boldsymbol{f}}^{\mathrm{cc}}$ *are generalized convex and concave relaxations of* $\widehat{\boldsymbol{f}}$ *on* $X \times R$*, respectively. Let* $\boldsymbol{r}^{cv}, \boldsymbol{r}^{cc} : \mathbb{R}^{n_p} \to \mathbb{R}^{n_r}$ *be convex and concave relaxations of* $\boldsymbol{r}$ *on* P*, respectively. Consider* $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$ *such that, for each* $i \in \{1, \ldots, n_x\}$ *and* $\boldsymbol{p} \in P$*,*

$$x_i^{cv}(\boldsymbol{p}) = \min_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \tag{2.8a}$$
$$\widehat{\boldsymbol{f}}^{cv}((\boldsymbol{\xi}, \boldsymbol{r}^{cv}(\boldsymbol{p})), (\boldsymbol{\xi}, \boldsymbol{r}^{cc}(\boldsymbol{p}))) \leq \boldsymbol{0} \leq \widehat{\boldsymbol{f}}^{cc}((\boldsymbol{\xi}, \boldsymbol{r}^{cv}(\boldsymbol{p})), (\boldsymbol{\xi}, \boldsymbol{r}^{cc}(\boldsymbol{p}))),$$
$$x_i^{cc}(\boldsymbol{p}) = \max_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \tag{2.8b}$$
$$\widehat{\boldsymbol{f}}^{cv}((\boldsymbol{\xi}, \boldsymbol{r}^{cv}(\boldsymbol{p})), (\boldsymbol{\xi}, \boldsymbol{r}^{cc}(\boldsymbol{p}))) \leq \boldsymbol{0} \leq \widehat{\boldsymbol{f}}^{cc}((\boldsymbol{\xi}, \boldsymbol{r}^{cv}(\boldsymbol{p})), (\boldsymbol{\xi}, \boldsymbol{r}^{cc}(\boldsymbol{p}))).$$

Under Assumption 2.2, with $\boldsymbol{x}$ *defined in* (2.7)*,* $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$ *in* (2.8) *are convex and concave relaxations of* $\boldsymbol{x}$ *on* P*, respectively.*

Proof. Consider $\boldsymbol{f}, \boldsymbol{f}^{cv}, \boldsymbol{f}^{cc} : \mathbb{R}^{n_x+n_p} \to \mathbb{R}^{n_x}$ such that, for each $\boldsymbol{\xi} \in \mathbb{R}^{n_x}$ and $\boldsymbol{p} \in \mathbb{R}^{n_p}$,

$$\begin{aligned}
\boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{p}) &= \widehat{\boldsymbol{f}}(\boldsymbol{\xi}, \boldsymbol{r}(\boldsymbol{p})), \\
\boldsymbol{f}^{\mathrm{cv}}(\boldsymbol{\xi}, \boldsymbol{p}) &= \widehat{\boldsymbol{f}}^{\mathrm{cv}}((\boldsymbol{\xi}, \boldsymbol{r}^{cv}(\boldsymbol{p})), (\boldsymbol{\xi}, \boldsymbol{r}^{cc}(\boldsymbol{p}))), \\
\boldsymbol{f}^{\mathrm{cc}}(\boldsymbol{\xi}, \boldsymbol{p}) &= \widehat{\boldsymbol{f}}^{\mathrm{cc}}((\boldsymbol{\xi}, \boldsymbol{r}^{cv}(\boldsymbol{p})), (\boldsymbol{\xi}, \boldsymbol{r}^{cc}(\boldsymbol{p}))).
\end{aligned}$$

Substituting $\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{f}}^{cv}, \widehat{\boldsymbol{f}}^{cc}$ with $\boldsymbol{f}, \boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$, respectively, then (2.6) and (2.8) becomes (2.1) and (2.3), respectively. Since $\widehat{\boldsymbol{f}}^{cv}$ and $\widehat{\boldsymbol{f}}^{cc}$ are generalized convex and concave

relaxations of $\widehat{\boldsymbol{f}}$ on $X \times R$, respectively, $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ are convex and concave relaxations of $\boldsymbol{f}$ on $X \times P$ following Definition 2.2. Assumption 2.2 shows that there exists one differentiable function $\boldsymbol{x} : P \to X$ such that

$$\widehat{\boldsymbol{f}}(\widehat{\boldsymbol{x}}(\boldsymbol{r}(\boldsymbol{p})), \boldsymbol{r}(\boldsymbol{p})) = \boldsymbol{f}(\boldsymbol{x}(\boldsymbol{p}), \boldsymbol{p}) = \boldsymbol{0},$$

which ensures Assumption 2.1. Therefore, Theorems 2.1 ensures that $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$ in (2.8) are convex and concave relaxations of $\boldsymbol{x}$ defined in (2.7) on P, respectively. □

Generalized convex and concave relaxations, $\widehat{\boldsymbol{f}}^{\text{cv}}$ and $\widehat{\boldsymbol{f}}^{\text{cc}}$, may be constructed using GM and DM.

2.3.3 Relaxations of Inverse Functions

Theorem 2.1 may be adapted to generate convex and concave relaxations for inverse functions. Suppose that $\boldsymbol{v} : X \to P$ is an invertible function, and so there exists an inverse function $\boldsymbol{v}^{-1} : P \to X$ of $\boldsymbol{v}$ such that, for each $\boldsymbol{p} \in P$,

$$\boldsymbol{v}(\boldsymbol{v}^{-1}(\boldsymbol{p})) = \boldsymbol{p}.$$

Observe that $\boldsymbol{v}^{-1}$ may also be expressed as an implicit function defined by the equation system:

$$\bar{\boldsymbol{f}}(\boldsymbol{v}^{-1}(\boldsymbol{p}), \boldsymbol{p}) = \boldsymbol{v}(\boldsymbol{v}^{-1}(\boldsymbol{p})) - \boldsymbol{p} = \boldsymbol{0} \quad \forall \boldsymbol{p} \in P. \tag{2.9}$$

So, correct convex and concave relaxations of $\boldsymbol{v}^{-1}$ on P may be constructed with a formulation adapted from (2.3).

Corollary 2.1. *Let $\boldsymbol{v}^{cv}, \boldsymbol{v}^{cc} : X \to P$ be convex and concave relaxations of $\boldsymbol{v}$ on X, respectively. Consider functions $\boldsymbol{v}^{-\mathrm{cv}}, \boldsymbol{v}^{-\mathrm{cc}} : P \to \mathbb{R}^{n_x}$ such that, for each $i \in \{1, \ldots, n_x\}$ and $\boldsymbol{p} \in P$,*

$$v_i^{-\mathrm{cv}}(\boldsymbol{p}) = \min_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \quad \boldsymbol{v}^{cv}(\boldsymbol{\xi}) \leq \boldsymbol{p} \leq \boldsymbol{v}^{cc}(\boldsymbol{\xi}), \tag{2.10a}$$

$$v_i^{-\mathrm{cc}}(\boldsymbol{p}) = \max_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \quad \boldsymbol{v}^{cv}(\boldsymbol{\xi}) \leq \boldsymbol{p} \leq \boldsymbol{v}^{cc}(\boldsymbol{\xi}). \tag{2.10b}$$

Then, $\boldsymbol{v}^{-\mathrm{cv}}, \boldsymbol{v}^{-\mathrm{cc}}$ in (2.10) *are convex and concave relaxations of $\boldsymbol{v}^{-1}$ on P.*

Proof. The desired result can be verified by showing that the hypotheses in Theorem 2.1 holds with $\boldsymbol{v}^{-1}$ in place of $\boldsymbol{x}$, and $\boldsymbol{v}^{-\mathrm{cv}}, \boldsymbol{v}^{-\mathrm{cc}}$ in place of $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$, respectively. Since $\boldsymbol{v}$ is invertible and $\boldsymbol{v}^{-1}$ is its inverse, Assumption 2.1 is satisfied. Define functions $\bar{\boldsymbol{f}}^{\mathrm{cv}}, \bar{\boldsymbol{f}}^{\mathrm{cc}} : X \times P \to P$ such that, for each $\boldsymbol{p} \in P$,

$$\bar{\boldsymbol{f}}^{\mathrm{cv}}(\boldsymbol{z}, \boldsymbol{p}) = \boldsymbol{v}^{\mathrm{cv}}(\boldsymbol{z}) - \boldsymbol{p},$$
$$\bar{\boldsymbol{f}}^{\mathrm{cc}}(\boldsymbol{z}, \boldsymbol{p}) = \boldsymbol{v}^{\mathrm{cc}}(\boldsymbol{z}) - \boldsymbol{p}.$$

Because $\boldsymbol{v}^{cv}$ and $\boldsymbol{v}^{cc}$ are convex and concave relaxations of $\boldsymbol{v}$ on X, respectively, $\bar{\boldsymbol{f}}^{\mathrm{cv}}$ and $\bar{\boldsymbol{f}}^{\mathrm{cc}}$ are respective convex and concave relaxations of $\bar{\boldsymbol{f}}$. Thus, Theorem 2.1 applies in this case, and $\boldsymbol{v}^{-\mathrm{cv}}, \boldsymbol{v}^{-\mathrm{cc}}$ in (2.10) are convex and concave relaxations of $\boldsymbol{v}^{-1}$ on P. □

As with Theorem 2.1, note that optimization problems in (2.10a) and (2.10b) are convex NLPs. In addition, if $\boldsymbol{v}^{cv}, \boldsymbol{v}^{cc}$ are affine or piecewise-affine relaxations, then (2.10a) and (2.10b) are actually LPs, which maybe be solved efficiently.

2.4 Convex Relaxations of Constraint Satisfaction Problems

In this section, we generalize the convex relaxation methodology described in (2.3) from implicit functions to constraint satisfaction problems (CSPs). Relaxations of the point-to-set mappings in CSPs will be presented, and the directional derivatives of these relaxations will be constructed as well. Throughout this section, consider continuously differentiable mappings $\boldsymbol{g} : \mathbb{R}^{n_x+n_p} \to \mathbb{R}^{n_g}$ and $\boldsymbol{h} : \mathbb{R}^{n_x+n_p} \to \mathbb{R}^{n_h}$. Unlike the function $\boldsymbol{f}$ considered in Section 2.3, the dimensions of the codomains of $\boldsymbol{g}$ and $\boldsymbol{h}$ are arbitrary and may be distinct from n_p. Given known intervals $X \in \mathbb{IR}^{n_x}$ and $P \in \mathbb{IR}^{n_p}$, consider the following CSP:

$$\begin{aligned} \min_{\boldsymbol{z}\in X,\, \boldsymbol{p}\in P} \quad & 0 \\ \text{subject to} \quad & \boldsymbol{g}(\boldsymbol{z},\boldsymbol{p}) \leq \boldsymbol{0}, \\ & \boldsymbol{h}(\boldsymbol{z},\boldsymbol{p}) = \boldsymbol{0}. \end{aligned} \tag{2.11}$$

Let the set of feasible $\boldsymbol{z}$-values in X be expressed as a point-to-set map Ξ from $\mathbb{R}^{n_p}$ to $\mathbb{R}^{n_x}$ such that, for each $\boldsymbol{p} \in P$,

$$\Xi(\boldsymbol{p}) := \{\boldsymbol{\xi} \in X \mid \boldsymbol{g}(\boldsymbol{\xi},\boldsymbol{p}) \leq \boldsymbol{0},\ \boldsymbol{h}(\boldsymbol{\xi},\boldsymbol{p}) = \boldsymbol{0}\}. \tag{2.12}$$

Let $\boldsymbol{g}^{cv} : \mathbb{R}^{n_x+n_p} \to \mathbb{R}^{n_p}$ be a convex relaxation of $\boldsymbol{g}$ on $X \times P$, and $\boldsymbol{h}^{cv}, \boldsymbol{h}^{cc} : \mathbb{R}^{n_x+n_p} \to \mathbb{R}^{n_p}$ be convex and concave relaxations of $\boldsymbol{h}$ on $X \times P$, respectively.

Define $\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc} : \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$ such that, for each $i \in \{1, \ldots, n_x\}$ and $\boldsymbol{p} \in P$,

$$\begin{aligned} \xi_i^{cv}(\boldsymbol{p}) = \quad & \min_{\boldsymbol{\xi} \in X} \quad \xi_i \\ & \text{subject to} \quad \boldsymbol{g}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0}, \\ & \qquad\qquad \boldsymbol{h}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0} \leq \boldsymbol{h}^{cc}(\boldsymbol{\xi}, \boldsymbol{p}), \end{aligned} \tag{2.13a}$$

$$\begin{aligned} \xi_i^{cc}(\boldsymbol{p}) = \quad & \max_{\boldsymbol{\xi} \in X} \quad \xi_i \\ & \text{subject to} \quad \boldsymbol{g}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0}, \\ & \qquad\qquad \boldsymbol{h}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0} \leq \boldsymbol{h}^{cc}(\boldsymbol{\xi}, \boldsymbol{p}). \end{aligned} \tag{2.13b}$$

The optimization problems in (2.13a) and (2.13b) are convex NLPs, which are generally easier to solve than the original nonconvex CSP in (2.11).

Define an interval-valued point-to-set map $\Xi^R : \mathbb{R}^{n_p} \rightrightarrows \mathbb{R}^{n_x}$ such that, for each $\boldsymbol{p} \in P$,

$$\Xi^R(\boldsymbol{p}) \equiv [\boldsymbol{\xi}^{cv}(\boldsymbol{p}), \boldsymbol{\xi}^{cc}(\boldsymbol{p})].$$

We will verify that Ξ^R is a convex relaxation of Ξ on P. According to Definition 2.3, it suffices to show that, for all $\boldsymbol{p} \in P$, $\Xi(\boldsymbol{p})$ is a subset of $\Xi^R(\boldsymbol{p})$ and Ξ^R is convex on P.

Theorem 2.3. *Suppose that $\Xi(\boldsymbol{p})$ is nonempty for all $\boldsymbol{p} \in P$. Then, Ξ^R is a convex relaxation of Ξ on P.*

Proof. According to Definition 2.3, we will proceed by showing that $\Xi(\boldsymbol{p}) \subseteq \Xi^R(\boldsymbol{p})$ for each $\boldsymbol{p} \in P$, and that Ξ^R is convex on P.

First, choose any $\boldsymbol{p} \in P$. Since $\Xi(\boldsymbol{p})$ is nonempty, consider an arbitrary $\boldsymbol{z} \in$

$\Xi(\boldsymbol{p})$. Define $\boldsymbol{\omega} : P \to X$ such that, for $i \in \{1, \ldots, n_x\}$,

$$\omega_i(\boldsymbol{q}) = \min_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{q}) \leq \boldsymbol{0}, \ \boldsymbol{h}(\boldsymbol{\xi}, \boldsymbol{q}) = \boldsymbol{0}. \tag{2.14}$$

Choose any $i \in \{1, \ldots, n_x\}$. Since $\boldsymbol{z} \in \Xi(\boldsymbol{p})$ as in (2.12), $\omega_i(\boldsymbol{p}) \leq z_i$. Moreover, observe that the optimization problem in (2.13a) is a convex relaxation of (2.14) in the sense of [133, Definition 2], so $\xi_i^{cv}(\boldsymbol{p}) \leq \omega_i(\boldsymbol{p}) \leq z_i$. It is analogous to show that $\xi_i^{cc}(\boldsymbol{p}) \geq z_i$ for each $i \in \{1, \ldots, n_x\}$. Hence, $\boldsymbol{\xi}^{cv}(\boldsymbol{p}) \leq \boldsymbol{z} \leq \boldsymbol{\xi}^{cc}(\boldsymbol{p})$, and so $\boldsymbol{z} \in \Xi^R(\boldsymbol{p})$. Thus, $\Xi(\boldsymbol{p})$ is a subset of $\Xi^R(\boldsymbol{p})$ for each $\boldsymbol{p} \in P$.

Next, we demonstrate the convexity of Ξ^R on P. Define $\boldsymbol{\phi} : X \times P \to \mathbb{R}^{3n_x}$ such that, for each $\boldsymbol{\xi} \in X$ and $\boldsymbol{p} \in P$, $\boldsymbol{\phi}(\boldsymbol{\xi}, \boldsymbol{p}) = (\boldsymbol{g}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}), \boldsymbol{h}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}), -\boldsymbol{h}^{cc}(\boldsymbol{\xi}, \boldsymbol{p}))$, which is convex on $X \times P$. For each $i \in \{1, \ldots, n_x\}$, (2.13a) is equivalent to

$$\xi_i^{cv}(\boldsymbol{p}) = \min_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \quad \boldsymbol{\phi}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0}. \tag{2.15}$$

Observe that any point $\boldsymbol{\xi} \in \Xi(\boldsymbol{p})$ is feasible in the optimization problem (2.15). Since the objective function of (2.15) is linear, $\boldsymbol{\phi}$ is convex on $X \times P$, and X, P are convex, the convexity of ξ_i^{cv} on P follows from Proposition 2.1. It is analogous to show that ξ_i^{cc} is concave on P.

Consider any $\boldsymbol{p}_A, \boldsymbol{p}_B \in P$ and $\lambda \in (0, 1)$. The convexity of $\boldsymbol{\xi}^{cv}$ and the concavity of $\boldsymbol{\xi}^{cc}$ ensure that

$$\lambda \boldsymbol{\xi}^{cv}(\boldsymbol{p}_A) + (1 - \lambda)\boldsymbol{\xi}^{cv}(\boldsymbol{p}_B) \geq \boldsymbol{\xi}^{cv}(\lambda \boldsymbol{p}_A + (1 - \lambda)\boldsymbol{p}_B),$$
$$\lambda \boldsymbol{\xi}^{cc}(\boldsymbol{p}_A) + (1 - \lambda)\boldsymbol{\xi}^{cc}(\boldsymbol{p}_B) \leq \boldsymbol{\xi}^{cc}(\lambda \boldsymbol{p}_A + (1 - \lambda)\boldsymbol{p}_B).$$

Consider any $\boldsymbol{z}_{\boldsymbol{p}_A} \in \Xi(\boldsymbol{p}_A)$ and $\boldsymbol{z}_{\boldsymbol{p}_B} \in \Xi(\boldsymbol{p}_B)$. $\Xi(\boldsymbol{p})$ being a subset of $\Xi^R(\boldsymbol{p})$ for each $\boldsymbol{p} \in P$ ensures that $\boldsymbol{z}_{\boldsymbol{p}_A} \in \Xi^R(\boldsymbol{p}_A)$ and $\boldsymbol{z}_{\boldsymbol{p}_B} \in \Xi^R(\boldsymbol{p}_B)$. Then,

$$\lambda \boldsymbol{z}_{\boldsymbol{p}_A} + (1-\lambda)\boldsymbol{z}_{\boldsymbol{p}_B} \geq \lambda \boldsymbol{\xi}^{cv}(\boldsymbol{p}_A) + (1-\lambda)\boldsymbol{\xi}^{cv}(\boldsymbol{p}_B) \geq \boldsymbol{\xi}^{cv}(\lambda \boldsymbol{p}_A + (1-\lambda)\boldsymbol{p}_B),$$
$$\lambda \boldsymbol{z}_{\boldsymbol{p}_A} + (1-\lambda)\boldsymbol{z}_{\boldsymbol{p}_B} \leq \lambda \boldsymbol{\xi}^{cc}(\boldsymbol{p}_A) + (1-\lambda)\boldsymbol{\xi}^{cc}(\boldsymbol{p}_B) \leq \boldsymbol{\xi}^{cc}(\lambda \boldsymbol{p}_A + (1-\lambda)\boldsymbol{p}_B),$$

which shows that

$$\lambda \boldsymbol{z}_{\boldsymbol{p}_A} + (1-\lambda)\boldsymbol{z}_{\boldsymbol{p}_B} \in \Xi^R(\lambda \boldsymbol{p}_A + (1-\lambda)\boldsymbol{p}_B).$$

Since $\lambda, \boldsymbol{z}_{\boldsymbol{p}_A}, \boldsymbol{z}_{\boldsymbol{p}_B}$ were arbitrarily chosen, and since $\lambda \boldsymbol{z}_{\boldsymbol{p}_A} + (1-\lambda)\boldsymbol{z}_{\boldsymbol{p}_B}$ is an arbitrary point in the Minkowski sum $\lambda \Xi^R(\boldsymbol{p}_A) + (1-\lambda)\Xi^R(\boldsymbol{p}_B)$, it follows that

$$\lambda \Xi^R(\boldsymbol{p}_A) + (1-\lambda)\Xi^R(\boldsymbol{p}_B) \subset \Xi^R(\lambda \boldsymbol{p}_A + (1-\lambda)\boldsymbol{p}_B).$$

Thus, according to Definition 2.3, Ξ^R is convex on P. □

2.4.1 Directional Derivatives

In the previous section, we constructed a pair of convex and concave functions to enclose the point-to-set mapping defined by a CSP. In this section, we describe the directional derivatives of these convex and concave relaxations. Since implicit functions may be considered as a special type of CSPs with only equality constraints, and since (2.3) is a variant of (2.13), the method presented in this section also applies to the relaxations in (2.3).

Define a function $\boldsymbol{\psi} : X \times P \to \mathbb{R}^{n_g + 2n_h}$ such that, for each $\boldsymbol{\xi} \in X$ and $\boldsymbol{p} \in P$,

$\psi(\boldsymbol{\xi}, \boldsymbol{p}) = (\boldsymbol{g}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}), \boldsymbol{h}^{cv}(\boldsymbol{\xi}, \boldsymbol{p}), -\boldsymbol{h}^{cc}(\boldsymbol{\xi}, \boldsymbol{p}))$. Then, (2.13) becomes

$$\xi_i^{cv}(\boldsymbol{p}) = \min_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \quad \psi(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0}, \tag{2.16a}$$

$$\xi_i^{cc}(\boldsymbol{p}) = \max_{\boldsymbol{\xi} \in X} \ \xi_i \quad \text{subject to} \quad \psi(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0}. \tag{2.16b}$$

We use results from [107] to evaluate directional derivatives of the constructed convex relaxations in (2.16a) for an arbitrary $i \in \{1, \ldots, n_x\}$; a similar approach can be adapted for concave relaxations in (2.16b). Given any $\boldsymbol{p}^* \in P$, it will be shown that for each $\boldsymbol{p}$ near $\boldsymbol{p}^*$, (2.16a) has a (global) solution $\boldsymbol{\xi}(\boldsymbol{p})$ near $\boldsymbol{\xi}^*$ under relevant assumptions. Furthermore, $\boldsymbol{\xi}(\cdot)$ is B-differentiable, and its directional derivative may be computed by solving a convex quadratic program. To proceed, we first introduce notations that will be used in this section.

For each $\boldsymbol{p} \in P$, let $\boldsymbol{y}(\boldsymbol{p}) = (\boldsymbol{\xi}(\boldsymbol{p}), \boldsymbol{p})$. Since (2.16a) is a convex optimization problem with inequality constraints only, let $M(\boldsymbol{p})$ be the set of multipliers $\boldsymbol{\lambda} \in \mathbb{R}^{n_g+2n_h}$ that satisfies the Karush-Kuhn-Tucker (KKT) conditions at $\boldsymbol{y}(\boldsymbol{p})$:

$$\begin{aligned} \boldsymbol{e}_{(i)} + \sum_{k=1}^{n_g+2n_h} \lambda_k \nabla_{\boldsymbol{\xi}} \psi_k(\boldsymbol{y}(\boldsymbol{p})) = \boldsymbol{0}, \\ \psi(\boldsymbol{y}(\boldsymbol{p})) \leq \boldsymbol{0}, \quad \boldsymbol{\lambda} \geq \boldsymbol{0}, \quad \langle \boldsymbol{\lambda}, \psi(\boldsymbol{y}(\boldsymbol{p})) \rangle = 0, \end{aligned} \tag{2.17}$$

where $\boldsymbol{e}_{(i)}$ is the ith column of the identity matrix $E \in \mathbb{R}^{n_x \times n_x}$.

For each $\boldsymbol{p} \in P$, denote the set of active inequality constraint indices [107] as $I(\boldsymbol{p}) = \{j : \psi_j(\boldsymbol{y}(\boldsymbol{p})) = 0\}$. Given $\boldsymbol{\lambda} \in M(\boldsymbol{p})$, let $I_\lambda^+(\boldsymbol{p}) = \{j : \lambda_j > 0\}$ and $I_\lambda^0(\boldsymbol{p}) = I(\boldsymbol{p}) \backslash I_\lambda^+(\boldsymbol{p})$. The *critical cone* of the constraints $\psi \leq \boldsymbol{0}$ at $\boldsymbol{y}(\boldsymbol{p})$ with respect

to $\boldsymbol{\lambda}$ is

$$K_\lambda(\boldsymbol{p}) := \{\boldsymbol{\omega} \in \mathbb{R}^{n_x+n_p} : \forall k \in I_\lambda^0(\boldsymbol{p}), \langle \nabla \psi_k(\boldsymbol{y}(\boldsymbol{p})), \boldsymbol{\omega} \rangle \leq 0, \\ \forall j \in I_\lambda^+(\boldsymbol{p}), \langle \nabla \psi_j(\boldsymbol{y}(\boldsymbol{p})), \boldsymbol{\omega} \rangle = 0, \}.$$

The critical cone at $\boldsymbol{y}(\boldsymbol{p})$ with respect to $\boldsymbol{\lambda}$ in the direction $\boldsymbol{d} \in \mathbb{R}^{n_p}$ is

$$K_\lambda(\boldsymbol{p}; \boldsymbol{d}) := \{\boldsymbol{\nu} \in \mathbb{R}^{n_x} \mid (\boldsymbol{\nu}, \boldsymbol{d}) \in K_\lambda(\boldsymbol{p})\}.$$

The Lagrangian of (2.16a) at $\boldsymbol{p}$ is:

$$L(\boldsymbol{y}; \lambda) \equiv y_i + \langle \boldsymbol{\lambda}, \boldsymbol{\psi}(\boldsymbol{y}) \rangle.$$

Lastly, given $\boldsymbol{d} \in \mathbb{R}^{n_p}$, define a subset $S(\boldsymbol{p}; \boldsymbol{d})$ of $M(\boldsymbol{p})$ such that

$$S(\boldsymbol{p}; \boldsymbol{d}) \equiv \arg\max_{\boldsymbol{\lambda}} \ \boldsymbol{\lambda}^\top \nabla_{\boldsymbol{p}} \boldsymbol{\psi}(\boldsymbol{y}(\boldsymbol{p}))\, \boldsymbol{d} \quad \text{subject to} \quad \boldsymbol{\lambda} \in M(\boldsymbol{p}).$$

The following assumption is adapted from Assumptions (A1)-(A4) in [107] for the convex optimization problem (2.16a).

Assumption 2.3. *For each $j \in \{1, \ldots, n_g + 2n_h\}$, assume the following conditions hold:*

1. *$\boldsymbol{\psi}$ is twice-continuously differentiable near $(\boldsymbol{\xi}^*, \boldsymbol{p}^*) \in \mathbb{R}^{n_x+n_p}$, where $\boldsymbol{\xi}^*$ is a locally optimal objective value of* (2.16a) *of $\xi_i^{cv}(\boldsymbol{p}^*)$.*

2. *There exists $\boldsymbol{\nu} \in \mathbb{R}^{n_x}$ such that, if $\psi_i(\boldsymbol{y}(\boldsymbol{p}^*)) = 0$, then $\langle \nabla_{\boldsymbol{\xi}} \psi_i(\boldsymbol{y}(\boldsymbol{p}^*)), \boldsymbol{\nu} \rangle < 0$.*

3. *For each $\boldsymbol{\lambda}$ that satisfies the KKT conditions* (2.17) *at $\boldsymbol{y}(\boldsymbol{p}^*)$, and each $\boldsymbol{\nu} \neq \boldsymbol{0}$ such*

that $\langle \nabla_{\boldsymbol{\xi}} \psi_i(\boldsymbol{\xi}, \boldsymbol{p}), \boldsymbol{\nu} \rangle = 0$ *if* $\lambda_i > 0$*, it holds that*

$$\boldsymbol{\nu}^\top \nabla^2_{\boldsymbol{\xi}\boldsymbol{\xi}} \langle \boldsymbol{\lambda}, \psi(\boldsymbol{y}(\boldsymbol{p}^*)) \rangle \boldsymbol{\nu} > 0.$$

4. *There exists a neighborhood* W *of* $\boldsymbol{y}(\boldsymbol{p}^*)$ *such that for any subset* I *of* $I(\boldsymbol{p}^*) \equiv \{j : \psi_j(\boldsymbol{p}^*) = 0\}$*, the collection of partial derivative matrices* $\{\nabla_{\boldsymbol{\xi}} \psi_j(\boldsymbol{y}) : j \in I\}$ *has the same rank for all vectors* $\boldsymbol{y} \in W$*.*

The following theorem describes directional derivatives for the optimization problem in (2.16a). It is adapted from Theorem 2 in [107].

Theorem 2.4. *Suppose that Assumption 2.3 holds. Then, for respective neighborhoods* P^* *of* $\boldsymbol{p}^*$ *and* X^* *of* $\boldsymbol{\xi}^*$*, there is a function* $\boldsymbol{\xi} : P^* \to X^*$ *that satisfies all of the following conditions:*

1. $\boldsymbol{\xi}$ *is continuous, and for each* $\boldsymbol{p} \in P^*$*,* $\boldsymbol{\xi}(\boldsymbol{p})$ *is the unique solution of* (2.16a) *in* X^**,*
2. $\boldsymbol{\xi}$ *is a piecewise-differentiable function, and hence locally Lipschitz continuous, and*
3. *The directional derivative* $\boldsymbol{\xi}'(\boldsymbol{p}; \cdot)$ *is a piecewise linear function such that for each* $\boldsymbol{p} \in P^*$*,* $\boldsymbol{d} \in \mathbb{R}^{n_p}$*, and* $\boldsymbol{\lambda} \in S(\boldsymbol{p}; \boldsymbol{d})$*,* $\boldsymbol{\xi}'(\boldsymbol{p}; \boldsymbol{d})$ *is the unique solution of the following convex quadratic program:*

$$\begin{aligned} \min_{\boldsymbol{\nu}} \quad & \frac{1}{2} \boldsymbol{\nu}^\top \nabla^2_{\boldsymbol{\xi}\boldsymbol{\xi}} \langle \boldsymbol{\lambda}, \psi(\boldsymbol{y}(\boldsymbol{p})) \rangle \boldsymbol{\nu} + \boldsymbol{d}^\top \nabla^2_{\boldsymbol{\xi}\boldsymbol{p}} \langle \boldsymbol{\lambda}, \psi(\boldsymbol{y}(\boldsymbol{p})) \rangle \boldsymbol{\nu} \\ \text{subject to} \quad & \boldsymbol{\nu} \in K_\lambda(\boldsymbol{p}; \boldsymbol{d}). \end{aligned} \tag{2.18}$$

2.5 Tightening Interval Bounds

In the previous sections, convex and concave relaxations of implicit functions and CSPs are constructed on known intervals, i.e., X in Assumption 2.1 and (2.11). In this section, we adapt the formulation in (2.13) to generate new interval bounds for implicit functions and CSPs that are at least as tight as the original bounds. These tighter intervals can then be used to construct relaxations of implicit functions and CSPs that are tighter than those constructed with the original intervals. This is because the convex relaxations of the original residual function, constructed using methods like αBB and McCormick relaxations, will converge quickly to the original function when the intervals shrink. Examples 2.1 and 2.3 in Section 2.7 illustrate such applications. With this foundation, we allow the interval X that convex and concave relaxations were constructed on to be varied, and add it as a superscript to these relaxations.

Define $\Xi^B \equiv [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U] \in \mathbb{IR}^{n_x}$ such that for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} \xi_i^L = \quad & \min_{\boldsymbol{\xi} \in X,\, \boldsymbol{p} \in P} \quad \xi_i \\ & \text{subject to} \quad \boldsymbol{g}^{\mathrm{cv},X}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0}, \\ & \qquad\qquad \boldsymbol{h}^{\mathrm{cv},X}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0} \leq \boldsymbol{h}^{\mathrm{cc},X}(\boldsymbol{\xi}, \boldsymbol{p}), \end{aligned} \tag{2.19a}$$

$$\begin{aligned} \xi_i^U = \quad & \max_{\boldsymbol{\xi} \in X,\, \boldsymbol{p} \in P} \quad \xi_i \\ & \text{subject to} \quad \boldsymbol{g}^{\mathrm{cv},X}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0}, \\ & \qquad\qquad \boldsymbol{h}^{\mathrm{cv},X}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0} \leq \boldsymbol{h}^{\mathrm{cc},X}(\boldsymbol{\xi}, \boldsymbol{p}). \end{aligned} \tag{2.19b}$$

We will show that, given a rough enclosure X, (2.19) describes refined interval

bounds for the feasible region in (2.12) for all $\boldsymbol{p} \in P$, and they are as least as tight as X.

Theorem 2.5. *Let* $\Xi^R(\boldsymbol{p}) \equiv [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}]$ *be a solution of* (2.13). *Then,* $\Xi^B \equiv [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U]$ *in* (2.19) *satisfy satisfies the following inclusions. For all* $\boldsymbol{p} \in P$,

$$\Xi^R(\boldsymbol{p}) \subseteq \Xi^B \subseteq X.$$

Proof. From (2.13) and (2.19), observe that, for any $i \in \{1, \ldots, n_x\}$,

$$\xi_i^L = \min_{\boldsymbol{p} \in P} \ \xi_i^{cv}(\boldsymbol{p}), \quad \text{and} \quad \xi_i^U = \max_{\boldsymbol{p} \in P} \ \xi_i^{cc}(\boldsymbol{p}).$$

Hence, $\Xi^R(\boldsymbol{p}) \subseteq \Xi^B$ for all $\boldsymbol{p} \in P$. Since (2.19) guarantees that $\boldsymbol{\xi}^L, \boldsymbol{\xi}^U \in X$, $\Xi^B \subseteq X$. □

The approach described in (2.19) may be used iteratively to further improve the tightness of intervals that enclose implicit functions and the point-to-set mappings in CSPs. Let interval $\Xi^{B,0}$ be an initial rough enclosure of implicit functions or the point-to-set mappings in CSPs, in place of X. Let $\Xi^{B,k} \equiv [\boldsymbol{\xi}^{L,k}, \boldsymbol{\xi}^{U,k}]$ be computed

by solving the following for $k \in \{1, 2, \dots\}$: for each $i \in \{1, \dots, n_x\}$,

$$\begin{aligned} \xi_i^{L,k} = \quad & \min_{\boldsymbol{\xi} \in \Xi^{B,k-1}, \boldsymbol{p} \in P} \quad \xi_i \\ & \text{subject to} \quad \boldsymbol{g}^{\mathrm{cv},\Xi^{B,k-1}}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0}, \qquad (2.20a) \\ & \qquad\qquad \boldsymbol{h}^{\mathrm{cv},\Xi^{B,k-1}}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0} \leq \boldsymbol{h}^{\mathrm{cc},\Xi^{B,k-1}}(\boldsymbol{\xi}, \boldsymbol{p}), \\ \xi_i^{U,k} = \quad & \max_{\boldsymbol{\xi} \in \Xi^{B,k-1}, \boldsymbol{p} \in P} \quad \xi_i \\ & \text{subject to} \quad \boldsymbol{g}^{\mathrm{cv},\Xi^{B,k-1}}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0}, \qquad (2.20b) \\ & \qquad\qquad \boldsymbol{h}^{\mathrm{cv},\Xi^{B,k-1}}(\boldsymbol{\xi}, \boldsymbol{p}) \leq \boldsymbol{0} \leq \boldsymbol{h}^{\mathrm{cc},\Xi^{B,k-1}}(\boldsymbol{\xi}, \boldsymbol{p}). \end{aligned}$$

Theorem 2.5 illustrates that $\Xi^{B,k} \subseteq \Xi^{B,k-1} \subseteq \cdots \subseteq \Xi^{B,0} \equiv X$. Thus, (2.20) presents a method to iteratively compute intervals for implicit functions and CSPs that are at least as tight as a known rough enclosure.

2.6 Relaxations of Numerical ODE Solutions

In this section, we construct convex and concave relaxations for implicit functions that are numerical solutions of parametric ordinary differential equations (ODEs), computed using implicit integration methods. Compared with explicit integration methods, implicit integration methods are more stable when dealing with stiff ODEs [136]. While methods have been established in [136, 153] to construct convex relaxations for implicit numerical solutions of ODEs, this section introduces an alternative approach that may construct tighter relaxations, as illustrated in Example 2.3 in Section 2.7. Such convex relaxations are useful in the deterministic global optimization of dynamic systems.

Define $t_0, t_f \in \mathbb{R}$ such that $t_0 < t_f$, and let $I = (t_0, t_f]$. Given $\boldsymbol{z}^0 \in \mathbb{R}^{n_z}$ and a continuous function $\boldsymbol{u} : I \times P \times \mathbb{R}^{n_z} \to \mathbb{R}^{n_z}$, consider an initial-value problem:

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}t}(t, \boldsymbol{p}) &= \boldsymbol{u}(t, \boldsymbol{p}, \boldsymbol{z}(t, \boldsymbol{p})), \quad t \in I, \\ \boldsymbol{z}(t_0, \boldsymbol{p}) &= \boldsymbol{z}^0. \end{aligned} \tag{2.21}$$

According to Peano's Theorem summarized in [60, Theorem 2.1, Chapter II], (2.21) has at least one solution. We will use the implicit Euler method to obtain a numerical solution and generate it convex relaxations using the approach of Section 2.3. Similar approaches can be applied to other implicit integration methods, such as the Adams–Moulton method and the BDF method [153]. To solve (2.21) with the implicit Euler method at an arbitrary $\boldsymbol{p} \in P$, we first discretize I into n evenly spaced intervals with length $\Delta t = (t_f - t_0)/n$ and $\{0, \ldots, n\}$ mesh points. Denote the ODE solution value at a mesh point as $\boldsymbol{z}^m$ for each $m \in \{0, \ldots, n\}$. Then, (2.21) can be approximated by a nonlinear equation system: for all $m \in \{1, \ldots, n\}$ and $\boldsymbol{p} \in P$,

$$\boldsymbol{z}^m(\boldsymbol{p}) - \boldsymbol{z}^{m-1}(\boldsymbol{p}) - \Delta t\, \boldsymbol{u}(m\,\Delta t, \boldsymbol{p}, \boldsymbol{z}^m(\boldsymbol{p})) = \boldsymbol{0}. \tag{2.22}$$

where $\boldsymbol{z}^0(\boldsymbol{p}) = \boldsymbol{z}^0$ is the known initial condition. (2.22) actually defines an implicit function

$$\boldsymbol{x}(\boldsymbol{p}) \equiv \begin{bmatrix} \boldsymbol{z}^1(\boldsymbol{p}) \\ \vdots \\ \boldsymbol{z}^n(\boldsymbol{p}) \end{bmatrix}$$

following (2.2) if we let

$$\begin{aligned}\boldsymbol{f}((\boldsymbol{z}^1(\boldsymbol{p}),\ldots,\boldsymbol{z}^n(\boldsymbol{p}))^\top,\boldsymbol{p}) &\equiv \begin{bmatrix} \boldsymbol{z}^1(\boldsymbol{p}) - \boldsymbol{z}^0(\boldsymbol{p}) - \Delta t\,\boldsymbol{u}(\Delta t,\boldsymbol{p},\boldsymbol{z}^1(\boldsymbol{p})) \\ \vdots \\ \boldsymbol{z}^n(\boldsymbol{p}) - \boldsymbol{z}^{n-1}(\boldsymbol{p}) - \Delta t\,\boldsymbol{u}(n\Delta t,\boldsymbol{p},\boldsymbol{z}^n(\boldsymbol{p})) \end{bmatrix} \\ &= \boldsymbol{0}. \end{aligned} \tag{2.23}$$

Thus, we can use Theorem 2.1 to construct convex and concave relaxations for $\boldsymbol{z}^n$ on P, where $\boldsymbol{z}^n$ is the numerical solution value of ODE (2.21) at its terminal time.

Let $Z \equiv [\boldsymbol{z}^L, \boldsymbol{z}^U] \in \mathbb{R}^{n_z}$ be a known interval bound so that $\boldsymbol{z}(t,\boldsymbol{p}) \in Z$ for all $(t,\boldsymbol{p}) \in I \times P$. Define $Z^{m,0} \equiv [\boldsymbol{z}^{m,0,L}, \boldsymbol{z}^{m,0,U}] \subseteq \mathbb{IR}^{n_z}$ as the initial interval bounds of $\boldsymbol{z}^m$ for each $m \in \{1,\ldots,n\}$, where the superscript m represents the mesh point index and 0 means that its is a initial rough enclosure (similar to the notation in Section 2.5). Since a conservative interval bound Z is known, we set $Z^{m,0} = Z$ for each $m \in \{1,\ldots,n\}$, and it follows that $\boldsymbol{z}^m(\boldsymbol{p}) \in Z^{m,0}$ for each $m \in \{1,\ldots,n\}$ and $\boldsymbol{p} \in P$. Then, convex and concave relaxations of $\boldsymbol{z}^n$ on P can be computed using

Theorem 2.1: for each $j \in \{1, \ldots, n_z\}$,

$$
\begin{aligned}
z_j^{n,\mathrm{cv}}(\boldsymbol{p}) = \min_{\substack{\boldsymbol{\zeta}^m \in Z^{m,0}, \\ m \in \{1,\ldots,n\}}} \quad & \zeta_j^n, \\
\text{subject to} \quad & f_{i+n_z(m-1)}^{\mathrm{cv},Z^{m,0}}((\boldsymbol{\zeta}^1, \ldots, \boldsymbol{\zeta}^n)^\top, \boldsymbol{p}) \\
& \qquad \leq 0 \leq f_{i+n_z(m-1)}^{\mathrm{cc},Z^{m,0}}((\boldsymbol{\zeta}^1, \ldots, \boldsymbol{\zeta}^n)^\top, \boldsymbol{p}), \\
& \forall i \in \{1, \ldots, n_z\}, m \in \{1, \ldots, n\}, \\
z_j^{n,\mathrm{cc}}(\boldsymbol{p}) = \max_{\substack{\boldsymbol{\zeta}^m \in Z^{m,0}, \\ m \in \{1,\ldots,n\}}} \quad & \zeta_j^n, \\
\text{subject to} \quad & f_{i+n_z(m-1)}^{\mathrm{cv},Z^{m,0}}((\boldsymbol{\zeta}^1, \ldots, \boldsymbol{\zeta}^n)^\top, \boldsymbol{p}) \\
& \qquad \leq 0 \leq f_{i+n_z(m-1)}^{\mathrm{cc},Z^{m,0}}((\boldsymbol{\zeta}^1, \ldots, \boldsymbol{\zeta}^n)^\top, \boldsymbol{p}), \\
& \forall i \in \{1, \ldots, n_z\}, m \in \{1, \ldots, n\},
\end{aligned}
\tag{2.24}
$$

where $\boldsymbol{f}^{\mathrm{cv},Z^{m,0}}$ and $\boldsymbol{f}^{\mathrm{cc},Z^{m,0}} : \mathbb{R}^{n_z \times n + n_p} \to \mathbb{R}^{n_z \times n}$ are convex and concave relaxations of $\boldsymbol{f}$ in (2.23), respectively, constructed on interval $Z^{m,0}$.

Furthermore, we may use the formulation in (2.20) to construct improved interval bounds $Z^{m,1} \equiv [\boldsymbol{z}^{m,L,1}, \boldsymbol{z}^{m,U,1}]$ of $\boldsymbol{z}^m$ for each $m \in \{1, \ldots, n\}$, where m denotes the index of mesh point and 1 is used in place of k in (2.20) to represent one iteration of refinement. As discussed in Section 2.5, these improved intervals are at least as tight as the original interval $Z^{m,0}$, and they will lead to tighter relaxations for implicit functions and CSPs. In this case, we can use these tighter intervals to generate tighter relaxations for the numerical solutions of ODEs by replacing $Z^{m,0}$ in (2.24) with $Z^{m,1}$. This result is illustrated in Example 2.3 in Section 2.7.

Similar to Section 2.5, the intervals where convex and concave relaxations are

constructed on are added as superscripts to these relaxations in the following formulation. For each $m \in \{1,\ldots,n\}$, consider $Z^{m,1} \equiv [\boldsymbol{z}^{m,L,1}, \boldsymbol{z}^{m,U,1}] \in \mathbb{R}^{n_z}$ such that, for each $j \in \{1,\ldots,n_z\}$,

$$\begin{aligned}
z_j^{m,L,1} = \min_{\substack{\boldsymbol{p}\in P, \boldsymbol{\zeta}^\kappa \in Z^{\kappa,0},\\ \kappa\in\{1,\ldots,n\}}} \quad & \zeta_j^m, \\
\text{subject to} \quad & f^{\mathrm{cv},Z^{\kappa,0}}_{i+n_z(\kappa-1)}((\boldsymbol{\zeta}^1,\ldots,\boldsymbol{\zeta}^n)^\top, \boldsymbol{p}) \\
& \qquad \leq 0 \leq f^{\mathrm{cc},Z^{\kappa,0}}_{i+n_z(\kappa-1)}((\boldsymbol{\zeta}^1,\ldots,\boldsymbol{\zeta}^n)^\top, \boldsymbol{p}), \\
& \forall i \in \{1,\ldots,n_z\}, \kappa \in \{1,\ldots,n\}, \\
z_j^{m,U,1} = \max_{\substack{\boldsymbol{p}\in P, \boldsymbol{\zeta}^\kappa \in Z^{\kappa,0},\\ \kappa\in\{1,\ldots,n\}}} \quad & \zeta_j^m, \\
\text{subject to} \quad & f^{\mathrm{cv},Z^{\kappa,0}}_{i+n_z(\kappa-1)}((\boldsymbol{\zeta}^1,\ldots,\boldsymbol{\zeta}^n)^\top, \boldsymbol{p}) \\
& \qquad \leq 0 \leq f^{\mathrm{cc},Z^{\kappa,0}}_{i+n_z(\kappa-1)}((\boldsymbol{\zeta}^1,\ldots,\boldsymbol{\zeta}^n)^\top, \boldsymbol{p}), \\
& \forall i \in \{1,\ldots,n_z\}, \kappa \in \{1,\ldots,n\}.
\end{aligned} \tag{2.25}$$

2.7 Numerical Examples

In this section, we use the approaches in previous sections to construct convex and concave relaxations, as well as improved interval bounds, for various implicit functions and parametric ODEs. These approaches were implemented in the programming language Julia [20]. The `McCormick.jl` package [152] was used to construct convex relaxations of nonconvex factorable functions following either the standard McCormick relaxations [92, 123] or the differentiable McCormick relaxations [72, 73]. All convex nonlinear programs were solved with IPOPT v3.13.2 [147] via JuMP v0.21.4 [49]. The numerical results reported below were obtained

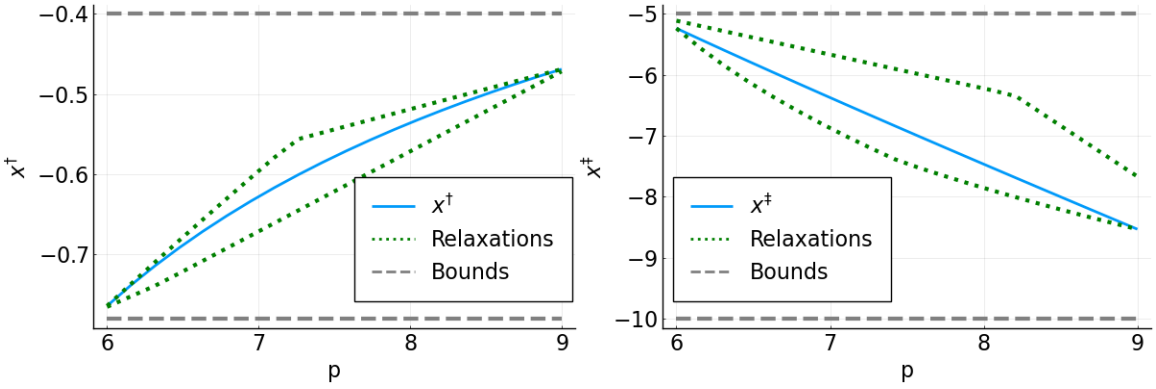
by running this implementation on a Windows 10 machine with a 3.6 GHz AMD Ryzen 5 2600X CPU and 8 GB memory.

The following example is adapted from [136, Example 3.26].

Example 2.1. *Let* $P := [6, 9]$*, and consider a function* $f(z, p) = z^2 + pz + 4$ *where the parameter* p *is an element of* P*. According to the quadratic formula, for each* $p \in P$*, there are two real roots* z^* *of the equation* $f(z, p) = 0$*. It was reported in [136] that* $X^{\dagger,0} = [-0.78, -0.4]$ *and* $X^{\ddagger,0} = [-10.0, -5.0]$ *are two interval bounds of these two real roots, respectively. In both* $X^{\dagger,0}$ *and* $X^{\ddagger,0}$*, there is a single real root* z^* *of* $f(z, p) = 0$ *for each* $p \in P$*, so we have two injective implicit functions* $x^\dagger : P \to X^{\dagger,0}$ *and* $x^\ddagger : P \to X^{\ddagger,0}$ *such that* $f(x^\dagger(p), p) = 0$ *and* $f(x^\ddagger(p), p) = 0$*.*

We generated convex and concave relaxations of $x^\dagger$ and $x^\ddagger$ on P using Theorem 2.1, and compared them with relaxations constructed using the method established in [136]. The convex and concave relaxations of f were constructed with standard McCormick relaxations [92, 123]. The minimization and maximization problems in (2.3) were solved at different $p \in P$. Their optimal values, plotted as functions of p in Figure 2.1, are convex and concave relaxations of the implicit functions $x^\dagger$ and $x^\ddagger$ on P. Moreover, it was observed that these relaxations are significantly tighter than the relaxations in [136, Figure 1] at each $p \in P$. For example, in Figure 2.1(b), our method generated a concave relaxation $x^{cc}(p) \approx -6.25$ at $p = 8$. However, in [136, Figure 1(b)], the concave relaxation of x at $p = 8$ is around -5.

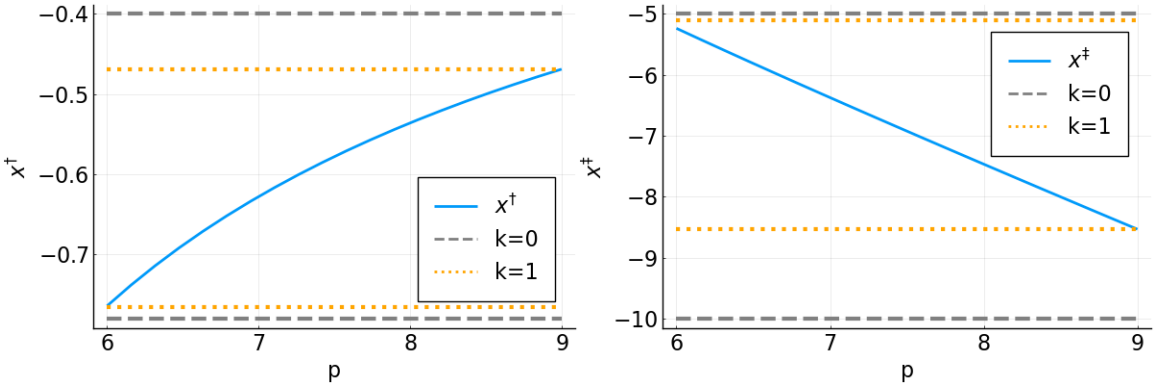
Next, we constructed improved interval bounds of $x^\dagger$ and $x^\ddagger$ on P separately. Since an implicit function may be considered as a CSP with equality constraints only, we applied the formulation in (2.20) with $k = 1$ to generate interval bounds



(a) (b)

Figure 2.1: The implicit functions $x^{\dagger}$ and $x^{\ddagger}$ in Example 2.1 (solid), along with their interval bounds (dashed) reported in [136] and convex and concave relaxations (dotted) constructed with the new method on P, plotted as functions of p.

that are tighter than the original interval bounds $X^{\dagger,0}$ and $X^{\ddagger,0}$. As shown in Figure 2.2, these improved interval bounds are significantly tighter than the original bounds.



(a) (b)

Figure 2.2: The implicit functions $x^{\dagger}$ and $x^{\ddagger}$ in Example 2.1 (solid), along with their original interval bounds $X^{\dagger,0}$ and $X^{\ddagger,0}$ (dashed) and improved interval bounds $X^{\dagger,1}$ and $X^{\ddagger,1}$ (dotted) on P, plotted as functions of p.

Furthermore, we used the improved interval bounds $X^{\dagger,1}$ and $X^{\ddagger,1}$ to generate improved relaxations for $x^\dagger$ and $x^\ddagger$, respectively, on P. These relaxations are plotted in Figure 2.3, along with the original relaxations constructed with $X^{\dagger,0}$ and $X^{\ddagger,0}$. This illustrates that tighter interval bounds produce tighter convex and concave relaxations.

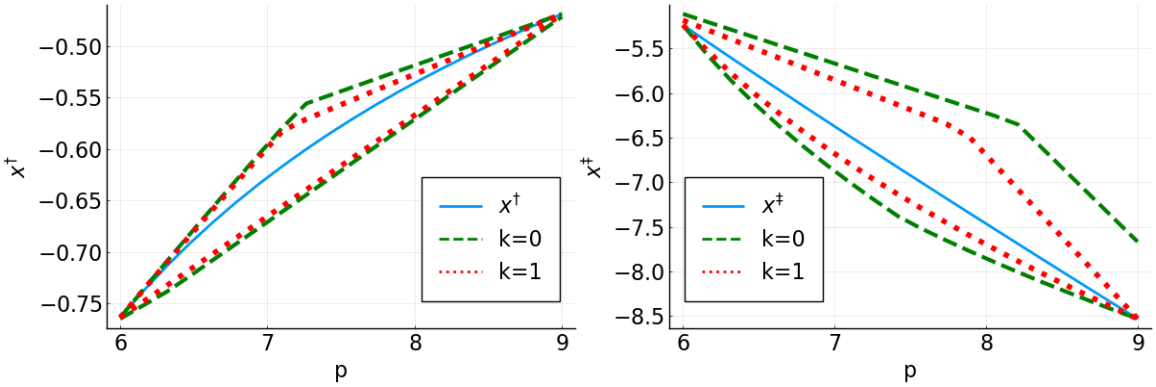
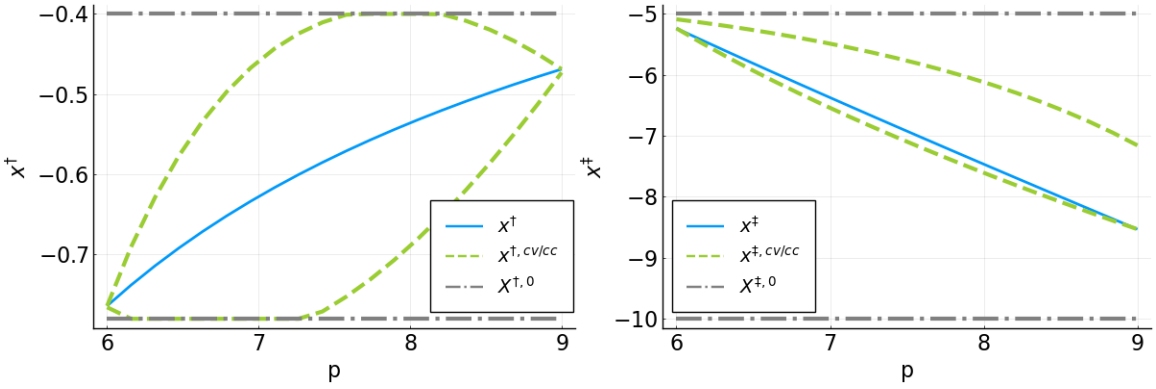


Figure 2.3: The implicit functions $x^\dagger$ and $x^\ddagger$ in Example 2.1 (solid), along with their relaxation constructed on $X^{\dagger,0}$ and $X^{\ddagger,0}$ (dashed) and improved relaxations constructed on $X^{\dagger,1}$ and $X^{\ddagger,1}$ (dotted) on P, plotted as functions of p.

In addition to McCormick relaxations, we also used αBB relaxations [3] to construct convex and concave relaxations of f. The resulting convex and concave relaxations of x on $X^{\dagger,0}$ and $X^{\ddagger,0}$ are illustrated in Figure 2.4. This illustrates the versatility of our relaxation approach. Any valid convex and concave relaxations of f can be used in (2.3), while the established method in [136] is limited to GM.

Example 2.2. *The van der Waals equation is a physical property model for describing the behavior of non-ideal gases in chemical engineering. It establishes the relationship between pressure P (*atm*), volume V (*L*), temperature T (*K*), and amount of gas n (*mole*) using*



(a) (b)

Figure 2.4: The implicit functions $x^{\dagger}$ and $x^{\ddagger}$ in Example 2.1 (solid), along with their interval bounds (dot-dashed) and convex and concave relaxations (dashed) where the relaxations of the original residual function f are constructed with αBB relaxations, plotted as functions of p.

the nonlinear equation below:

$$f(P,V) := \left(P + a\frac{n^2}{V^2}\right)(V - nb) - nRT = 0. \tag{2.26}$$

where $R = 0.082\,06\,\frac{\mathrm{L\,atm}}{\mathrm{K\,mol}}$ *is the gas constant, and* a, b *are van der Waals constants. We study the behavior of 1 mole of hydrogen gas that undergoes reversible isothermal compression from* 23.0 L *to* 22.0 L *at* 273 K. *We would like to compute guaranteed bounds on the pressures obtained during this conversion, which may be used to verify that the process operates safely. In this case,* $n = 1\,\mathrm{mole}$, $T = 273\,\mathrm{K}$, $a = 0.2444\,\frac{\mathrm{L^2\,atm}}{\mathrm{mol^2}}$, *and* $b = 0.026\,61\,\frac{\mathrm{L}}{\mathrm{mol}}$ *are constants.*

To study how pressure varies during this compression process, we consider the implicit function of pressure in terms of volume, defined to satisfy (2.26). Thus, we construct convex and concave relaxations of P on $[22.0, 23.0]$ using Theorem 2.1. The interval bound X that encloses P on $[22.0, 23.0]$ is set to $[0.9, 1.1]$ and the convex

and concave relaxations of f are constructed with GM. The generated convex and concave relaxations of P are illustrated in Figure 2.5, and appear to provide tight enclosures of the graph of P.

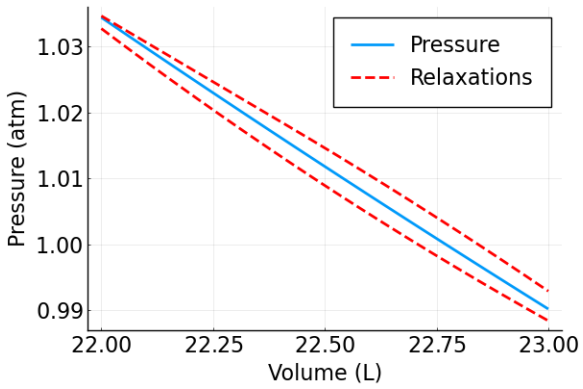


Figure 2.5: The implicit function of pressure with respect to volume in Example 2.2 (solid), along with its convex and concave relaxations (dashed)

Example 2.3. *Consider the following parametric ODE:*

$$\begin{aligned} \frac{\mathrm{d}z}{\mathrm{d}t}(t,p) &= -z^2 + p, \quad t \in (0,1], \\ z(0,p) &= 9, \end{aligned} \tag{2.27}$$

where $p \in P \equiv [-1,1]$.

The convex and concave relaxations of this ODE system are generated according to Section 2.6. Similar work has been done in [111, Section 4.1] and [153, Example 1]. We first discretize I into 20 intervals, so that $n = 20$ and $\Delta t = (t_f - t_0)/n = 0.05$. Using the implicit Euler method, ODE solution $z(\cdot, p)$ can be numerically approximated by $z^1(p), \ldots, z^{20}(p)$ for all $p \in P$. In particular, $z^{20}(p)$ is the numerical approximation of the ODE solution $z(t_f, p)$ at the terminal time for all $p \in P$. A known conservative interval bound for the ODE (2.27) is $Z = [0.1, 9]$ according to

[153], so the interval bounds of z^m on P, $Z^{m,0}$, are set to Z for each $m \in \{1,\ldots,20\}$, where the superscript 0 means that they are initial rough enclosures. Then, we generated convex and concave relaxations $z^{20,\mathrm{cv},0}(p), z^{20,\mathrm{cc},0}(p)$ on P using (2.24) to, where $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ were constructed with GM. These relaxations are plotted in Figure 2.6, and appear to be valid convex and concave relaxations of $z^{20}(p)$ on P.

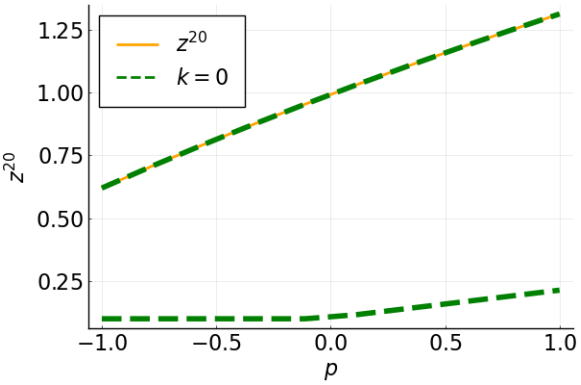
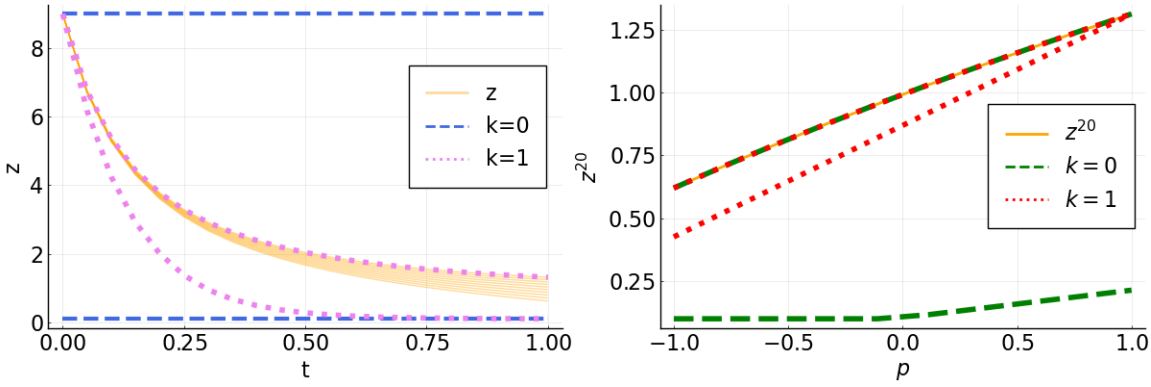


Figure 2.6: The numerical solution of (2.27) via the implicit Euler method (solid) at $t = 1$, along with its convex and concave relaxations (dashed), plotted as a function of p

Next, the formulation in (2.25) was employed to construct improved interval bounds $Z^{m,1} \equiv [\boldsymbol{z}^{m,L,1}, \boldsymbol{z}^{m,U,1}]$ of $\boldsymbol{z}^m$ for each $m \in \{1,\ldots,20\}$, where the last superscript 1 stands for one iteration of refinement. The generated lower bounds $\boldsymbol{z}^{1,L,1},\ldots,\boldsymbol{z}^{20,L,1}$ and upper bounds $\boldsymbol{z}^{1,U,1},\ldots,\boldsymbol{z}^{20,U,1}$ are plotted as the lower-bounding and upper-bounding trajectories in Figure 2.7a. Furthermore, these tighter interval bounds are used to generated tighter convex and concave relaxations $z^{20,\mathrm{cv},1}(p), z^{20,\mathrm{cc},1}(p)$ by replacing $Z^{m,0}$ in (2.24) with $Z^{m,1}$ for each $m \in \{1,\ldots,20\}$. The improved relaxations are illustrated in Figure 2.7(b).

Lastly, we compare the convex and concave relaxations illustrated in Figure 2.7(b) with those constructed with established methods [111, 153]. When $k = 0$, we



(a) (b)

Figure 2.7: (a) Interval bounds $Z^{m,0}$ (dashed) and tighter interval bounds $Z^{m,1}$ (dotted), $m \in \{1, \ldots, 20\}$, in Example 2.3. Solid lines are trajectories of $z(\cdot, p)$ in (2.27) with different p. (b) The parametric solution of (2.21) (solid), along with its convex and concave relaxations constructed on conservative interval bounds (dashed) and improved interval bounds (dotted), plotted as a function of p at $t = 1$

used very conservative interval bounds $Z^{1,0}, \ldots, Z^{20,0}$ that are much looser than the bounds used in [111]. In this case, the convex relaxation $z^{20,\mathrm{cv},0}$ in Figure 2.7(b) is looser than the convex relaxation in [111, Figure 5], but the concave relaxation $z^{20,\mathrm{cc},0}$ overlaps with the numerical solution z^{20}, and is significantly tighter than the concave relaxation in [111, Figure 5]. When $k = 1$, we used tighter interval bounds $Z^{1,1}, \ldots, Z^{20,1}$. In this case, the convex and concave relaxations, $z^{20,\mathrm{cv},1}$ and $z^{20,\mathrm{cc},1}$, are both significantly tighter than the relaxations in [111, Figure 5]. Compared with the lower and upper bounds shown in [153, Figure 4, lower left and lower right], the bounds in Figure 2.7(a) are looser. This is probably due to the difference in numerical integration methods. Instead of the naive implicit Euler method used in this work, more advanced Adams–Moulton (AM) and backward difference formula (BDF) methods were used in the implementation of [153]. In principle, the approach in Section 2.6 can be trivially adapted for the AM and BDF

methods, but we won't attempt it here due to the implementation complexity.

2.8 Conclusion

A novel approach for generating convex and concave relaxations of implicit functions has been developed in this article. These relaxations are described by the convex parametric programs shown in Theorem 2.1, whose constraints are arbitrary convex and concave relaxations of the original residual function. Using the Tsoukalas-Mitsos relaxations of compositions [144], Section 2.3.2 demonstrated that *a priori* convex and concave relaxations can be used to generate relaxations for composite functions that involve implicit functions. Furthermore, this new approach was extended to construct convex relaxations for inverse functions (Section 2.3.3) and feasible set mappings in CSPs (Section 2.4). Directional derivatives of these convex relaxations are available through solving auxiliary convex quadratic programs described in Section 2.4.1. Section 2.5 illustrated that tighter interval bounds of implicit functions and feasible regions in CSPs can be obtained by further optimizing their convex relaxations with respect to parameters. These improved interval bounds can then be used to generate tighter relaxations. Lastly, Section 2.6 demonstrated constructing convex relaxations and interval bounds for the numerical solutions of parametric ODEs using our new approach.

Unlike established methods that construct relaxations for implicit functions and CSPs, our new approach does not assume uniqueness of a solution and does not require the original residual function to be factorable. While the method in [136] requires GM and the method in [151] requires RM, our new approach admits any valid convex relaxations of the original residual function, including McCormick

relaxations [92, 123, 72], αBB relaxations [3], convex envelopes, and the pointwise best among multiple relaxations. Furthermore, while the established method in [136] depends on one particular nonlinear equation solution approach, i.e. fixed-point iteration, our new approach may employ various methods to solve the embedded optimization problems, such as LP algorithms and NLP algorithms, or even solve them analytically. This optimization-based approach is also easy to implement. A proof-of-concept Julia implementation of this approach was developed. As illustrated by the numerical examples in Section 2.7, our new approach may construct tighter relaxations of implicit functions and parametric ODEs than established methods. These properties are beneficial in applications such as global optimization and reachability analysis.

Future work may include describing subgradients for the new convex relaxations of implicit functions, for use when minimizing these relaxations during global optimization, or when constructing outer approximations. Tsoukalas and Mitsos has described subgradients of their convex relaxations in [144]. Another potential direction of future research is to extend the approach of generating convex relaxations for parametric ODEs to construct relaxations for parametric differential algebraic equations (DAEs). Semi-explicit index-1 DAEs can be approximated as CSPs using similar implicit numerical methods as discussed in Section 2.6. Then, convex relaxations can be generated following Theorem 2.3 if the parametric DAE has a solution for each $\boldsymbol{p} \in P$. However, compared with ODEs, it may be difficult to verify the existence of solutions of DAEs on the entire parameter domain due to these additional algebraic equations.

Chapter 3

A Smoothing Method for Generating Tighter Reachable Set Enclosures for Parametric Ordinary Differential Equations

This chapter is to be submitted to a journal before my anticipated thesis defense.

3.1 Introduction

The reachable set of a dynamic system is the set of possible final states that the system may attain, given a range of permitted initial conditions, parameters, or controls. This article focues on dynamic systems that are represented as parametric systems of ordinary differential equations (ODEs), as formalized in Section 3.2

below. Methods for constructing reachable sets are useful for estimating the influence of uncertainty on the dynamic system. Moreover, generating convex enclosures of reachable sets is fundamental to methods for deterministic global dynamic optimization [101, 81]. Direct analysis of reachable sets is also important in various applications, such as uncertainty evaluation [59], parameter estimation [75], state estimation [69], safety verification [66, 121], and fault detection [82].

Several approaches for describing reachable sets have been established. The Hamilton-Jacobi equation, which is a partial differential equation (PDE), characterizes the reachable set accurately (as summarized by [91]). However, solving such a PDE is still a computationally intensive task with current technology. Other methods require conservative approximations of the original nonlinear system with linearized models, such as a linearization method by [8] with the linearization error rigorously bounded. To reduce over-approximation of the actual reachable set, [12] described a method for splitting state space into smaller regions and computed linear approximations in these partitions. Many types of enclosures have been applied to such linear systems, such as hyper-rectangles [46], zonotopes [7], and ellipsoids [79].

Taylor series methods provide a second way to compute reachable sets for parametric ODEs. They involve computing a validated solution (i.e. a guaranteed enclosure of the true solution) for ODEs by constructing high-order Taylor expansions of the system states with respect to time in discrete time steps, and then bounding the coefficients and remainder terms with interval arithmetic [96]. To

overcome the dependency problem of classic Taylor series methods, in which repeated terms in algebraic representations of functions can lead to significant overestimation in interval arithmetic, Taylor models were introduced by [88]. These methods bound the the Taylor remainder error by propagating an auxiliary model consisting of a Taylor polynomial and an interval remainder bound. [84] used Taylor models to enclose the reachable sets for parametric ODEs. [112] extended these further by replacing interval arithmetic with McCormick relaxations, yielding tighter enclosures in general. However, Taylor series methods may be limited in computational efficiency because of the complexity of constructing and evaluating high-order Taylor expansions. The number of Taylor coefficients involved grows exponentially with the numbers of states and inputs.

A third major category of methods for describing reachable sets involves differential inequalities. Differential inequality-based methods use an auxiliary system of ODEs obtained from the original system to describe the reachable sets. The right-hand side (RHS) functions in the auxiliary relaxation system are modified enclosures of the original ODE RHS function. The solutions of this auxiliary system are guaranteed to be component-wise lower and upper bounds for the reachable set. Such auxiliary systems can be solved via off-the-shelf numerical solvers with adaptive time-stepping, while Taylor series methods require integration procedures with manually configured step-size. Several methods in this different inequality category have been developed in the past decades. A major distinction among these methods is how the original RHS functions are handled. We now briefly review some established methods for constructing the new auxiliary RHS in chronological order. Harrison [59] used interval arithmetic [94] to construct the

auxiliary RHS and computed interval bounds for the original states. A flattening technique was applied in this method (see Section 3.2 below) to reduce the so-called wrapping effect of interval arithmetic. An affine relaxation-based method was introduced by [128], in which the auxiliary RHS functions are constructed via linearizing the classic McCormick relaxation [89]. Although the solutions of such auxiliary systems are rigorous bounds for the original states, their existence and uniqueness are not guaranteed without additional assumptions. Scott and Barton [120] proposed a method for computing component-wise convex and concave relaxations for the final states of parametric ODEs, which are guaranteed to be at least as tight as the Harrison's interval bounds. Scott and Barton's construction of the auxiliary RHS functions involves applying Harrison's flattening technique to generalized McCormick relaxations (GMR) [123]. However, these RHS functions are typically discontinuous, which may hinder methods for solving the ODEs and evaluating subgradients for use in dynamic global optimization. More details about this approach are presented in Section 3.3. Harwood et al. [62] proposed a method that embeds linear programs into the auxiliary RHS functions to improve the enclosures. A special relaxation technique is used to ensure the uniqueness of ODE solutions. Moreover, Harwood et al. considered leveraging an *a priori* enclosure to reduce the conservatism in the relaxation of the original RHS functions. This strategy was further developed by [126, 125, 124]. Shen and Scott made use of known information of the original system, including physical bounds, model redundancy, and path constraints, to further tighten the reachable sets.

In this article, a new differential inequality method is proposed with the goal of generating tight enclosures of reachable sets for parametric ODEs automatically.

This method improves Scott and Barton's method by adding a new square-root term to the auxiliary RHS functions based on kinematic intuition. This modification eliminates the discrete jumps from the auxiliary RHS functions and further tightening them. The solutions of the new auxiliary system are verified to be tighter convex and concave relaxations of the original states. Moreover, under mild assumptions, they are differentiable with respect to parameters. The improved tightness and smoothness of these new convex relaxations are desirable for a number of reasons, such as permitting sensitivity analysis and supplying tighter global bounds for use in global optimization algorithms.

This article is organized as follows. In Section 3.2, we introduce the problem formulation. Necessary mathematical background is summarized in Section 3.3. Our new approach is then presented in Section 3.4, and Section 3.5 establishes useful properties of this approach. In Section 3.6, we introduce a practical numerical method for computing these new relaxations automatically. Finally, numerical examples are presented in Section 3.7 to demonstrate the tighter and smooth convex relaxations generated with this new method.

The following notation conventions are used in this article. The set of positive real numbers is represented as $\mathbb{R}_{>0}$, and the set of non-negative real numbers is represented as $\mathbb{R}_{\geq 0}$. The standard Euclidean norm $\|\cdot\|$ is adopted on $\mathbb{R}^n$. Vectors are denoted with boldface lower-case letters (e.g. $\boldsymbol{x}$). Given vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, inequalities such as $\boldsymbol{x} < \boldsymbol{y}$ or $\boldsymbol{x} \leq \boldsymbol{y}$ are to be interpreted componentwise. Moreover, $\boldsymbol{x}_{(-i)} \in \mathbb{R}^{n-1}$ denotes the vector $\boldsymbol{x} \in \mathbb{R}^n$ except with its ith component excluded. Throughout this article, convexity of a vector-valued function $\boldsymbol{f}$ refers to convexity of all components f_i, and concavity is analogous. An interval in $\mathbb{R}^n$ is a nonempty

subset of $\mathbb{R}^n$ of the form $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}\}$, which is denoted as $[\boldsymbol{a}, \boldsymbol{b}]$. $\mathbb{IR}^n$ denotes the set of all intervals in $\mathbb{R}^n$.

3.2 Problem Formulation

Consider scalars $t_0, t_f' \in \mathbb{R}$ and $t_f \in \mathbb{R} \cup \{+\infty\}$ with $t_0 < t_f' \leq t_f$. Set $I := [t_0, t_f]$ (or $[t_0, +\infty)$ if $t_f = +\infty$). Set $I' := [t_0, t_f'] \subset I$. Let $P \subset \mathbb{R}^{n_p}$ be an interval, and $D \subset \mathbb{R}^{n_x}$ be open. Given a continuous mapping $\boldsymbol{x}_0 : P \to D$ and a Lipschitz continuous function $\boldsymbol{f} : I \times P \times D \to \mathbb{R}^{n_x}$, the remainder of this article considers an initial-value problem

$$\dot{\boldsymbol{x}}(t, \boldsymbol{p}) = \boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{x}(t, \boldsymbol{p})), \qquad \boldsymbol{x}(t_0, \boldsymbol{p}) = \boldsymbol{x}_0(\boldsymbol{p}). \tag{3.1}$$

Assume that there exists a solution of (3.1) on $I \times P$. Moreover, suppose that $k^{\boldsymbol{x}} \in \mathbb{R}_{>0}$ is a Lipschitz constant of $\boldsymbol{f}(t, \boldsymbol{p}, \cdot)$ over D for all $(t, \boldsymbol{p}) \in I \times P$. Thus, (3.1) has a unique solution $\boldsymbol{x}$ by the Picard-Lindelöf Theorem summarized in [60, Theorem 1.1, Chapter II].

State relaxations, defined below, are convex and concave relaxations of the state variable $\boldsymbol{x}$, respectively, which provide valid enclosures for the reachable set of (3.1). Besides that, they are also required in generating convex relaxed problems for nonconvex dynamic optimization problems with ODEs embedded [120]. Minimizing these relaxed convex problems locally generates guaranteed lower bounds of the original nonconvex objective function, which are desired in deterministic global dynamic optimization [131].

Definition 3.1. *Suppose that $Z \subset \mathbb{R}^n$ is convex and $\boldsymbol{h} : Z \to \mathbb{R}^m$.*

1. $\boldsymbol{h}^{cv} : Z \to \mathbb{R}^m$ *is a* convex relaxation *of* $\boldsymbol{h}$ *on* Z *if* $\boldsymbol{h}^{cv}(\boldsymbol{z}) \leq \boldsymbol{h}(\boldsymbol{z})$ *for all* $\boldsymbol{z} \in Z$ *and* $\boldsymbol{h}^{cv}$ *is convex on* Z.

2. $\boldsymbol{h}^{cc} : Z \to \mathbb{R}^m$ *is a* concave relaxation *of* $\boldsymbol{h}$ *on* Z *if* $\boldsymbol{h}^{cc}(\boldsymbol{z}) \geq \boldsymbol{h}(\boldsymbol{z})$ *for all* $\boldsymbol{z} \in Z$ *and* $\boldsymbol{h}^{cc}$ *is concave on* Z.

Definition 3.2. *Functions* $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc} : I' \times P \to \mathbb{R}^{n_x}$ *are* state relaxations *for* (3.1) *on* $I' \times P$*, if, for every* $t \in I'$*,*

1. $\boldsymbol{x}^{cv}(t, \cdot)$ *is convex on* P*,*

2. $\boldsymbol{x}^{cc}(t, \cdot)$ *is concave on* P*, and*

3. $\boldsymbol{x}^{cv}(t, \boldsymbol{p}) \leq \boldsymbol{x}(t, \boldsymbol{p}) \leq \boldsymbol{x}^{cc}(t, \boldsymbol{p})$*, for all* $\boldsymbol{p} \in P$.

The objective of this work is to generate improved state relaxations for the ODE (3.1) on $I' \times P$. We will achieve this by adapting and tightening a state-of-the-art method developed by [120]. Their ODE relaxation method will be referred as the *Scott-Barton method* hereafter, and their corresponding state relaxations will be called *Scott-Barton relaxations*. These relaxations are described by an auxiliary system of ODEs, whose right-hand side (RHS) functions are constructed with generalized McCormick relaxations (GMR) [123]. Intuitively, the usage of GMR indicates that such RHS functions depend on interval bounds of $\boldsymbol{x}$, motivating the following definition.

Definition 3.3. *Functions* $\boldsymbol{x}^L, \boldsymbol{x}^U : I \to \mathbb{R}^{n_x}$ *are* state bounds *for* (3.1) *over* P *if* $\boldsymbol{x}^L(t) \leq \boldsymbol{x}(t, \boldsymbol{p}) \leq \boldsymbol{x}^U(t)$ *for all* $(t, \boldsymbol{p}) \in I \times P$*. For all* $t \in I$*, the interval* $[\boldsymbol{x}^L(t), \boldsymbol{x}^U(t)]$ *is denoted by* $X^B(t) \in \mathbb{IR}^{n_x}$.

Similar to [120], we assume that state bounds for (3.1) are always available. We also assume similar requirements on these state bounds. Define the following variant of the ODE (3.1):

$$\begin{aligned} \dot{\boldsymbol{\xi}}(t) &= \boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{\xi}(t)), \quad t \in I, \\ \boldsymbol{\xi}(\tau) &= \boldsymbol{\xi}_0, \end{aligned} \tag{3.2}$$

where $\tau \in I$ and $\boldsymbol{\xi}_0 \in X^B(\tau)$.

Assumption 3.1. *Assume the following conditions hold:*

1. *State bounds of* (3.1) *over* P, $\boldsymbol{x}^L$ *and* $\boldsymbol{x}^U$, *are available and differentiable on* I.
2. $\dot{\boldsymbol{x}}^L(\cdot)$ *and* $\dot{\boldsymbol{x}}^U(\cdot)$ *are measurable on* I.
3. *There exists an integrable function* $\tilde{m} : I \to \mathbb{R}$ *such that* $\|\dot{\boldsymbol{x}}^L(t)\| \leq \tilde{m}(t)$ *and* $\|\dot{\boldsymbol{x}}^U(t)\| \leq \tilde{m}(t)$ *for each* $t \in I$.
4. $X^B \equiv [\boldsymbol{x}^L, \boldsymbol{x}^U]$ *is a state bound of* (3.2) *over* P *for all* $t \in I$ *and* $\boldsymbol{\xi}_0 \in X^B(\tau)$.

An established approach by [59] was used in the Scott-Barton method to compute state bounds for (3.1). This approach solves an auxiliary system of ODEs with RHS function constructed with a variant of natural interval extension (NIE) [94]. This approach satisfies Assumption 3.1, and will be adopted in the implementation of this work as well. To summarize the relevant details of Harrison state bounds, we adapt the description from [120, 31], which involves a flattening operation over state bounds.

Definition 3.4. *Define flattening operators* $B_i^L, B_i^U : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ *such that, for each* $i \in \{1, \ldots, n_x\}$ *and* $\boldsymbol{\phi}, \boldsymbol{\psi} \in \mathbb{R}^{n_x}$,

1. $B_i^L(\boldsymbol{\phi}, \boldsymbol{\psi}) = (\boldsymbol{\phi}, \boldsymbol{\psi}')$, *where* $\psi_i' = \phi_i$ *and* $\boldsymbol{\psi}'_{(-i)} = \boldsymbol{\psi}_{(-i)}$.

2. $B_i^U(\boldsymbol{\phi}, \boldsymbol{\psi}) = (\boldsymbol{\phi}', \boldsymbol{\psi})$, *where* $\phi_i' = \psi_i$ *and* $\boldsymbol{\phi}'_{(-i)} = \boldsymbol{\phi}_{(-i)}$.

Define $\boldsymbol{f}^L, \boldsymbol{f}^U : I \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ so that, for all $t \in I$, $\boldsymbol{p} \in P$, and $\boldsymbol{z}, \boldsymbol{\phi}, \boldsymbol{\psi} \in \mathbb{R}^{n_x}$ such that $\boldsymbol{\phi} \leq \boldsymbol{z} \leq \boldsymbol{\psi}$,

$$\boldsymbol{f}^L(t, \boldsymbol{\phi}, \boldsymbol{\psi}) \leq \boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{z}) \leq \boldsymbol{f}^U(t, \boldsymbol{\phi}, \boldsymbol{\psi}).$$

Define $\bar{\boldsymbol{x}}_0^L, \bar{\boldsymbol{x}}_0^U \in \mathbb{R}^{n_x}$ so that, for all $\boldsymbol{p} \in P$,

$$\bar{\boldsymbol{x}}_0^L \leq \boldsymbol{x}_0(\boldsymbol{p}) \leq \bar{\boldsymbol{x}}_0^U.$$

The functions $\boldsymbol{f}^L, \boldsymbol{f}^U$ and $\bar{\boldsymbol{x}}_0^L, \bar{\boldsymbol{x}}_0^U$ described above may be constructed via applying NIE to $\boldsymbol{f}$ and $\boldsymbol{x}_0$, respectively.

Definition 3.5. *Harrison's state bounds for* (3.1)*, denoted as* $\bar{X}^B \equiv [\bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U]$*, are computed as the solutions of the following auxiliary system of ODEs: for each* $i \in \{1, \ldots, n_x\}$*,*

$$\begin{aligned} \dot{\bar{x}}_i^L &= f_i^L(t, B_i^L(\bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U)), \quad \bar{x}_i^L(t_0) = \bar{x}_{0,i}^L, \\ \dot{\bar{x}}_i^U &= f_i^U(t, B_i^U(\bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U)), \quad \bar{x}_i^U(t_0) = \bar{x}_{0,i}^U. \end{aligned} \tag{3.3}$$

Define $\bar{\boldsymbol{f}}^L, \bar{\boldsymbol{f}}^U : I \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ such that, for $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} \bar{f}_i^L(t, \boldsymbol{\phi}, \boldsymbol{\psi}) &:= f_i^L(t, B_i^L(\boldsymbol{\phi}, \boldsymbol{\psi})), \\ \bar{f}_i^U(t, \boldsymbol{\phi}, \boldsymbol{\psi}) &:= f_i^U(t, B_i^U(\boldsymbol{\phi}, \boldsymbol{\psi})). \end{aligned}$$

Then, (3.3) is equivalent to

$$\begin{aligned} \dot{\bar{x}}_i^L &= \bar{f}_i^L(t, \bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U), \quad \bar{x}_i^L(t_0) = \bar{x}_{0,i}^L, \\ \dot{\bar{x}}_i^U &= \bar{f}_i^U(t, \bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U), \quad \bar{x}_i^U(t_0) = \bar{x}_{0,i}^U. \end{aligned} \tag{3.4}$$

Analogous to Harrison's bounding method, the Scott-Barton method applies the flattening operators to GMR and constructs auxiliary RHS functions. This technique reduces the overestimation in state relaxations [113]. Desired bounding and convexity properties of the generated state relaxations are ensured by verifying that the flattened relaxations describe *bound-preserving dynamics* and *convexity-preserving dynamics*, under the following definitions adapted from [120].

Definition 3.6. *Functions* $\boldsymbol{u}, \boldsymbol{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ *describe* bound-preserving dynamics *for* (3.1) *if, for any* $\boldsymbol{p} \in P$*, each* $i \in \{1, \ldots, n_x\}$*, a.e.* $t \in I$ *(in the Lebesgue sense), and any* $\boldsymbol{z}, \boldsymbol{\phi}, \boldsymbol{\psi} \in X^B(t)$ *such that* $\boldsymbol{\phi} \leq \boldsymbol{z} \leq \boldsymbol{\psi}$*,* $\boldsymbol{u}$ *and* $\boldsymbol{o}$ *satisfy the following conditions:*

1. *If* $z_i = \phi_i$*, then* $u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \leq f_i(t, \boldsymbol{p}, \boldsymbol{z})$.
2. *If* $z_i = \psi_i$*, then* $o_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \geq f_i(t, \boldsymbol{p}, \boldsymbol{z})$.

Definition 3.7. *Functions* $\boldsymbol{u}, \boldsymbol{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ *describe* convexity-preserving dynamics *for* (3.1) *if, for any* $(\lambda, \boldsymbol{p}_1, \boldsymbol{p}_2) \in (0,1) \times P \times P$*,* $\bar{\boldsymbol{p}} := \lambda \boldsymbol{p}_1 + (1-\lambda)\boldsymbol{p}_2$*, each* $i \in \{1, \ldots, n_x\}$*, a.e.* $t \in I$*, and any* $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \bar{\boldsymbol{\phi}}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \bar{\boldsymbol{\psi}} \in X^B(t)$ *such that the following three conditions all hold:*

1. $\bar{\boldsymbol{\phi}} \leq \lambda \boldsymbol{\phi}_1 + (1-\lambda)\boldsymbol{\phi}_2$*,*
2. $\bar{\boldsymbol{\psi}} \geq \lambda \boldsymbol{\psi}_1 + (1-\lambda)\boldsymbol{\psi}_2$*, and*

3. $\boldsymbol{\phi}_1 \leq \boldsymbol{\psi}_1$, $\boldsymbol{\phi}_2 \leq \boldsymbol{\psi}_2$, $\bar{\boldsymbol{\phi}} \leq \bar{\boldsymbol{\psi}}$,

$\boldsymbol{u}$ and $\boldsymbol{o}$ satisfy the following conditions:

1. *If $\bar{\phi}_i = \lambda\phi_{1,i} + (1-\lambda)\phi_{2,i}$, then*

$$u_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \leq \lambda u_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) + (1-\lambda) u_i(t, \boldsymbol{p}_2, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2).$$

2. *If $\bar{\psi}_i = \lambda\psi_{1,i} + (1-\lambda)\psi_{2,i}$, then*

$$o_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \geq \lambda o_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) + (1-\lambda) o_i(t, \boldsymbol{p}_2, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2).$$

Assumption 3.2. *There exist particular functions $\boldsymbol{u}, \boldsymbol{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ satisfying the following conditions:*

1. *$\boldsymbol{u}, \boldsymbol{o}$ are continuous,*

2. *There exists a positive constant $k^r \in \mathbb{R}_{>0}$ such that, for all $t \in I$, $\boldsymbol{p} \in P$, and $\boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger \in \mathbb{R}^{n_x}$,*

$$\begin{aligned} &\left\| \boldsymbol{u}(t, \boldsymbol{p}, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) - \boldsymbol{u}(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) \right\| + \left\| \boldsymbol{o}(t, \boldsymbol{p}, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) - \boldsymbol{o}(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) \right\| \\ &\quad \leq k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right), \end{aligned} \tag{3.5}$$

3. *$\boldsymbol{u}, \boldsymbol{o}$ describe bound-preserving dynamics for* (3.1)*,*

4. *$\boldsymbol{u}, \boldsymbol{o}$ describe convexity-preserving dynamics for* (3.1)*.*

In addition to flattened GMR, functions $\boldsymbol{u}, \boldsymbol{o}$ satisfying Assumption 3.2 may be generated with flattened differentiable McCormick relaxations (DMR) [72, 73].

Moreover, [131] developed an optimization-based approach to construct functions $\boldsymbol{u}, \boldsymbol{o}$ satisfying Assumption 3.2 with the following formulation. For each $i \in \{1, \ldots, n_x\}$,

$$u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) = \min_{\boldsymbol{z} \in [\boldsymbol{\phi}, \boldsymbol{\psi}], z_i = \phi_i} f^{cv}(t, \boldsymbol{p}, \boldsymbol{z}),$$
$$o_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) = \max_{\boldsymbol{z} \in [\boldsymbol{\phi}, \boldsymbol{\psi}], z_i = \psi_i} f^{cc}(t, \boldsymbol{p}, \boldsymbol{z}),$$

where $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc} : I \times P \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ are modified convex and concave relaxations of $\boldsymbol{f}$ [131]. Available methods for generating $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ from $\boldsymbol{f}$ include αBB relaxations [3] and McCormick-based relaxations [92].

3.3 Background

Given the problem formulation in the previous section, and under Assumptions 3.1 and 3.2, we now briefly introduce the state relaxation method developed by [120]. Their method lays the foundation of our new approach, and motivates its structure.

Consider the following auxiliary system of ODEs: for each $i \in \{1, \ldots, n_x\}$,

$$\dot{x}_i^{cv}(t,\boldsymbol{p}) = \begin{cases} u_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) & \text{if } x_i^{cv}(t,\boldsymbol{p}) > x_i^L(t), \\ \max\left\{\dot{x}_i^L(t), u_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p}))\right\} & \text{if } x_i^{cv}(t,\boldsymbol{p}) = x_i^L(t), \end{cases} \tag{3.6a}$$

$$x_i^{cv}(t_0,\boldsymbol{p}) = x_{0,i}^{cv}(\boldsymbol{p}),$$

$$\dot{x}_i^{cc}(t,\boldsymbol{p}) = \begin{cases} o_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) & \text{if } x_i^{cc}(t,\boldsymbol{p}) < x_i^U(t), \\ \min\left\{\dot{x}_i^U(t), o_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p}))\right\} & \text{if } x_i^{cc}(t,\boldsymbol{p}) = x_i^U(t), \end{cases} \tag{3.6b}$$

$$x_i^{cc}(t_0,\boldsymbol{p}) = x_{0,i}^{cc}(\boldsymbol{p}),$$

where $\boldsymbol{x}_0^{cv}$, $\boldsymbol{x}_0^{cc} : P \to \mathbb{R}^{n_x}$ are respectively convex and concave relaxations of $\boldsymbol{x}_0$ on P such that $\boldsymbol{x}^L(t_0) \leq \boldsymbol{x}_0^{cv}(\boldsymbol{p})$ and $\boldsymbol{x}^U(t_0) \geq \boldsymbol{x}_0^{cc}(\boldsymbol{p})$. These inequality requirements may be enforced by setting $\boldsymbol{x}_0^{cv}(\boldsymbol{p}) \leftarrow \max\{\boldsymbol{x}^L(t_0), \boldsymbol{x}_0^{cv}(\boldsymbol{p})\}$ and $\boldsymbol{x}_0^{cc}(\boldsymbol{p}) \leftarrow \min\{\boldsymbol{x}^U(t_0), \boldsymbol{x}_0^{cc}(\boldsymbol{p})\}$, where max and min are computed componentwise.

Scott and Barton [120] showed that valid state relaxations of (3.1) are given by the unique Carathéodory solutions of (3.6). Moreover, [113] verified the following result. If we construct $\boldsymbol{u}, \boldsymbol{o}$ with flattened GMR, then the Scott-Barton relaxations will have *second-order pointwise convergence* to $\boldsymbol{x}$ in the sense of [24]. This convergence result is critical for using state relaxations in deterministic global optimization without invoking the "cluster problem" [48, 150] in which a branch-and-bound algorithm must branch many times before terminating.

Although the solutions of (3.6) provide state relaxations for (3.1), the if-statements in the RHS of the ODE system (3.6) will typically create discontinuity in the RHS.

To numerically solve (3.6), [120] proposed to use the event detection feature of CVODES [44] to handle these discontinuities, but this approach increases the difficulty of implementation and limits the use of other off-the-shelf ODE solvers. Without event detection, the numerical error resulting from the integration process will likely be worse than when solving similar ODEs with continuous RHS. Another limitation of the Scott-Barton method is the difficulty of evaluating gradient or subgradient information for state relaxations, again due to those discrete jumps. To avoid any possibility of discrete RHS jumps and to provide improved relaxations, a new relaxation system is proposed in the next section.

Finally, we introduce a definition adapted from [115, Section 3.1], which will be used for the automatic computation of the new state relaxations.

Definition 3.8. *Let $\boldsymbol{h} : D \to \mathbb{R}^n$ be a function. If for every $\boldsymbol{y} \in \mathbb{R}^n$ the limit*

$$\boldsymbol{h}'(\boldsymbol{z}_0; \boldsymbol{y}) = \lim_{\lambda \downarrow 0} \frac{1}{\lambda}(\boldsymbol{h}(\boldsymbol{z}_0 + \lambda \boldsymbol{y}) - \boldsymbol{h}(\boldsymbol{z}))$$

exists, then $\boldsymbol{h}$ is directionally differentiable *at $\boldsymbol{z}_0$ and the function $\boldsymbol{h}'(\boldsymbol{z}_0; \cdot)$ is the directional derivative of $\boldsymbol{h}$ at $\boldsymbol{z}_0$. Moreover, if $\boldsymbol{h}$ is directionally differentiable at $\boldsymbol{z}_0$ and also Lipschitz continuous near $\boldsymbol{z}_0$, then $\boldsymbol{h}$ is* B-differentiable *at $\boldsymbol{z}_0$ and the function $\boldsymbol{h}'(\boldsymbol{z}_0; \cdot)$ is the B-derivative of $\boldsymbol{h}$ at $\boldsymbol{z}_0$.*

3.4 New State Relaxation Formulation

This section presents a new formulation for generating state relaxations without discrete jumps in the auxiliary RHS function. Useful properties of this formulation are then established in Section 3.5. This formulation employs a *safe-landing*

mechanism to avoid the if-statements in (3.6), based on a kinematic intuition that is explained in the end of this section. This safe-landing mechanism requires positive constants $\underline{\boldsymbol{k}}, \overline{\boldsymbol{k}} \in \mathbb{R}^{n_x}_{>0}$ satisfying the following definition. Section 3.5.5 will present automatable approaches for computing these constants.

Definition 3.9. *Suppose that Assumption 3.1 holds. Let $\boldsymbol{\xi}(t)$ be a solution of* (3.2). *$\underline{\boldsymbol{k}}, \overline{\boldsymbol{k}}$ are* safe-landing constants *for* (3.1) *over X^B if the following holds. For any $\tau \in I'$ and $i \in \{1, \ldots, n_x\}$,*

1. *If $\dot{\xi}_i(\tau) < \dot{x}_i^L(\tau)$, then for each $t \in [\tau, \tau + \frac{\dot{x}_i^L(\tau) - \dot{\xi}_i(\tau)}{\underline{k}_i}]$,*

$$\left(\dot{\xi}_i(t) - \dot{x}_i^L(t)\right) - \left(\dot{\xi}_i(\tau) - \dot{x}_i^L(\tau)\right) \leq \underline{k}_i(t - \tau).$$

2. *If $\dot{\xi}_i(\tau) > \dot{x}_i^U(\tau)$, then for each $t \in [\tau, \tau + \frac{\dot{\xi}_i(\tau) - \dot{x}_i^U(\tau)}{\overline{k}_i}]$,*

$$\left(\dot{x}_i^U(t) - \dot{\xi}_i(t)\right) - \left(\dot{x}_i^U(\tau) - \dot{\xi}_i(\tau)\right) \leq \overline{k}_i(t - \tau).$$

Remark 3.1. *To apply Definition 3.9, it is necessary to ensure that the time intervals $[\tau, \tau + \frac{\dot{x}_i^L(\tau) - \dot{\xi}_i(\tau)}{\underline{k}_i}]$ and $[\tau, \tau + \frac{\dot{\xi}_i(\tau) - \dot{x}_i^U(\tau)}{\overline{k}_i}]$ are contained in I for any $\tau \in I'$, $\boldsymbol{\xi}_0 \in X^B(\tau)$, and $i \in \{1, \ldots, n_x\}$. If $I = [t_0, +\infty)$, then this requirement is immediately satisfied. Otherwise, let $\dot{\xi}_i^L$ and $\dot{\xi}_i^U$ be lower and upper bounds of $\dot{\xi}_i(t'_f)$, respectively, for all $\boldsymbol{\xi}_0 \in X^B(\tau)$. These can be calculated from the NIE of $f_i(t'_f, \cdot, \cdot)$ on $P \times X^B(t'_f)$. Then, we need only to choose t_f large enough so that $t_f \geq \max\{t'_f, t'_f + \frac{\dot{x}_i^L(t'_f) - \dot{\xi}_i^L}{\underline{k}_i}, t'_f + \frac{\dot{\xi}_i^U - \dot{x}_i^U(t'_f)}{\overline{k}_i}\}$. This claim is supported by the following lemma.*

Lemma 3.1. *Suppose that $t_f := \max\{t'_f, t'_f + \frac{-\dot{y}(t'_f)}{k}\}$. Consider a positive constant $k \in \mathbb{R}_{>0}$ and a non-negative differentiable function $y : [t_0, t_f] \to \mathbb{R}_{\geq 0}$. For any $\tau \in I' =$*

$[t_0, t'_f]$ *such that* $\dot{y}(\tau) < 0$, *if*

$$\dot{y}(t) - \dot{y}(\tau) \leq k(t - \tau), \quad \forall t \in [\tau, \tau + \frac{-\dot{y}(\tau)}{k}], \tag{3.7}$$

then

$$\tau + \frac{-\dot{y}(\tau)}{k} \leq t_f.$$

Proof. Choose an arbitrary $\hat{\tau} \in I'$ such that $\dot{y}(\hat{\tau}) < 0$. To achieve a contradiction, suppose that

$$t_f < \hat{\tau} + \frac{-\dot{y}(\hat{\tau})}{k}. \tag{3.8}$$

Reformulating (3.8) yields:

$$-\dot{y}(\hat{\tau}) > k(t_f - \hat{\tau}). \tag{3.9}$$

Contradictions can be found in the following exhaustive cases:

1. If $\dot{y}(t'_f) \geq 0$, then $t_f = \max\{t'_f, t'_f + \frac{-\dot{y}(t'_f)}{k}\}$ implies that $t_f = t'_f$. Moreover, in this case, (3.9) implies

$$\dot{y}(t'_f) - \dot{y}(\hat{\tau}) > k(t'_f - \hat{\tau}). \tag{3.10}$$

   Now, (3.8) ensures that $t'_f \in [\hat{\tau}, \hat{\tau} + \frac{-\dot{y}(\hat{\tau})}{k}]$. Therefore, (3.10) contradicts (3.7).

2. If $\dot{y}(t'_f) < 0$, then $t_f = t'_f + \frac{-\dot{y}(t'_f)}{k}$. (3.9) becomes

$$-\dot{y}(\hat{\tau}) > k(t'_f + \frac{-\dot{y}(t'_f)}{k} - \hat{\tau}).$$

Rearranging the above inequality yields (3.10), which contradicts (3.7).

□

The following assumption is imposed in the remainder of this article.

Assumption 3.3. *Suppose Assumption 3.1 holds, and that* $\underline{\boldsymbol{k}}, \overline{\boldsymbol{k}} \in \mathbb{R}^{n_x}_{>0}$ *are safe-landing constants for* (3.1) *over* X^B.

Definition 3.10. *Define a scalar-valued mapping* $\sigma : \mathbb{R} \to \mathbb{R}_{\geq 0}$ *for which*

$$\sigma(\theta) \equiv \sqrt{\max\{\theta,\ 0\}}. \tag{3.11}$$

Definition 3.11. *Under Assumption 3.3, define functions* $\boldsymbol{\alpha}, \boldsymbol{\beta} : I \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ *so that, for each* $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} \alpha_i(t, \boldsymbol{\theta}) &\equiv \dot{x}_i^L(t) - \sigma\left(2\underline{k}_i(\theta_i - x_i^L(t))\right), \\ \beta_i(t, \boldsymbol{\theta}) &\equiv \dot{x}_i^U(t) + \sigma\left(2\overline{k}_i(x_i^U(t) - \theta_i)\right). \end{aligned} \tag{3.12}$$

Definition 3.12. *Under Assumptions 3.2 and 3.3, define functions* $\boldsymbol{v}, \boldsymbol{w} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ *so that, for each* $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} v_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) &\equiv \max\left\{u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}),\ \alpha_i(t, \boldsymbol{\phi})\right\}, \\ w_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) &\equiv \min\left\{o_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}),\ \beta_i(t, \boldsymbol{\psi})\right\}. \end{aligned} \tag{3.13}$$

Substituting (3.11) and (3.12) into (3.13) yields: for each $i \in \{1, \ldots, n_x\}$,

$$v_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}) = \max\left\{u_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}),\ \dot{x}_i^L(t) - \sqrt{2\underline{k}_i \max\{\phi_i - x_i^L(t), 0\}}\right\}, \quad (3.14a)$$

$$w_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}) = \min\left\{o_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}),\ \dot{x}_i^U(t) + \sqrt{2\overline{k}_i \max\{x_i^U(t) - \psi_i, 0\}}\right\}. \quad (3.14b)$$

The following definition constructs our new state relaxations. Their validity will be established in Section 3.5.

Definition 3.13. *Under Assumptions 3.2 and 3.3, and with $\boldsymbol{v}$ and $\boldsymbol{w}$ as in Definition 3.12, define an auxiliary initial-value problem in parametric ODEs:*

$$\dot{\boldsymbol{x}}^{cv}(t,\boldsymbol{p}) = \boldsymbol{v}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})), \quad \boldsymbol{x}^{cv}(t_0,\boldsymbol{p}) = \boldsymbol{x}_0^{cv}(\boldsymbol{p}), \quad (3.15a)$$

$$\dot{\boldsymbol{x}}^{cc}(t,\boldsymbol{p}) = \boldsymbol{w}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})), \quad \boldsymbol{x}^{cc}(t_0,\boldsymbol{p}) = \boldsymbol{x}_0^{cc}(\boldsymbol{p}). \quad (3.15b)$$

Remark 3.2. *It will be shown in Section 3.5.3 that $\boldsymbol{x}^{cv}(t,\boldsymbol{p}) \geq \boldsymbol{x}^L(t)$ and $\boldsymbol{x}^{cc}(t,\boldsymbol{p}) \leq \boldsymbol{x}^U(t)$ for all $(t,\boldsymbol{p}) \in I' \times P$. Thus, the arguments of σ in* (3.12) *can never be negative. So $(\boldsymbol{\alpha},\boldsymbol{\beta})$ can be simplified by replacing the function σ with the square root, as will be shown in* (3.24).

Next, we explain the kinematic intuition behind our new formulation in (3.14a) and (3.15a). For any $i \in \{1, \ldots, n_x\}$ and $(t,\boldsymbol{p}) \in I' \times P$, we image a drone, which flies at the altitude $x_i^{cv}(t,\boldsymbol{p}) - x_i^L(t)$, is trying to land safely on the ground with a constant acceleration rate. This is a metaphor for describing that the state relaxation x_i^{cv} is moving towards the state bound x_i^L and trying to achieve the same first-order time derivative. Unlike the Scott-Barton method in (3.6a) which changes the drone's trajectory suddenly using if-statements, our new formulation makes this

landing process happen smoothly. The motion of this imaginary drone is designed to follow this kinematic equation:

$$v_f^2 = v_i^2 + 2ad,$$

where $v_i \equiv \dot{x}_i^{cv}(t, \boldsymbol{p}) - \dot{x}_i^L(t) < 0$ stands for initial vertical velocity, v_f stands for final vertical velocity, $d \equiv x_i^{cv}(t, \boldsymbol{p}) - x_i^L(t)$ stands for displacement, and a stands for acceleration rate. We expect that when the drone's vertical velocity v_f is 0, its altitude is nonnegative, which implies that the drone will not crash to the ground. This requires the drone to pull up quickly with a sufficiently large acceleration rate, which corresponds to the requirement of safe-landing constants in Definition 3.9.

3.5 Theoretical Development

Under Assumptions 3.2 and 3.3, this section establishes the following features of the ODE system (3.15).

- The ODE system (3.15) has a unique solution on $I' \times P$.
- The solutions of (3.15) are valid state relaxations for (3.1) on $I' \times P$.
- The new state relaxations generated with (3.15) are at least as tight as Scott-Barton relaxations when the same $\boldsymbol{u}, \boldsymbol{o}$ are used, and therefore have analogous second-order pointwise convergence.
- Under additional mild assumptions, these state relaxations are differentiable with respect to parameters.

We also discuss how to compute the safe-landing constants required to formulate and solve the ODE system (3.15).

3.5.1 Existence of a solution

Theorem 3.1. *Under Assumptions 3.2 and 3.3, there exist a solution of* (3.15) *on* $I' \times P$.

Proof. Under Assumption 3.3, it is trivially verified that $\boldsymbol{v}$ and $\boldsymbol{w}$ in (3.14) satisfy the Carathéodory conditions [53, page 3]. Theorem 1 in [53, Chapter 1] then ensures that there exists a solution of (3.15) on $I' \times P$. □

3.5.2 Uniqueness of a solution

Classic uniqueness results for ODEs, e.g., the Picard-Lindelöf Theorem, typically require the ODE right-hand side (RHS) functions to be Lipschitz continuous. However, $\boldsymbol{v}, \boldsymbol{w}$ in (3.15) do not satisfy this condition due to the square root functions. To address this, a uniqueness theorem by [148] is used here instead. This result instead requires the RHS functions to satisfy a one-sided Lipschitz-like condition.

Theorem 3.2. *Under Assumptions 3.2 and 3.3, the solution of* (3.15) *on* $I' \times P$ *is unique.*

Proof. Theorem 3.1 shows that such a solution exits. Consider the Lipschitz constant $k^r \in \mathbb{R}_{>0}$ from Assumption 3.2. According to the uniqueness result in [148, p. 88], it suffices to show that, for all $t \in I'$, $\boldsymbol{p} \in P$, $\boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger \in \mathbb{R}^{n_x}$, and $i \in \{1, \ldots, n_x\}$, the following two conditions hold.

- If $\phi_i^\dagger \geq \phi_i^\ddagger$, then

$$v_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) - v_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) \leq k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right). \quad (3.16)$$

- If $\psi_i^\dagger \geq \psi_i^\ddagger$, then

$$w_i(t,\boldsymbol{p},\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger) - w_i(t,\boldsymbol{p},\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger) \leq k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right). \quad (3.17)$$

Now, suppose that $\phi_i^\dagger \geq \phi_i^\ddagger$. (3.16) is equivalent to

$$\begin{aligned} & k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right) \\ & \geq \max \left\{ u_i(t,\boldsymbol{p},\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger),\ \alpha_i(t,\boldsymbol{\phi}^\dagger) \right\} - \max \left\{ u_i(t,\boldsymbol{p},\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger),\ \alpha_i(t,\boldsymbol{\phi}^\ddagger) \right\}. \end{aligned} \quad (3.18)$$

Since the right-hand side of (3.18) has two bivariate "max" operations, we consider the corresponding four cases separately.

1. Suppose $u_i(t,\boldsymbol{p},\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger) \geq \alpha_i(t,\boldsymbol{\phi}^\dagger)$ and $u_i(t,\boldsymbol{p},\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger) \geq \alpha_i(t,\boldsymbol{\phi}^\ddagger)$.

   In this case, demonstrating (3.18) is equivalent to demonstrating that

$$k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right) \geq u_i(t,\boldsymbol{p},\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger) - u_i(t,\boldsymbol{p},\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger), \quad (3.19)$$

   which always holds according to (3.5).

2. Suppose $u_i(t,\boldsymbol{p},\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger) \geq \alpha_i(t,\boldsymbol{\phi}^\dagger)$ and $u_i(t,\boldsymbol{p},\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger) < \alpha_i(t,\boldsymbol{\phi}^\ddagger)$.

   In this case, it suffices to show that

$$\begin{aligned} & k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right) \\ & \geq u_i(t,\boldsymbol{p},\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger) - \left(\dot{x}_i^L(t) - \sqrt{2\underline{k}_i \max\{\phi_i^\ddagger - x_i^L(t), 0\}} \right). \end{aligned} \quad (3.20)$$

According to Assumption (3.5),

$$k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right) \geq u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) - u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger).$$

Since in this case

$$0 > u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) - \left(\dot{x}_i^L(t) - \sqrt{2\underline{k}_i \max\{\phi_i^\ddagger - x_i^L(t), 0\}} \right),$$

adding the above two inequalities yields

$$\begin{aligned} & k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right) \\ & \quad > u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) - u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) + u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) \\ & \qquad - \left(\dot{x}_i^L(t) - \sqrt{2\underline{k}_i \max\{\phi_i^\ddagger - x_i^L(t), 0\}} \right) \\ & \quad = u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) - \left(\dot{x}_i^L(t) - \sqrt{2\underline{k}_i \max\{\phi_i^\ddagger - x_i^L(t), 0\}} \right). \end{aligned}$$

(3.20) follows.

3. Suppose $u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) < \alpha_i(t, \boldsymbol{\phi}^\dagger)$ and $u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) \geq \alpha_i(t, \boldsymbol{\phi}^\ddagger)$.

   The argument for the previous case applies here, after interchanging $\boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger$ and $\boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger$.

4. Suppose $u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) < \alpha_i(t, \boldsymbol{\phi}^\dagger)$ and $u_i(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) < \alpha_i(t, \boldsymbol{\phi}^\ddagger)$.

In this case, it suffices to show that

$$\begin{aligned} & k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right) \\ & \quad \geq \left(\dot{x}_i^L(t) - \sqrt{2\underline{k}_i \max\{\phi_i^\dagger - x_i^L(t), 0\}} - \left(\dot{x}_i^L(t) - \sqrt{2\underline{k}_i \max\{\phi_i^\ddagger - x_i^L(t), 0\}} \right) \right) \\ & \quad = -\sqrt{2\underline{k}_i} \left(\sqrt{\max\{\phi_i^\dagger - x_i^L(t), 0\}} - \sqrt{\max\{\phi_i^\ddagger - x_i^L(t), 0\}} \right) \end{aligned} \tag{3.21}$$

We now divide this case into several further cases depending on the "max" terms in (3.21), given that $\phi_i^\dagger \geq \phi_i^\ddagger$.

(a) Suppose $\phi_i^\dagger \geq \phi_i^\ddagger > x_i^L(t)$. Then, (3.21) becomes

$$\begin{aligned} & k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right) \\ & \quad \geq -\sqrt{2\underline{k}_i} \left(\sqrt{\phi_i^\dagger - x_i^L(t)} - \sqrt{\phi_i^\ddagger - x_i^L(t)} \right). \end{aligned} \tag{3.22}$$

Because $\phi_i^\dagger \geq \phi_i^\ddagger$, the right-hand side of (3.22) cannot be positive and so (3.22) holds.

(b) Suppose $\phi_i^\dagger \geq x_i^L(t) \geq \phi_i^\ddagger$. Then, (3.21) becomes

$$\begin{aligned} k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right) &\geq -\sqrt{2\underline{k}_i} \left(\sqrt{\phi_i^\dagger - x_i^L(t)} - 0 \right) \\ &= -\sqrt{2\underline{k}_i} \sqrt{\phi_i^\dagger - x_i^L(t)}. \end{aligned} \tag{3.23}$$

Since the right-hand side of (3.23) cannot be positive, (3.23) always holds.

(c) Suppose $x_i^L(t) > \phi_i^\dagger \geq \phi_i^\ddagger$. Then, (3.21) becomes

$$k^r \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\| + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\| \right) \geq 0,$$

which is always true.

Combining Cases 1-4, the inequality (3.16) is established. A similar argument establishes (3.17). □

The previous two subsections showed that, under Assumptions 3.1, 3.2, and 3.3, (3.15) has a unique solution on $I' \times P$. This result is used implicitly in the remaining results.

3.5.3 Obedience of state bounds

This subsection shows that the unique solutions of (3.15) lie within the state bounds $\boldsymbol{x}^L, \boldsymbol{x}^U$ from Assumption 3.1.

Lemma 3.2. *Under Assumptions 3.2 and 3.3, let* $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ *be the unique solution of* (3.15) *on* $I' \times P$. *Then* $\boldsymbol{x}^{cv}(t, \boldsymbol{p}) \geq \boldsymbol{x}^L(t)$ *and* $\boldsymbol{x}^{cc}(t, \boldsymbol{p}) \leq \boldsymbol{x}^U(t)$ *for all* $(t, \boldsymbol{p}) \in I' \times P$.

Proof. We will show that $\boldsymbol{x}^{cv}(t, \boldsymbol{p}) \geq \boldsymbol{x}^L(t)$ for every $(t, \boldsymbol{p}) \in I' \times P$. An analogous argument then shows that $\boldsymbol{x}^{cc}(t, \boldsymbol{p}) \leq \boldsymbol{x}^U(t)$.

To achieve a contradiction, suppose that there exist $\hat{\boldsymbol{p}} \in P$, $i \in \{1, \ldots, n_x\}$, and $\tau \in I'$, for which $x_i^{cv}(\tau, \hat{\boldsymbol{p}}) < x_i^L(\tau)$. Define $S := \{s \in [t_0, \tau] : x_i^{cv}(s, \hat{\boldsymbol{p}}) \geq x_i^L(s)\}$ and $t_1 := \sup S$. Since $x_i^{cv}(t_0, \hat{\boldsymbol{p}}) \geq x_i^L(t_0)$ by construction, the set S is non-empty, and the continuity of $x_i^{cv}(\cdot, \hat{\boldsymbol{p}})$ and x_i^L ensures that $t_1 \in [t_0, \tau)$. Because t_1 is an upper bound of S, we have $x_i^{cv}(t, \hat{\boldsymbol{p}}) < x_i^L(t)$ for all $t \in (t_1, \tau]$, and because t_1 is the least upper bound of S, the continuity of $x_i^{cv}(\cdot, \hat{\boldsymbol{p}})$ and $x_i^L(\cdot)$ implies that $x_i^{cv}(t_1, \hat{\boldsymbol{p}}) = x_i^L(t_1)$. Then, for all $t \in [t_1, \tau]$, (3.11) implies that $\sigma(2\underline{k}_i(x_i^{cv}(t, \hat{\boldsymbol{p}}) - x_i^L(t))) = 0$.

(3.13) and (3.15) yield that, for all $t \in [t_1, \tau]$,

$$\dot{x}_i^{cv}(t, \hat{\boldsymbol{p}}) = \max\left\{u_i(t, \hat{\boldsymbol{p}}, \boldsymbol{x}^{cv}(t, \hat{\boldsymbol{p}}), \boldsymbol{x}^{cc}(t, \hat{\boldsymbol{p}})),\ \dot{x}_i^L(t)\right\} \geq \dot{x}_i^L(t),$$

Applying Theorem 3.1 in [137] (as used similarly by [120]), it follows that $(x_i^L - x_i^{cv}(\cdot, \hat{\boldsymbol{p}}))$ is non-increasing on $[t_1, \tau]$. Hence,

$$0 = x_i^L(t_1) - x_i^{cv}(t_1, \hat{\boldsymbol{p}}) \geq x_i^L(\tau) - x_i^{cv}(\tau, \hat{\boldsymbol{p}}).$$

Then, $x_i^L(\tau) \leq x_i^{cv}(\tau, \hat{\boldsymbol{p}})$, which yields the desired contradiction. □

Lemma 3.2 shows that the arguments of σ in (3.14) and (3.15) are always non-negative along solution trajectories of (3.15). So the functions $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in (3.14) and (3.15) can be simplified as

$$\begin{aligned} \alpha_i(t, \boldsymbol{\theta}) &= \dot{x}_i^L(t) - \sqrt{2\underline{k}_i(\theta_i - x_i^L(t))}, \\ \beta_i(t, \boldsymbol{\theta}) &= \dot{x}_i^U(t) + \sqrt{2\overline{k}_i(x_i^U(t) - \theta_i)}, \end{aligned} \tag{3.24}$$

for $i \in \{1, \ldots, n_x\}$. In the remaining parts, we will use (3.24) instead of (3.12).

3.5.4 Enclosing the reachable set

This subsection verifies that, under Assumptions 3.1 and 3.3, $\boldsymbol{v}, \boldsymbol{w}$ in (3.14) describe bound-preserving dynamics for (3.1). A differential inequality-based result is then developed to show that solutions of (3.15) encloses enclose the reachable set of the original ODE (3.1).

Lemma 3.3. *Suppose that Assumptions 3.1 and 3.3 hold, and that $\boldsymbol{\xi}(t)$ solves the ODE (3.2). Then, for any $\tau \in I'$, $\boldsymbol{\xi}_0 \in X^B(\tau)$, $\boldsymbol{p} \in P$, and $i \in \{1, \ldots, n_x\}$, the following holds:*

1. *If $f_i(\tau, \boldsymbol{p}, \boldsymbol{\xi}_0) < \dot{x}_i^L(\tau)$,*

$$2\underline{k}_i \left(\xi_{0,i} - x_i^L(\tau)\right) \geq \left(f_i(\tau, \boldsymbol{p}, \boldsymbol{\xi}_0) - \dot{x}_i^L(\tau)\right)^2.$$

2. *If $f_i(\tau, \boldsymbol{p}, \boldsymbol{\xi}_0) > x_i^U(\tau)$,*

$$2\overline{k}_i \left(x_i^U(\tau) - \xi_{0,i}\right) \geq \left(\dot{x}_i^U(\tau) - f_i(\tau, \boldsymbol{p}, \boldsymbol{\xi}_0)\right)^2.$$

Proof. It will be shown that the first result holds; it is analogous to verify the second.

Suppose we choose arbitrary $\hat{\tau} \in I'$, $\boldsymbol{\xi}_0 \in X^B(\tau)$, and $\boldsymbol{p} \in P$ for which $f_i(\hat{\tau}, \boldsymbol{p}, \boldsymbol{\xi}_0) < \dot{x}_i^L(\hat{\tau})$. To achieve a contradiction, suppose that

$$2\underline{k}_i \left(\xi_{0,i} - x_i^L(\hat{\tau})\right) < \left(f_i(\hat{\tau}, \boldsymbol{p}, \boldsymbol{\xi}_0) - \dot{x}_i^L(\hat{\tau})\right)^2. \tag{3.25}$$

Let $t := \hat{\tau} + \frac{\dot{x}_i^L(\hat{\tau}) - f_i(\hat{\tau}, \boldsymbol{p}, \boldsymbol{\xi}_0)}{\underline{k}_i}$. Remark 3.1 ensures that $t \in I'$. Now,

$$\begin{aligned}
&\xi_i(t) - x_i^L(t) \\
&= \xi_i(\tau) - x_i^L(\hat{\tau}) + (t - \hat{\tau}) \int_0^1 \left(\dot{\xi}_i(\hat{\tau} + s(t - \hat{\tau})) - \dot{x}_i^L(\hat{\tau} + s(t - \hat{\tau}))\right) \mathrm{d}s, \\
&= \xi_i(\tau) - x_i^L(\hat{\tau}) + (t - \hat{\tau}) \left(\dot{\xi}_i(\hat{\tau}) - \dot{x}_i^L(\hat{\tau})\right) \\
&\quad + (t - \hat{\tau}) \int_0^1 \left(\dot{\xi}_i(\hat{\tau} + s(t - \hat{\tau}), \boldsymbol{p}) - \dot{x}_i^L(\hat{\tau} + s(t - \hat{\tau})) - \left(\dot{\xi}_i(\hat{\tau}) - \dot{x}_i^L(\hat{\tau})\right)\right) \mathrm{d}s.
\end{aligned}$$

The first condition in Definition 3.9 then implies:

$$\begin{aligned}
&\xi_i(t) - x_i^L(t) \\
&\leq \xi_i(\tau) - x_i^L(\hat{\tau}) + (t-\hat{\tau})\left(\dot{\xi}_i(\hat{\tau}) - \dot{x}_i^L(\hat{\tau})\right) + \underline{k}_i(t-\hat{\tau})^2 \int_0^1 s \, \mathrm{d}s \\
&= \xi_i(\tau) - x_i^L(\hat{\tau}) + (t-\hat{\tau})\left(\dot{\xi}_i(\hat{\tau}) - \dot{x}_i^L(\hat{\tau})\right) + \frac{\underline{k}_i}{2}(t-\hat{\tau})^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
2\underline{k}_i\left(\xi_i(t) - x_i^L(t)\right) &\leq 2\underline{k}_i\left(\xi_i(\tau) - x_i^L(\hat{\tau})\right) + 2\underline{k}_i\left(\dot{\xi}_i(\hat{\tau}) - \dot{x}_i^L(\hat{\tau})\right)(t-\hat{\tau}) \\
&\quad + \left(\underline{k}_i\,(t-\hat{\tau})\right)^2, \\
&= 2\underline{k}_i\left(\xi_i(\tau) - x_i^L(\hat{\tau})\right) - \left(\dot{\xi}_i(\hat{\tau}) - \dot{x}_i^L(\hat{\tau})\right)^2 \\
&\quad + \left(\underline{k}_i\,(t-\hat{\tau}) + \dot{\xi}_i(\hat{\tau}) - \dot{x}_i^L(\hat{\tau})\right)^2. \qquad (3.26)
\end{aligned}$$

Substituting $t := \hat{\tau} + \frac{\dot{x}_i^L(\hat{\tau}) - \dot{\xi}_i(\hat{\tau})}{\underline{k}_i}$ into (3.26) and applying (3.25) yields:

$$\begin{aligned}
2\underline{k}_i\left(\xi_i(t) - x_i^L(t)\right) &\leq 2\underline{k}_i\left(\xi_i(\hat{\tau}) - x_i^L(\hat{\tau})\right) - \left(f_i(\hat{\tau}, \boldsymbol{p}, \boldsymbol{\xi}_0) - \dot{x}_i^L(\hat{\tau})\right)^2 \\
&< 0, \qquad (3.27)
\end{aligned}$$

and so $\xi_i(t) < x_i^L(t)$, which contradicts the fact that x_i^L underestimates ξ_i. □

Lemma 3.4. *Under Assumptions 3.1, 3.2, and 3.3,* $(\boldsymbol{v}, \boldsymbol{w})$ *describe bound-preserving dynamics for* (3.1).

Proof. Consider any $t \in I'$, $\boldsymbol{z}, \boldsymbol{\phi}, \boldsymbol{\psi} \in X^B(t)$, and choose any fixed $\boldsymbol{p} \in P$. It is desired to show that, for each $i \in \{1, \ldots, n_x\}$,

1. if $z_i = \phi_i$, then $v_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \leq f_i(t, \boldsymbol{p}, \boldsymbol{z})$, and

2. if $z_i = \psi_i$, then $w_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \geq f_i(t, \boldsymbol{p}, \boldsymbol{z})$.

We will show that the first condition holds; showing the second is analogous.

Consider the following two scenarios. First, if $f_i(t, \boldsymbol{p}, \boldsymbol{z}) < \dot{x}_i^L(t)$, then according to Lemma 3.3,

$$2\underline{k}_i(z_i - x_i^L(t)) \geq (f_i(t, \boldsymbol{p}, \boldsymbol{z}) - \dot{x}_i^L(t))^2,$$

and so

$$\dot{x}_i^L(t) - \sqrt{2\underline{k}_i(z_i - x_i^L(t))} \leq f_i(t, \boldsymbol{p}, \boldsymbol{z}).$$

When $z_i = \phi_i$, the above inequality becomes

$$\dot{x}_i^L(t) - \sqrt{2\underline{k}_i(\phi_i - x_i^L(t))} = \alpha_i(t, \boldsymbol{\phi}) \leq f_i(t, \boldsymbol{p}, \boldsymbol{z}). \tag{3.28}$$

Next, if $f_i(t, \boldsymbol{p}, \boldsymbol{z}) \geq \dot{x}_i^L(t)$, then (3.28) trivially holds. Combining the above two scenarios, if $z_i = \phi_i$, then $\alpha_i(t, \boldsymbol{\phi}) \leq f_i(t, \boldsymbol{p}, \boldsymbol{z})$.

According to Assumption 3.2, $(\boldsymbol{u}, \boldsymbol{o})$ describe bounding-preserving dynamics for (3.1). So if $z_i = \phi_i$, then $u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \leq f_i(t, \boldsymbol{p}, \boldsymbol{z})$, and therefore

$$\max\{u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}), \alpha_i(t, \boldsymbol{\phi})\} = v_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \leq f_i(t, \boldsymbol{p}, \boldsymbol{z}).$$

□

For convenience, the following two propositions reproduce Lemma 4 and Theorem 2 from [120].

Proposition 3.1. *Let* $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ *be a solution of* (3.6) *on* $I' \times P$. *Then* $\boldsymbol{x}^{cv}(t, \boldsymbol{p}) \leq \boldsymbol{x}^{cc}(t, \boldsymbol{p})$ *for all* $(t, \boldsymbol{p}) \in I' \times P$.

Proposition 3.2. *Suppose that for each $t \in I'$, $\boldsymbol{x}(t,\boldsymbol{p}) \in X^B(t)$ for all $\boldsymbol{p} \in P$. Let $\boldsymbol{\phi}, \boldsymbol{\psi} : I' \times P \to \mathbb{R}^{n_x}$ be continuous functions satisfying*

$$\text{(EX)} : [\boldsymbol{\phi}(t), \boldsymbol{\psi}(t)] \cap X^B(t) \neq \emptyset, \ \forall t \in I'.$$

$$\text{(IC)} : \boldsymbol{\phi}(t_0) \leq \boldsymbol{x}_0(\boldsymbol{p}) \leq \boldsymbol{\psi}(t_0), \ \forall \boldsymbol{p} \in P.$$

(RHS) : *For a.e. $t \in I'$ and each index i,*

1. $\dot{\phi}_i(t) \leq f_i(t,\boldsymbol{p},\boldsymbol{z})$ *if* $\boldsymbol{p} \in P, \boldsymbol{z} \in B_i^L([\boldsymbol{\phi}(t), \boldsymbol{\psi}(t)]) \cap X^B(t) \cap D$ *and* $\phi_i(t) > x_i^L(t)$,
2. $\dot{\psi}_i(t) \geq f_i(t,\boldsymbol{p},\boldsymbol{z})$ *if* $\boldsymbol{p} \in P, \boldsymbol{z} \in B_i^U([\boldsymbol{\phi}(t), \boldsymbol{\psi}(t)]) \cap X^B(t) \cap D$ *and* $\psi_i(t) < x_i^U(t)$.

Then $\boldsymbol{\phi}(t) \leq \boldsymbol{x}(t,\boldsymbol{p}) \leq \boldsymbol{\psi}(t)$ for all $(t,\boldsymbol{p}) \in I' \times P$.

Theorem 3.3. *Under Assumptions 3.1, 3.2, and 3.3, let $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ be solutions of* (3.15) *on $I' \times P$. Then, $\boldsymbol{x}^{cv}(t,\boldsymbol{p}) \leq \boldsymbol{x}(t,\boldsymbol{p}) \leq \boldsymbol{x}^{cc}(t,\boldsymbol{p})$ for all $(t,\boldsymbol{p}) \in I' \times P$.*

Proof. Choose any $(t,\boldsymbol{p}) \in I' \times P$. We will show that the three conditions in Proposition 3.2 are satisfied with $\boldsymbol{\phi} := \boldsymbol{x}^{cv}(\cdot,\boldsymbol{p})$ and $\boldsymbol{\psi} := \boldsymbol{x}^{cc}(\cdot,\boldsymbol{p})$. Lemma 3.2 shows that $\boldsymbol{x}^{cv}(\tau,\boldsymbol{p}), \boldsymbol{x}^{cc}(\tau,\boldsymbol{p}) \in X^B(\tau)$ for all $\tau \in I'$, and so $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$ are solutions to (3.6). Proposition 3.1 shows that $\boldsymbol{x}^{cv}(t,\boldsymbol{p}) \leq \boldsymbol{x}^{cc}(t,\boldsymbol{p})$. Hence, (EX) holds. By construction, $\boldsymbol{x}_0^{cv}(\boldsymbol{p}) \leq \boldsymbol{x}_0(\boldsymbol{p}) \leq \boldsymbol{x}_0^{cc}(\boldsymbol{p})$, which shows that (IC) is satisfied. Lemma 3.4 shows that $(\boldsymbol{v}, \boldsymbol{w})$ describe bound-preserving dynamics for (3.1). Therefore, for

each $i \in \{1, \ldots, n_x\}$,

1. $\dot{x}_i^{cv}(t, \boldsymbol{p}) \leq f_i(t, \boldsymbol{p}, \boldsymbol{z})$
   if $\boldsymbol{z} \in B_i^L(\boldsymbol{x}^{cv}(t, \boldsymbol{p}), \boldsymbol{x}^{cc}(t, \boldsymbol{p})) \cap X^B(t) \cap D$ and $x_i^{cv}(t, \boldsymbol{p}) > x_i^L(t)$,
2. $\dot{x}_i^{cc}(t, \boldsymbol{p}) \geq f_i(t, \boldsymbol{p}, \boldsymbol{z})$
   if $\boldsymbol{z} \in B_i^U(\boldsymbol{x}^{cv}(t, \boldsymbol{p}), \boldsymbol{x}^{cc}(t, \boldsymbol{p})) \cap X^B(t) \cap D$ and $x_i^{cc}(t, \boldsymbol{p}) < x_i^U(t)$.

As a result, both (RHS) conditions hold. Thus, Proposition 3.2 yields the required result. □

3.5.5 Obtaining safe-landing constants

To ensure the bounding property developed in this section, appropriate safe-landing constants $\underline{\boldsymbol{k}}, \overline{\boldsymbol{k}}$ are desired. In the lemma below, we show that $\underline{\boldsymbol{k}}, \overline{\boldsymbol{k}}$ can be calculated using Lipschitz constants of $\dot{\boldsymbol{\xi}}(\cdot, \boldsymbol{p})$ and $\dot{\boldsymbol{x}}^L, \dot{\boldsymbol{x}}^U$.

Assumption 3.4. *Suppose Assumption 3.1 holds. Assume that there exist $\boldsymbol{k}^t, \boldsymbol{k}^L, \boldsymbol{k}^U \in \mathbb{R}^{n_x}_{>0}$ such that, for any $t_1, t_2 \in I$, $\boldsymbol{p} \in P$, and $i \in \{1, \ldots, n_x\}$,*

$$\begin{aligned} |\dot{\xi}_i(t_1) - \dot{\xi}_i(t_2)| &\leq k_i^t |t_1 - t_2|, \\ |\dot{x}_i^L(t_1) - \dot{x}_i^L(t_2)| &\leq k_i^L |t_1 - t_2|, \\ \text{and} \quad |\dot{x}_i^U(t_1) - \dot{x}_i^U(t_2)| &\leq k_i^U |t_1 - t_2|. \end{aligned}$$

Lemma 3.5. *Under Assumptions 3.1 and 3.4, define*

$$\underline{\boldsymbol{k}} := \boldsymbol{k}^t + \boldsymbol{k}^L, \qquad \overline{\boldsymbol{k}} := \boldsymbol{k}^t + \boldsymbol{k}^U. \tag{3.29}$$

Then, $(\underline{\boldsymbol{k}}, \overline{\boldsymbol{k}})$ *are safe-landing constants for* (3.1) *and* X^B.

Proof. For any $\tau \in I$, $\boldsymbol{p} \in P$, $\boldsymbol{\xi}_0 \in X^B(\tau)$, let $\boldsymbol{\xi}(t)$ solve (3.2). It is desired to verify the two conditions in Definition 3.9. For any $i \in \{1, \ldots, n\}$, it will be shown that the first condition in Assumption 3.3 holds: If $\dot{\xi}_i(\tau) < \dot{x}_i^L(\tau)$, then, for each $t \in [\tau, \tau + \frac{\dot{x}_i^L(\tau) - \dot{\xi}_i(\tau)}{\underline{k}_i}]$,

$$\left(\dot{\xi}_i(t) - \dot{x}_i^L(t)\right) - \left(\dot{\xi}_i(\tau) - \dot{x}_i^L(\tau)\right) \leq \underline{k}_i(t - \tau).$$

It is analogous to verify the second condition.

Consider the non-negative differentiable function $\boldsymbol{\xi} - \boldsymbol{x}^L$. Lemma 3.1 and Remark 3.1 guarantee that, if $\dot{\xi}_i(\tau) < \dot{x}_i^L(\tau)$,

$$\tau + \frac{\dot{x}_i^L(\tau) - \dot{\xi}_i(\tau)}{\underline{k}_i} \in I.$$

Assumption 3.4 ensures that, for any $t_1, t_2 \in I$, $\boldsymbol{p} \in P$, and $i \in \{1, \ldots, n_x\}$,

$$|\dot{\xi}_i(t_1) - \dot{\xi}_i(t_2)| \leq k_i^t |t_1 - t_2|, \quad \text{and}$$
$$|\dot{x}_i^L(t_1) - \dot{x}_i^L(t_2)| \leq k_i^L |t_1 - t_2|.$$

Thus, for any $t \in [\tau, \tau + \frac{x_i^L(\tau) - \dot{\xi}_i(\tau)}{\underline{k}_i}]$,

$$\begin{aligned}
&\left(\dot{\xi}_i(t) - \dot{x}_i^L(t)\right) - \left(\dot{\xi}_i(\tau) - \dot{x}_i^L(\tau)\right) \\
&\leq \left|\left(\dot{\xi}_i(t) - \dot{x}_i^L(t)\right) - \left(\dot{\xi}_i(\tau) - \dot{x}_i^L(\tau)\right)\right| \\
&= \left|\left(\dot{\xi}_i(t) - \dot{\xi}_i(\tau)\right) - \left(\dot{x}_i^L(t) - \dot{x}_i^L(\tau)\right)\right| \\
&\leq \left|\dot{\xi}_i(t) - \dot{\xi}_i(\tau)\right| + \left|\dot{x}_i^L(t) - \dot{x}_i^L(\tau)\right| \\
&\leq k_i^t|t - \tau| + k_i^L|t - \tau| \\
&= \underline{k}_i(t - \tau),
\end{aligned}$$

which ensures the first condition in Assumption 3.3.

□

3.5.6 Convexity and concavity

This subsection shows that the unique solutions of (3.15), $\boldsymbol{x}^{cv}(t, \cdot)$ and $\boldsymbol{x}^{cc}(t, \cdot)$, are convex and concave over P, respectively.

Lemma 3.6. *Under Assumption 3.1 and 3.3,* $(\boldsymbol{v}, \boldsymbol{w})$ *describe convexity-preserving dynamics for* (3.1)*.*

Proof. Consider any $t \in I'$, $(\lambda, \boldsymbol{p}_1, \boldsymbol{p}_2) \in (0,1) \times P \times P$, $\bar{\boldsymbol{p}} := \lambda \boldsymbol{p}_1 + (1-\lambda)\boldsymbol{p}_2$, and $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \bar{\boldsymbol{\phi}}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \bar{\boldsymbol{\psi}} \in X^B(t)$ such that

1. $\bar{\boldsymbol{\phi}} \leq \lambda \boldsymbol{\phi}_1 + (1-\lambda)\boldsymbol{\phi}_2$,
2. $\bar{\boldsymbol{\psi}} \geq \lambda \boldsymbol{\psi}_1 + (1-\lambda)\boldsymbol{\psi}_2$, and
3. $\boldsymbol{\phi}_1 \leq \boldsymbol{\psi}_1$, $\boldsymbol{\phi}_2 \leq \boldsymbol{\psi}_2$, $\bar{\boldsymbol{\phi}} \leq \bar{\boldsymbol{\psi}}$,

It suffices to show that, for any $i \in \{1, \ldots, n_x\}$,

1. If $\bar{\phi}_i = \lambda \phi_{1,i} + (1-\lambda)\phi_{2,i}$, then

$$v_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \leq \lambda v_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) + (1-\lambda) v_i(t, \boldsymbol{p}_2, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2). \tag{3.30}$$

2. If $\bar{\psi}_i = \lambda \psi_{1,i} + (1-\lambda)\psi_{2,i}$, then

$$w_i(t, \tilde{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \geq \lambda w_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) + (1-\lambda) w_i(t, \boldsymbol{p}_2, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2).$$

It will be shown that the first of these conditions holds; showing the second is analogous.

Consider $v_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) = \max\{u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}), \alpha_i(t, \boldsymbol{\phi})\}$ in the following cases:

1. If $u_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) \geq \alpha_i(t, \boldsymbol{\phi}_1)$ and $u_i(t, \boldsymbol{p}_2, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2) \geq \alpha_i(t, \boldsymbol{\phi}_2)$

   (i) if $u_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \geq \alpha_i(t, \bar{\boldsymbol{\phi}})$, (3.30) becomes

$$u_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \leq \lambda u_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) + (1-\lambda) u_i(t, \boldsymbol{p}_2, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2),$$

   which holds in this case because $\boldsymbol{u}$ describes convexity preserving dynamics.

   (ii) if $u_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) < \alpha_i(t, \bar{\boldsymbol{\phi}})$, (3.30) becomes

$$\alpha_i(t, \bar{\boldsymbol{\phi}}) \leq \lambda u_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) + (1-\lambda) u_i(t, \boldsymbol{p}_2, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2).$$

In this case, the above inequality will be proved by showing

$$\alpha_i(t, \bar{\phi}) \leq \lambda\alpha_i(t, \phi_1) + (1-\lambda)\alpha_i(t, \phi_2),$$

which is equivalent to

$$\begin{aligned}
&\dot{x}_i^L(t) - \sqrt{2\underline{k}_i(\bar{\phi}_i - x_i^L(t))} \\
&\leq \lambda\left(\dot{x}_i^L(t) - \sqrt{2\underline{k}_i(\phi_{1,i} - x_i^L(t))}\right) + (1-\lambda)\left(\dot{x}_i^L(t) - \sqrt{2\underline{k}_i(\phi_{2,i} - x_i^L(t))}\right) \\
&= \dot{x}_i^L(t) + \lambda\left(-\sqrt{2\underline{k}_i(\phi_{1,i} - x_i^L(t))}\right) + (1-\lambda)\left(-\sqrt{2\underline{k}_i(\phi_{2,i} - x_i^L(t))}\right).
\end{aligned}$$

Rearranging the above inequality,

$$\sqrt{(\bar{\phi}_i - x_i^L(t))} \geq \lambda\sqrt{\phi_{1,i} - x_i^L(t)} + (1-\lambda)\sqrt{\phi_{2,i} - x_i^L(t)}.$$

Because $\bar{\phi}_i = \lambda\phi_{1,i} + (1-\lambda)\phi_{2,i}$, the above inequality is equivalent to

$$\begin{aligned}
&\sqrt{\lambda(\phi_{1,i} - x_i^L(t)) + (1-\lambda)(\phi_{2,i} - x_i^L(t))} \\
&\quad \geq \lambda\sqrt{\phi_{1,i} - x_i^L(t)} + (1-\lambda)\sqrt{\phi_{2,i} - x_i^L(t)}.
\end{aligned}$$

Because both sides of the above inequalities are non-negative, it is equivalent to

$$\begin{aligned}
&\lambda\left(\phi_{1,i} - x_i^L(t)\right) + (1-\lambda)\left(\phi_{2,i} - x_i^L(t)\right) \\
&\quad \geq \lambda^2(\phi_{1,i} - x_i^L(t)) + (1-\lambda)^2(\phi_{2,i} - x_i^L(t)) \\
&\quad + 2\lambda(1-\lambda)\sqrt{(\phi_{1,i} - x_i^L(t))}\sqrt{(\phi_{2,i} - x_i^L(t))}.
\end{aligned}$$

Rearranging the above inequality,

$$\begin{aligned}&\phi_{1,i} - x_i^L(t) + \phi_{2,i} - x_i^L(t)\\ &\geq 2\sqrt{(\phi_{1,i} - x_i^L(t))}\sqrt{(\phi_{2,i} - x_i^L(t))}.\end{aligned}$$

The above inequality always holds, and is equivalent to (3.30) in this case.

2. If $u_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) \geq \alpha_i(t, \boldsymbol{\phi}_1)$ and $u_i(t, \boldsymbol{p}_2, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2) < \alpha_i(t, \boldsymbol{\phi}_2)$, then:

   (i) if $u_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \geq \alpha_i(t, \bar{\boldsymbol{\phi}})$, (3.30) becomes

   $$u_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \leq \lambda u_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) + (1-\lambda)\alpha_i(t, \boldsymbol{\phi}_2).$$

   In this case, the above inequality is true because $u_i(t, \boldsymbol{p}_2, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2) < \alpha_i(t, \boldsymbol{\phi}_2)$ and $\boldsymbol{u}$ describes convexity preserving dynamics.

   (ii) if $u_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) < \alpha_i(t, \bar{\boldsymbol{\phi}})$, (3.30) becomes

   $$\alpha_i(t, \bar{\boldsymbol{\phi}}) \leq \lambda u_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) + (1-\lambda)\alpha_i(t, \boldsymbol{\phi}_2).$$

   In this case, the above inequality can be proved by showing

   $$\alpha_i(t, \bar{\boldsymbol{\phi}}) \leq \lambda \alpha_i(t, \boldsymbol{\phi}_1) + (1-\lambda)\alpha_i(t, \boldsymbol{\phi}_2).$$

   This inequality holds here as shown in case 1(ii).

3. If $u_i(t, \boldsymbol{p}_1, \boldsymbol{\phi}_1, \boldsymbol{\psi}_1) < \alpha_i(t, \boldsymbol{\phi}_1)$ and $u_i(t, \boldsymbol{p}_2, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2) \geq \alpha_i(t, \boldsymbol{\phi}_2)$, then the argument in case 2 applies here, after interchanging $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$.

4. If $u_i(t, \boldsymbol{p}_1, \phi_1, \psi_1) < \alpha_i(t, \phi_1)$ and $u_i(t, \boldsymbol{p}_2, \phi_2, \psi_2) < \alpha_i(t, \phi_2)$, then:

   (i) if $u_i(t, \bar{\boldsymbol{p}}, \bar{\phi}, \bar{\psi}) \geq \alpha_i(t, \bar{\phi})$, (3.30) becomes

   $$u_i(t, \bar{\boldsymbol{p}}, \bar{\phi}, \bar{\psi}) \leq \lambda \alpha_i(t, \phi_1) + (1 - \lambda)\alpha_i(t, \phi_2).$$

   In this case, we only need to show

   $$u_i(t, \bar{\boldsymbol{p}}, \bar{\phi}, \bar{\psi}) \leq \lambda u_i(t, \boldsymbol{p}_1, \phi_1, \psi_1) + (1 - \lambda)u_i(t, \boldsymbol{p}_2, \phi_2, \psi_2).$$

   The above inequality holds here because $\boldsymbol{u}$ describes convexity preserving dynamics.

   (ii) if $u_i(t, \bar{\boldsymbol{p}}, \bar{\phi}, \bar{\psi}) < \alpha_i(t, \bar{\phi})$, (3.30) becomes

   $$\alpha_i(t, \bar{\phi}) \leq \lambda \alpha_i(t, \phi_1) + (1 - \lambda)\alpha_i(t, \phi_2).$$

   The above inequality holds here as shown in case 1(ii).

Combining the above cases, (3.30) holds when $\bar{\phi}_i = \lambda \phi_{1,i} + (1 - \lambda)\phi_{2,i}$. □

The following proposition is adapted from [120, Theorem 3].

Proposition 3.3. *Let* $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ *be a solution of* (3.6) *on* $I' \times P$. *Then,* $\boldsymbol{x}^{cv}(t, \cdot)$ *and* $\boldsymbol{x}^{cc}(t, \cdot)$ *are respectively convex and concave on* P, *for every* $t \in I'$.

Theorem 3.4. *Under Assumptions 3.1, 3.2, and 3.3, let* $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ *be a solution of* (3.15) *on* $I' \times P$. *Then,* $\boldsymbol{x}^{cv}(t, \cdot)$ *and* $\boldsymbol{x}^{cc}(t, \cdot)$ *are respectively convex and concave on* P, *for every* $t \in I'$.

Proof. Lemma 3.2 implies that $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ are solutions to (3.6) with $(\boldsymbol{v}, \boldsymbol{w})$ replacing $(\boldsymbol{u}, \boldsymbol{o})$. Lemmas 3.4 and 3.6 show that $(\boldsymbol{v}, \boldsymbol{w})$ describe bound-preserving dynamics and convexity-preserving dynamics for (3.1). Thus $\boldsymbol{x}^{cv}(t, \cdot)$ and $\boldsymbol{x}^{cc}(t, \cdot)$ are respectively convex and concave on P, for every $t \in I'$ according to Proposition 3.3. □

Theorems 3.3 and 3.4 show that the solutions to (3.15) are valid state relaxations of (3.1).

3.5.7 Tighter State Relaxations

This subsection shows that our new state relaxations are at least as tight as the Scott-Barton relaxations. In addition to Assumption 3.2, we suppose that $(\boldsymbol{u}, \boldsymbol{o})$ satisfy the following assumption.

Assumption 3.5. *Consider any $i \in \{1, \ldots, n_x\}$, $\boldsymbol{p} \in P$, and $t \in I$. For all $\boldsymbol{\phi}, \boldsymbol{\psi}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}} \in X^B(t)$ such that $\hat{\boldsymbol{\phi}} \leq \boldsymbol{\phi} \leq \boldsymbol{\psi} \leq \hat{\boldsymbol{\psi}}$,*

1. *if $\phi_i = \hat{\phi}_i$, then $u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \geq u_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}})$,*
2. *if $\psi_i = \hat{\psi}_i$, then $o_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \leq o_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}})$.*

Definition 3.14. *An interval function $H : \mathbb{IR}^n \to \mathbb{IR}^m$ is* inclusion monotonic *if for all $Z_1, Z_2 \in \mathbb{IR}^n$ such that $Z_1 \subset Z_2$, we have $H(Z_1) \subseteq H(Z_2)$.*

Assumption 3.5 holds if the interval-valued function $F^R \equiv [\boldsymbol{u}, \boldsymbol{o}] : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ is inclusion monotonic on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, which is always satisfied by GMR [117, 113], DMR [72], and an optimization-based relaxation approach by [131]. Moreover, if the flattening operators in Definition 3.4 are applied to any such $\boldsymbol{u}, \boldsymbol{o}$ on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, the resulting functions also satisfy Assumption 3.5. This result is justified in Section 3.6.2.

Theorem 3.5. *Under Assumptions 3.1, 3.2, 3.3, and 3.5, let* $(\hat{\boldsymbol{x}}^{cv}, \hat{\boldsymbol{x}}^{cc})$ *be a solution of* (3.6)*, and* $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ *be a solution of* (3.15) *on* $I' \times P$*. If the same state bounds* $(\boldsymbol{x}^L, \boldsymbol{x}^U)$ *and same initial conditions* $(\boldsymbol{x}_0^{cv}, \boldsymbol{x}_0^{cc})$ *are used for both* (3.6) *and* (3.15)*, then it holds that* $[\boldsymbol{x}^{cv}(t,\boldsymbol{p}), \boldsymbol{x}^{cc}(t,\boldsymbol{p})] \subseteq [\hat{\boldsymbol{x}}^{cv}(t,\boldsymbol{p}), \hat{\boldsymbol{x}}^{cc}(t,\boldsymbol{p})]$ *for all* $(t,\boldsymbol{p}) \in I' \times P$*.*

Proof. This theorem will proved by showing that all requirements in [130, Theorem 2] are satisfied with $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$ in place of $\boldsymbol{v}^B, \boldsymbol{w}^B$ and $\hat{\boldsymbol{x}}^{cv}, \hat{\boldsymbol{x}}^{cc}$ in place of $\boldsymbol{v}^A, \boldsymbol{w}^A$. Consider any $i \in \{1,\ldots,n_x\}$, $\boldsymbol{p} \in P$, $t \in I'$, and $\boldsymbol{\phi}, \boldsymbol{\psi}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}} \in X^B(t)$ such that $\hat{\boldsymbol{\phi}} \leq \boldsymbol{\phi} \leq \boldsymbol{\psi} \leq \hat{\boldsymbol{\psi}}$.

Lemma 1 in [120] and Lemma 3.2 in this article ensure Condition II.1 in [130, Theorem 2] with the state bound X^B in place of both C^A and C^B. For Condition II.2, we will show that the first inequality holds; showing the second is analogous. Consider the following exhaustive scenarios, with $\boldsymbol{d}^{L,A}$ defined as in [130].

- If $\phi_i \geq \hat{\phi}_i > x_i^L(t)$, then the first case of (3.6a) is selected for both $\dot{x}_i^{cv}, \dot{\hat{x}}_i^{cv}$ such that

$$d_i^{L,A}(t,\boldsymbol{\phi},\boldsymbol{\psi}) = u_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}), \quad d_i^{L,A}(t,\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}}) = u_i(t,\boldsymbol{p},\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}}).$$

  (3.5) implies that

$$u_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}) - u_i(t,\boldsymbol{p},\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}}) \leq k^r \left(\left\| \boldsymbol{\phi} - \hat{\boldsymbol{\phi}} \right\| + \left\| \boldsymbol{\psi} - \hat{\boldsymbol{\psi}} \right\| \right),$$

  which ensures the first inequality of Condition II.2.

- If $\phi_i > \hat{\phi}_i = x_i^L(t)$, then,

$$d_i^{L,A}(t,\boldsymbol{\phi},\boldsymbol{\psi}) = u_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}), \quad d_i^{L,A}(t,\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}}) = \max\left\{u_i(t,\boldsymbol{p},\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}}), \dot{x}_i^L(t)\right\}.$$

  The first inequality then follows from

$$\begin{aligned} u_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}) - \max\left\{u_i(t,\boldsymbol{p},\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}}), \dot{x}_i^L(t)\right\} &\leq u_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}) - u_i(t,\boldsymbol{p},\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}}) \\ &\leq k^r\left(\left\|\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}\right\| + \left\|\boldsymbol{\psi}-\hat{\boldsymbol{\psi}}\right\|\right). \end{aligned} \tag{3.31}$$

- If $\phi_i = \hat{\phi}_i = x_i^L(t)$, then,

$$\begin{aligned} d_i^{L,A}(t,\boldsymbol{\phi},\boldsymbol{\psi}) &= \max\left\{u_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}), \dot{x}_i^L(t)\right\}, \\ d_i^{L,A}(t,\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}}) &= \max\left\{u_i(t,\boldsymbol{p},\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}}), \dot{x}_i^L(t)\right\}. \end{aligned}$$

  We further consider the following two cases of $d_i^{L,A}(t,\boldsymbol{\phi},\boldsymbol{\psi})$. When $u_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}) \geq \dot{x}_i^L(t)$, we recover (3.31), which holds under Assumption 3.2. When $u_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}) < \dot{x}_i^L(t)$, the first inequality becomes

$$\begin{aligned} \dot{x}_i^L(t) - \max\left\{u_i(t,\boldsymbol{p},\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}}), \dot{x}_i^L(t)\right\} &\leq 0 \\ &\leq k^r\left(\left\|\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}\right\| + \left\|\boldsymbol{\psi}-\hat{\boldsymbol{\psi}}\right\|\right), \end{aligned}$$

  which holds in this case.

Combining the scenarios above, the first inequality in Condition II.2 in [130, Theorem 2] holds.

Next, we verify Condition II.3 in [130, Theorem 2] with $\boldsymbol{\phi}, \boldsymbol{\psi}$ in place of $\boldsymbol{\phi}^B, \boldsymbol{\psi}^B$, $\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}}$ in place of $\boldsymbol{\phi}^A, \boldsymbol{\psi}^A$, $\boldsymbol{u}, \boldsymbol{o}$ in place of $\boldsymbol{d}^{L,B}, \boldsymbol{d}^{U,B}$, and $\boldsymbol{v}, \boldsymbol{w}$ in place of $\boldsymbol{d}^{L,A}, \boldsymbol{d}^{U,A}$. Suppose that $\phi_i = \hat{\phi}_i$, and consider the following two scenarios. If $\phi_i = \hat{\phi}_i > x_i^L(t)$, then Assumption 3.5 ensures that

$$\begin{aligned} v_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) &= \max\left\{u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}), \dot{x}_i^L(t) - \sqrt{2\underline{k}_i(\phi_i - x_i^L(t))}\right\} \\ &\geq u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \\ &\geq u_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}}). \end{aligned}$$

If $\phi_i = \hat{\phi}_i = x_i^L(t)$, then

$$\begin{aligned} v_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) &= \max\left\{u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}), \dot{x}_i^L(t) - \sqrt{2\underline{k}_i(\phi_i - x_i^L(t))}\right\} \\ &\geq \max\left\{u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}), \dot{x}_i^L(t)\right\}. \end{aligned}$$

So Condition II.3(a) is satisfied. A similar argument yields Condition II.3(b). Lemma 4 in [120] and Theorem 3.3 ensure Condition II.4. Condition II.5. is trivially satisfied by these two systems. Therefore, all requirements are satisfied. □

One immediate consequence of the above theorem is that, the new state relaxations obtained by solving (3.15) also enjoy pointwise convergence with order 2 under the same assumptions as the Scott-Barton relaxations. This is important when using state relaxations in deterministic global optimization algorithms without invoking the "cluster problem" [48, 150].

Furthermore, choosing safe-landing constants $\underline{\boldsymbol{k}}$ and $\overline{\boldsymbol{k}}$ with smaller components will typically tighten state relaxations. This statement is supported by the

following theorem.

Theorem 3.6. *Under Assumptions 3.1, 3.2, 3.3, and 3.5, let* $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ *be a solution of* (3.15) *on* $I' \times P$. *Let* $\underline{\boldsymbol{k}}^*, \overline{\boldsymbol{k}}^* \in \mathbb{R}^{n_x}_{>0}$ *be another pair of safe-landing constants for* (3.1) *over* X^B. *Let* $\hat{\boldsymbol{v}}, \hat{\boldsymbol{w}} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ *be a variant of* $\boldsymbol{v}, \boldsymbol{w}$ *such that, for* $i \in \{1, \ldots, n_x\}$,

$$\hat{v}_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}}) = \max\left\{u_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}}), \dot{x}_i^L(t) - \sqrt{2\underline{k}_i^*(\hat{\phi}_i - x_i^L(t))}\right\},$$
$$\hat{w}_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}}) = \min\left\{o_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}}), \dot{x}_i^U(t) + \sqrt{2\overline{k}_i^*(x_i^U(t) - \hat{\psi}_i)}\right\}.$$

Let $(\hat{\boldsymbol{x}}^{cv}, \hat{\boldsymbol{x}}^{cc})$ *solve the following variant of* (3.15) *with the same initial conditions on* $I' \times P$:

$$\dot{\hat{\boldsymbol{x}}}^{cv}(t, \boldsymbol{p}) = \hat{\boldsymbol{v}}(t, \boldsymbol{p}, \hat{\boldsymbol{x}}^{cv}(t, \boldsymbol{p}), \hat{\boldsymbol{x}}^{cc}(t, \boldsymbol{p})), \quad \hat{\boldsymbol{x}}^{cv}(t_0, \boldsymbol{p}) = \boldsymbol{x}_0^{cv}(\boldsymbol{p}),$$
$$\dot{\hat{\boldsymbol{x}}}^{cc}(t, \boldsymbol{p}) = \hat{\boldsymbol{w}}(t, \boldsymbol{p}, \hat{\boldsymbol{x}}^{cv}(t, \boldsymbol{p}), \hat{\boldsymbol{x}}^{cc}(t, \boldsymbol{p})), \quad \hat{\boldsymbol{x}}^{cc}(t_0, \boldsymbol{p}) = \boldsymbol{x}_0^{cc}(\boldsymbol{p}).$$

If $\underline{\boldsymbol{k}} \leq \underline{\boldsymbol{k}}^*$ *and* $\overline{\boldsymbol{k}} \leq \overline{\boldsymbol{k}}^*$, *then* $[\boldsymbol{x}^{cv}(t, \boldsymbol{p}), \boldsymbol{x}^{cc}(t, \boldsymbol{p})] \subseteq [\hat{\boldsymbol{x}}^{cv}(t, \boldsymbol{p}), \hat{\boldsymbol{x}}^{cc}(t, \boldsymbol{p})]$ *for all* $(t, \boldsymbol{p}) \in I' \times P$.

Proof. To prove $[\boldsymbol{x}^{cv}(t, \boldsymbol{p}), \boldsymbol{x}^{cc}(t, \boldsymbol{p})] \subseteq [\hat{\boldsymbol{x}}^{cv}(t, \boldsymbol{p}), \hat{\boldsymbol{x}}^{cc}(t, \boldsymbol{p})]$ for all $(t, \boldsymbol{p}) \in I' \times P$, we will show that all the conditions in [130, Theorem 2] are satisfied with $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$ in place of $\boldsymbol{v}^B, \boldsymbol{w}^B$, $\hat{\boldsymbol{x}}^{cv}, \hat{\boldsymbol{x}}^{cv}$ in place of $\boldsymbol{v}^A, \boldsymbol{w}^A$, respectively. Consider any $i \in \{1, \ldots, n_x\}$, $\boldsymbol{p} \in P$, $t \in I'$, and $\boldsymbol{\phi}, \boldsymbol{\psi}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}} \in X^B(t)$ such that $\hat{\boldsymbol{\phi}} \leq \boldsymbol{\phi} \leq \boldsymbol{\psi} \leq \hat{\boldsymbol{\psi}}$.

Lemma 3.2 ensures Condition II.1 in in [130, Theorem 2] with state bound X^B in place of both C^A and C^B. In the proof of Theorem 3.2, we showed that, for every

$i \in \{1, \ldots, n_x\}$, if $\phi_i - \hat{\phi}_i \geq 0$ and $\hat{\psi}_i - \psi_i \geq 0$, then

$$\hat{v}_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) - \hat{v}_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}}) \leq k^r \left(\left\| \boldsymbol{\phi} - \hat{\boldsymbol{\phi}} \right\| + \left\| \boldsymbol{\psi} - \hat{\boldsymbol{\psi}} \right\| \right),$$
$$\text{and } \hat{w}_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}}) - \hat{w}_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \leq k^r \left(\left\| \boldsymbol{\phi} - \hat{\boldsymbol{\phi}} \right\| + \left\| \boldsymbol{\psi} - \hat{\boldsymbol{\psi}} \right\| \right).$$

Thus, Condition II.2. is satisfied with $\hat{\boldsymbol{v}}, \hat{\boldsymbol{w}}$ in place of $\boldsymbol{d}^{L,A}, \boldsymbol{d}^{U,A}$, respectively.

Next, suppose that $\phi_i = \hat{\phi}_i$. Assumption 3.5 provides

$$u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \geq u_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}}),$$

and $\underline{\boldsymbol{k}} \leq \underline{\boldsymbol{k}}^*$ provides

$$\dot{x}_i^L(t) - \sqrt{2\underline{k}_i(\phi_i - x_i^L(t))} \geq \dot{x}_i^L(t) - \sqrt{2\hat{\underline{k}}_i(\hat{\phi}_i - x_i^L(t))}.$$

Therefore, $v_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \geq \hat{v}_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}})$. Similarly, if $\psi_i = \hat{\psi}_i$, then $w_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \leq \hat{w}_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}})$. Condition II.3. is satisfied with $\boldsymbol{\phi}, \boldsymbol{\phi}'$ in place of $\boldsymbol{\phi}^A, \boldsymbol{\phi}^B$ and $\boldsymbol{\psi}, \boldsymbol{\psi}'$ in place of $\boldsymbol{\psi}^A, \boldsymbol{\psi}^B$, respectively.

Theorem 3.3 ensures Condition II.4. The Carathéodeory solution requirement in Condition II.5. always holds for these systems. Therefore, all conditions in [130, Theorem 2] are satisfied. □

3.5.8 Differentiability

This subsection shows that our new method can produce differentiable state relaxations. Throughout this section, continuous differentiability on closed sets is defined as in [71]. In addition to Assumptions 3.1, 3.2, and 3.3, we suppose that

the following differentiability and non-singularity conditions are satisfied.

Assumption 3.6. *$\boldsymbol{u}$ and $\boldsymbol{o}$ are continuously differentiable on their domain.*

Functions $\boldsymbol{u}, \boldsymbol{o}$ satisfying Assumption 3.2 may be constructed with DMR [72, 73, 71]. The assumption below is adapted from [57]. It ensures that small variations in the parameters will still lead to a unique solution when a discrete switch occurs in the max and min functions in (3.13).

Assumption 3.7. *Under Assumptions 3.1, 3.2, and 3.3, let $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ be a solution of (3.15) on $I' \times P$. For all $t \in I'$, $\boldsymbol{p} \in P$, and $i \in \{1, \ldots, n_x\}$, assume the following conditions hold. Arguments $(t, \boldsymbol{p}, \boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ are omitted for simplicity.*

1. *If $u_i(t, \boldsymbol{p}, \boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}) = \alpha_i(t, \boldsymbol{x}^{cv})$, then $x_i^{cv}(t, \boldsymbol{p}) \neq x_i^L(t)$, and*

$$\frac{\partial u_i}{\partial \boldsymbol{x}^{cv}}\boldsymbol{v} + \frac{\partial u_i}{\partial \boldsymbol{x}^{cc}}\boldsymbol{w} + \frac{\partial u_i}{\partial t} \neq -\frac{\underline{k}_i}{\sqrt{2\underline{k}_i(x_i^{cv}(t, \boldsymbol{p}) - x_i^L(t))}}v_i.$$

2. *If $o_i(t, \boldsymbol{p}, \boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}) = \beta_i(t, \boldsymbol{x}^{cc})$, then $x_i^{cc}(t, \boldsymbol{p}) \neq x_i^U(t)$, and*

$$\frac{\partial o_i}{\partial \boldsymbol{x}^{cv}}\boldsymbol{v} + \frac{\partial o_i}{\partial \boldsymbol{x}^{cc}}\boldsymbol{w} + \frac{\partial o_i}{\partial t} \neq -\frac{\overline{k}_i}{\sqrt{2\overline{k}_i(x_i^U(t) - x_i^{cc}(t, \boldsymbol{p}))}}w_i.$$

The following proposition is adapted from [60, Theorem 3.1, Section 3, Chapter V, p. 95].

Proposition 3.4. *Let $\boldsymbol{h} : I \times P \times \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function.*

Consider the following ODE system

$$\dot{\boldsymbol{z}}(t) = \boldsymbol{h}(t, \boldsymbol{p}, \boldsymbol{z}), \quad \forall t \in I,$$
$$\boldsymbol{z}(t_0) = \boldsymbol{z}_0.$$

Then, the solution $\boldsymbol{z}$ to the above system is continuously differentiable on $I \times P$.

To apply the sensitivity results by [57], we formulate (3.15) as a hybrid discrete/continuous system without jump discontinuities following [57, Section 2]. A hybrid state space $S = \cup_{k=1}^{n_k} S_k$ is used to describe this hybrid system, where each mode S_k corresponds to a smooth piece of those max/min functions. Since the switches between the smooth pieces in max/min functions are continuous, there are no discrete jumps in this hybrid system. Next, we use the following lemma to show that the derivatives of state relaxations with respect to parameters exist when these switches happen.

Lemma 3.7. *Under Assumptions 3.1, 3.2, 3.3, 3.6, and 3.7, let $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ be a solution of* (3.15) *on $I' \times P$. Consider any $\boldsymbol{p} \in P$, $m \in \{1, \ldots, n_p\}$, and $i \in \{1, \ldots, n_x\}$. Then, the partial derivative $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}(t, \boldsymbol{p})$ exists at every $t \in I'$ such that $u_i(t, \boldsymbol{p}, \boldsymbol{x}^{cv}(t, \boldsymbol{p}), \boldsymbol{x}^{cc}(t, \boldsymbol{p})) = \alpha_i(t, \boldsymbol{x}^{cv}(t, \boldsymbol{p}))$, and $\frac{\partial \boldsymbol{x}^{cc}}{\partial p_m}(t, \boldsymbol{p})$ exist at every $t \in I'$ such that $o_i(t, \boldsymbol{p}, \boldsymbol{x}^{cv}(t, \boldsymbol{p}), \boldsymbol{x}^{cc}(t, \boldsymbol{p})) = \beta_i(t, \boldsymbol{x}^{cc}(t, \boldsymbol{p}))$.*

Proof. Consider any $\boldsymbol{p} \in P$, $m \in \{1, \ldots, n_p\}$, and $i \in \{1, \ldots, n_x\}$. We will show that $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}(t, \boldsymbol{p})$ exists at every $t \in I'$ such that $u_i(t, \boldsymbol{p}, \boldsymbol{x}^{cv}(t, \boldsymbol{p}), \boldsymbol{x}^{cc}(t, \boldsymbol{p})) = \alpha_i(t, \boldsymbol{x}^{cc}(t, \boldsymbol{p}))$. It is analogous to show the second result.

According to Assumption 3.6, $\boldsymbol{u}$ is continuously differentiable. Assumption 3.7 and Theorem 3.3 ensure that, if $u_i(t, \boldsymbol{p}, \boldsymbol{x}^{cv}(t, \boldsymbol{p}), \boldsymbol{x}^{cc}(t, \boldsymbol{p})) = \alpha_i(t, \boldsymbol{x}^{cv}(t, \boldsymbol{p}))$, then

$x_i^{cv}(t,\boldsymbol{p}) > x_i^L(t)$. So the partial derivatives

$$\frac{\partial u_i}{\partial \boldsymbol{x}^{cv}}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc}),\ \frac{\partial u_i}{\partial p_m}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc}),\ \frac{\partial \alpha_i}{\partial \boldsymbol{x}^{cv}}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc}),\ \frac{\partial \alpha_i}{\partial p_m}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc})$$

exist and are continuous in a neighborhood of $x_i^{cv}(t,\boldsymbol{p})$ and $x_i^{cc}(t,\boldsymbol{p})$. When v_i changes between u_i and α_i, we consider the hybrid system to be moving from some mode S_j to some mode S_{j+1}. Assumption 3.7 ensures that the Jacobian matrix corresponding to $\boldsymbol{x}_j^{cv}, \boldsymbol{x}_j^{cc}, \boldsymbol{x}_{j+1}^{cv}, \boldsymbol{x}_{j+1}^{cc}, t$ is invertible. According to Theorem 1 and Remark 6 in [57], the partial derivatives in mode S_j and mode S_{j+1} are equal such that

$$\frac{\partial x_{i,j}^{cv}}{\partial p_m}(t,\boldsymbol{p}) = \frac{\partial x_{i,j+1}^{cv}}{\partial p_m}(t,\boldsymbol{p}).$$

So $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}(t,\boldsymbol{p})$ exists at every $t \in I'$ such that $u_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) = \alpha_i(t,\boldsymbol{x}^{cv}(t,\boldsymbol{p}))$.

□

Theorem 3.7. *Under Assumptions 3.1, 3.2, 3.3, 3.6, and 3.7, let* $(\boldsymbol{x}^{cv},\boldsymbol{x}^{cc})$ *be a solution of* (3.15) *on* $I' \times P$. *For any* $m \in \{1,\dots,n_p\}$, *the partial derivatives* $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}$ *and* $\frac{\partial \boldsymbol{x}^{cc}}{\partial p_m}$ *exist and are continuous on* $I' \times P$.

Proof. Consider any $m \in \{1,\dots,n_p\}$, and $\boldsymbol{p} \in P$. Let $[S_1, S_2, \dots, S_j, \dots, S_{n_k}]$ be the sequence of modes visited by a hybrid system described by (3.15). At each event time $t \in I'$ between two successive modes, Lemma 3.7 ensures that $\frac{\partial x_i^{cv}}{\partial p_m}(t,\boldsymbol{p})$ or $\frac{\partial x_i^{cc}}{\partial p_m}(t,\boldsymbol{p})$ exists, and respectively

$$\frac{\partial x_{i,j}^{cv}}{\partial p_m}(t,\boldsymbol{p}) = \frac{\partial x_{i,j+1}^{cv}}{\partial p_m}(t,\boldsymbol{p}) \quad \text{or} \quad \frac{\partial x_{i,j}^{cc}}{\partial p_m}(t,\boldsymbol{p}) = \frac{\partial x_{i,j+1}^{cc}}{\partial p_m}(t,\boldsymbol{p}).$$

Next, we consider each mode as an individual ODE system. The RHS function of the ODE is either $\boldsymbol{u}$ or $\boldsymbol{\alpha}$, and the initial condition is the terminal state of the previous mode. When $\boldsymbol{u}$ and $\boldsymbol{\alpha}$ are differentiable, Proposition 3.4 ensures the existence and continuity of $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}$ and $\frac{\partial \boldsymbol{x}^{cc}}{\partial p_m}$ within each mode. Therefore, $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}$ and $\frac{\partial \boldsymbol{x}^{cc}}{\partial p_m}$ exist and are continuous on $I' \times P$. □

In the formulation of our new method (3.15), there is no discrete jump in the RHS functions (3.14). Thus, the partial derivatives of state relaxations with respect to parameters may in principle be calculated using the approach developed in [132]. This result is beyond the scope of this work and will not be included here. Nevertheless, a demonstration of the corresponding evaluated partial derivatives will be presented in Section 3.7.

3.6 Implementation Considerations

This section discusses additional considerations that arise when implementing the new relaxation method in Section 3.4. Approaches for calculating safe-landing constants and constructing functions $\boldsymbol{u}, \boldsymbol{o}$ are introduced. To proceed, we assume that Harrison's method in Definition 3.5 is used to compute state bounds.

3.6.1 Calculating safe-landing constants

First, we discuss the calculation of safe-landing constants $\underline{\boldsymbol{k}}, \overline{\boldsymbol{k}}$ according to Definition 3.9, as required by Assumption 3.3. Following Assumption 3.4 and Lemma 3.5, we can set $\underline{\boldsymbol{k}} = \boldsymbol{k}^t + \boldsymbol{k}^L$ and $\overline{\boldsymbol{k}} = \boldsymbol{k}^t + \boldsymbol{k}^U$. Here, we introduce methods to calculate the respective Lipschitz constants $\boldsymbol{k}^t$, $\boldsymbol{k}^L$, $\boldsymbol{k}^U$ of $\dot{\boldsymbol{\xi}}(\cdot, \boldsymbol{p})$, $\dot{\boldsymbol{x}}^L(\cdot)$, $\dot{\boldsymbol{x}}^U(\cdot)$ for any $\boldsymbol{p} \in P$.

Suppose that $\boldsymbol{f}$ is directionally differentiable. Since $\boldsymbol{f}^L$ and $\boldsymbol{f}^U$ are derived from $\boldsymbol{f}$ with natural interval extension (NIE) [94], they are also directionally differentiable.

We start with $\boldsymbol{k}^t$. With $\boldsymbol{g}(t,\boldsymbol{p}) := \dot{\boldsymbol{\xi}}(t,\boldsymbol{p})$ for all $(t,\boldsymbol{p}) \in I \times P$, we are looking for a valid Lipschitz constant of $g_i(\cdot,\boldsymbol{p})$ for any $i \in \{1,\ldots,n_x\}$ and $\boldsymbol{p} \in P$. If a Lipschitz constant $k^{\boldsymbol{f}} \in \mathbb{R}_{>0}$ of $\boldsymbol{f}$ on $I' \times P \times D$ is known, then Lemma 3.8 below shows that each k_i^t can be set to $k^{\boldsymbol{f}} + \frac{M(e^{k^{\boldsymbol{f}}(t_f-t_0)}-1)}{t_f-t_0}$, where $M \geq \|\boldsymbol{f}(t,\boldsymbol{p},\boldsymbol{z})\|$ for any $t \in I, \boldsymbol{p} \in P$, and $\boldsymbol{z} \in D$.

The following proposition is adapted from Theorem 2.3, Chapter 1 by [43].

Proposition 3.5. *Suppose $\boldsymbol{h} : I \times D \to \mathbb{R}^{n_x}$ is a Lipschitz continuous function with a Lipschitz constant $k^{\boldsymbol{h}}$, and let*

$$M = \max \|\boldsymbol{h}(t,\boldsymbol{y})\|, \quad \forall (t,\boldsymbol{y}) \in I \times D.$$

Then the problem

$$\dot{\boldsymbol{y}} = \boldsymbol{h}(t,\boldsymbol{y}), \qquad \boldsymbol{y}(\tau) = \boldsymbol{y}_0$$

has a unique solution $\bar{\boldsymbol{y}}$ on I, and

$$\|\bar{\boldsymbol{y}}(t) - \boldsymbol{y}_0\| \leq \frac{M(e^{k^{\boldsymbol{h}}|t-\tau|}-1)}{k^{\boldsymbol{h}}}, \quad \forall t \in I.$$

Lemma 3.8. *Suppose that $I = [t_0, t_f]$. Let $k^{\boldsymbol{f}}$ be a Lipschitz constant of $\boldsymbol{f}$ on $I \times P \times D$ such that, for all $t_1, t_2 \in I$, $\boldsymbol{p}_1, \boldsymbol{p}_2$, and $\boldsymbol{z}_1, \boldsymbol{z}_2 \in D$,*

$$\|\boldsymbol{f}(t_1,\boldsymbol{p}_1,\boldsymbol{z}_1) - \boldsymbol{f}(t_2,\boldsymbol{p}_2,\boldsymbol{z}_2)\| \leq k^{\boldsymbol{f}}(|t_1-t_2| + \|\boldsymbol{p}_1-\boldsymbol{p}_2\| + \|\boldsymbol{z}_1-\boldsymbol{z}_2\|).$$

Moreover, suppose that there is $M \in \mathbb{R}$ *such that,* $M \geq \|\boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{z})\|$ *for any* $t \in I$, $\boldsymbol{p} \in P$, *and* $\boldsymbol{z} \in D$. *Then,* $k^{\boldsymbol{f}} + \frac{M(e^{k^{\boldsymbol{f}}|t_f - t_0|} - 1)}{t_f - t_0}$ *is a Lipschitz constant of* $\boldsymbol{g}(\cdot, \boldsymbol{p})$ *on* I *for any* $\boldsymbol{p} \in P$.

Proof. Choose an arbitrary $\boldsymbol{p} \in P$. To achieve a contradiction, suppose that there exist $t_1, t_2 \in I, t_1 \neq t_2$, such that

$$\|\boldsymbol{g}(t_1, \boldsymbol{p}) - \boldsymbol{g}(t_2, \boldsymbol{p})\| > (k^{\boldsymbol{f}} + \frac{M(e^{k^{\boldsymbol{f}}(t_f - t_0)} - 1)}{t_f - t_0})|t_1 - t_2|.$$

Since $\eta(z) := M(e^{k^{\boldsymbol{f}} z} - 1)/z$ is a monotonically increasing function, $(t_f - t_0) \geq |t_1 - t_2|$ provides that

$$\frac{M(e^{k^{\boldsymbol{f}}(t_f - t_0)} - 1)}{t_f - t_0} \geq \frac{M(e^{k^{\boldsymbol{f}}|t_1 - t_2|} - 1)}{|t_1 - t_2|}.$$

Therefore,

$$\begin{aligned} \|\boldsymbol{g}(t_1, \boldsymbol{p}) - \boldsymbol{g}(t_2, \boldsymbol{p})\| &> (k^{\boldsymbol{f}} + \frac{M(e^{k^{\boldsymbol{f}}|t_1 - t_2|} - 1)}{|t_1 - t_2|})|t_1 - t_2| \\ &= k^{\boldsymbol{f}}|t_1 - t_2| + M(e^{k^{\boldsymbol{f}}|t_1 - t_2|} - 1). \end{aligned} \tag{3.32}$$

According to Proposition 3.5,

$$k^{\boldsymbol{f}} \|\boldsymbol{x}(t_1, \boldsymbol{p}) - \boldsymbol{x}(t_2, \boldsymbol{p})\| \leq M(e^{k^{\boldsymbol{f}}|t_1 - t_2|} - 1).$$

Then, (3.32) yields

$$\begin{aligned}\|\boldsymbol{g}(t_1,\boldsymbol{p})-\boldsymbol{g}(t_2,\boldsymbol{p})\| &= \|\boldsymbol{f}(t_1,\boldsymbol{p},\boldsymbol{x}(t_1,\boldsymbol{p}))-\boldsymbol{f}(t_2,\boldsymbol{p},\boldsymbol{x}(t_2,\boldsymbol{p}))\| \\ &> k^{\boldsymbol{f}}|t_1-t_2|+k^{\boldsymbol{f}}\|\boldsymbol{x}(t_1,\boldsymbol{p})-\boldsymbol{x}(t_2,\boldsymbol{p})\| \\ &= k^{\boldsymbol{f}}\left(|t_1-t_2|+\|\boldsymbol{x}(t_1,\boldsymbol{p})-\boldsymbol{x}(t_2,\boldsymbol{p})\|\right),\end{aligned}$$

which contradicts that $k^{\boldsymbol{f}}$ is a Lipschitz constant of $\boldsymbol{f}$. □

Otherwise, if $k^{\boldsymbol{f}}$ is not available in advance, we can proceed according to the following steps. Denote some particular lower and upper bounds of $\bar{\boldsymbol{x}}^L(\cdot)$ and $\bar{\boldsymbol{x}}^U(\cdot)$ on I as $\bar{\boldsymbol{x}}^{L*}$ and $\bar{\boldsymbol{x}}^{U*}$, respectively. This bounding information can be obtained by solving the ODEs in Definition 3.5 from t_0 to t_f, and keeping the lowest and highest values of state variables $\bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U$.

The following proposition is adapted from [115, Proposition 3.1.1].

Proposition 3.6. *Let $U \subseteq \mathbb{R}^n$ be an open set, and $\boldsymbol{z}_0, \boldsymbol{z}_1 \in U$. Let function $\boldsymbol{h} : U \to \mathbb{R}^m$ be locally Lipschitz continuous and B-differentiable. The function $\boldsymbol{\psi} : [0,1] \to \mathbb{R}^m$ defined by $\boldsymbol{\psi}(t) = \boldsymbol{h}'(\boldsymbol{z}_0 + \lambda(\boldsymbol{z}_1 - \boldsymbol{z}_0); \boldsymbol{z}_1 - \boldsymbol{z}_0)$ is Lebesgue integrable and*

$$\boldsymbol{h}(\boldsymbol{z}_1) = \boldsymbol{h}(\boldsymbol{z}_0) + \int_0^1 \boldsymbol{h}'(\boldsymbol{z}_0 + \lambda(\boldsymbol{z}_1 - \boldsymbol{z}_0); \boldsymbol{z}_1 - \boldsymbol{z}_0)\,\mathrm{d}\,\lambda.$$

According to Proposition 3.6, for any $t_1, t_2 \in I$ and $\boldsymbol{p} \in P$,

$$g_i(t_2,\boldsymbol{p}) = g_i(t_1,\boldsymbol{p}) + \int_0^1 g_i'(t_1+\lambda(t_2-t_1),\boldsymbol{p};t_2-t_1,\boldsymbol{0})\,\mathrm{d}\,\lambda,$$

and so

$$|g_i(t_2, \boldsymbol{p}) - g_i(t_1, \boldsymbol{p})| \leq |t_2 - t_1| \sup_{0 \leq \lambda \leq 1} |g_i'(t_1 + \lambda(t_2 - t_1), \boldsymbol{p}; 1, \boldsymbol{0})|.$$

A sufficient condition for k_i^t being a Lipschitz constant of $g_i(\cdot, \boldsymbol{p})$ over I is

$$k_i^t \geq |g_i'(t, \boldsymbol{p}; 1, \boldsymbol{0})|, \quad \forall (t, \boldsymbol{p}) \in I \times P.$$

The following proposition is adapted from [115, Theorem 3.1.1].

Proposition 3.7. *If $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^p$ are open sets and $\pi : U \to \mathbb{R}^m$ and $\rho : V \to \mathbb{R}^n$ are continuous and B-differentiable at the points $z_0 \in V$ and $g(z_0) \in U$, respectively, then the function $\pi \circ \rho$ is B-differentiable at z_0 and*

$$(\pi \circ \rho)'(z_0; y) = \pi'(\rho(z_0); \rho'(z_0; y)).$$

According to Proposition 3.7,

$$\begin{aligned} |g_i'(t, \boldsymbol{p}; 1, 0)| &= |f_i'(t, \boldsymbol{p}, \boldsymbol{\xi}; 1, \boldsymbol{0}, \boldsymbol{0}) + f_i'(t, \boldsymbol{p}, \boldsymbol{\xi}; 0, \boldsymbol{0}, \boldsymbol{\xi}'(t, \boldsymbol{p}; 1, \boldsymbol{0}))| \\ &= |f_i'(t, \boldsymbol{p}, \boldsymbol{\xi}; 1, \boldsymbol{0}, \boldsymbol{0}) + f_i'(t, \boldsymbol{p}, \boldsymbol{\xi}; 0, \boldsymbol{0}, \boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{\xi}))|. \end{aligned} \tag{3.33}$$

Provided with the interval bounds of t, $\boldsymbol{p}$, and $\boldsymbol{\xi}$ such that $t \in I$, $\boldsymbol{p} \in P$, and $\boldsymbol{\xi}(t, \boldsymbol{p}) \in [\bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}]$ for all $(t, \boldsymbol{p}) \in I \times P$, we apply interval extensions to (3.33). Let $[\cdot]^U$ denote the upper bound of an interval extension. We may then set

$$k_i^t = [\ |f_i'(t, \boldsymbol{p}, \boldsymbol{\xi}; 1, \boldsymbol{0}, \boldsymbol{0}) + f_i'(t, \boldsymbol{p}, \boldsymbol{\xi}; 0, \boldsymbol{0}, \boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{\xi}))|\ ]^U. \tag{3.34}$$

Similar approaches can be used to compute $\boldsymbol{k}^L$ and $\boldsymbol{k}^U$, respectively. According to Proposition 3.7, and recalling the notation of Definition 3.4,

$$\begin{aligned}
&|(\dot{\bar{x}}_i^L)'(t;1)| \\
&\quad = |(\bar{f}_i^L)'(t, \bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U; 1, (\bar{\boldsymbol{x}}^L)'(t;1), (\bar{\boldsymbol{x}}^U)'(t;1))| \\
&\quad = |(\bar{f}_i^L)'(t, \bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U; 1, \bar{\boldsymbol{f}}^L(t, \bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U), \bar{\boldsymbol{f}}^U(t, \bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U))| \\
&\quad = |(f_i^L)'(t, B_i^L(\bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U); 1, \boldsymbol{f}^L(t, B_i^L(\bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U)), \boldsymbol{f}^U(t, B_i^L(\bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U)))|. \qquad (3.35)
\end{aligned}$$

Since $B_i^L(\bar{\boldsymbol{x}}^L(t), \bar{\boldsymbol{x}}^U(t))$ is a subset of $[\bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}]$ for all $t \in I$, the inclusion monotonicity of NIE [117, Theorem 2.3.11] suggests that, for each $t \in I$,

$$\begin{aligned}
&[|(f_i^L)'(t, B_i^L(\bar{\boldsymbol{x}}^L(t), \bar{\boldsymbol{x}}^U(t)); 1, \boldsymbol{f}^L(t, B_i^L(\bar{\boldsymbol{x}}^L(t), \bar{\boldsymbol{x}}^U(t))), \boldsymbol{f}^U(t, B_i^L(\bar{\boldsymbol{x}}^L(t), \bar{\boldsymbol{x}}^U(t))))|]^U \\
&\quad \leq [|(f_i^L)'(t, \bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}; 1, \boldsymbol{f}^L(t, \bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}), \boldsymbol{f}^U(t, \bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}))|]^U.
\end{aligned}$$

Thus, k_i^L can be set to

$$k_i^L = [|(f_i^L)'(t, \bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}; 1, \boldsymbol{f}^L(t, \bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}), \boldsymbol{f}^U(t, \bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}))|]^U. \qquad (3.36)$$

Similarly, we may set

$$k_i^U = [|(f_i^U)'(t, \bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}; 1, \boldsymbol{f}^L(t, \bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}), \boldsymbol{f}^U(t, \bar{\boldsymbol{x}}^{L*}, \bar{\boldsymbol{x}}^{U*}))|]^U. \qquad (3.37)$$

Setting $\underline{\boldsymbol{k}} = \boldsymbol{k}^t + \boldsymbol{k}^L$ and $\overline{\boldsymbol{k}} = \boldsymbol{k}^t + \boldsymbol{k}^U$ is only one way to satisfy Assumption 3.3, but it is broadly applicable. Theorem 3.6 shows that smaller $\underline{\boldsymbol{k}}$ and $\overline{\boldsymbol{k}}$ will help generate tighter relaxations. With additional knowledge of the original system

and the state bounds, it may indeed be possible to compute smaller safe-landing constants.

Next, we discuss the influence of state bounds on the choice of $\underline{\boldsymbol{k}}$ and $\overline{\boldsymbol{k}}$. When Harrison's method [59] is used for state bounds, its flattened RHS function has a special property, which is introduced in the following theorem.

Theorem 3.8. *Suppose that Harrison's method is used to computed the state bounds in Assumption 3.1. For any $i \in \{1, \ldots, n_x\}$, if f_i does not depend directly on its x_i argument, then Assumption 3.3 is satisfied with any $\underline{k}_i, \overline{k}_i \in \mathbb{R}^{n_x}_{>0}$.*

Proof. Consider any $i \in \{1, \ldots, n_x\}$. According to Definition 3.4, the flattening operator B_i^L only modifies the ith component of its interval argument. When $f_i(t, \boldsymbol{p}, \boldsymbol{x})$ does not depend on x_i, the flattened interval extension $\bar{f}_i^L(t, \bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U)$ is equivalent to $f_i^L(t, \bar{\boldsymbol{x}}^L, \bar{\boldsymbol{x}}^U)$, which is always a lower bound of $f_i(t, \boldsymbol{p}, \boldsymbol{x}(t, \boldsymbol{p}))$ on $I \times P$. That is, $\dot{x}_i(t, \boldsymbol{p}) - \dot{\bar{x}}_i^L(t)$ is non-negative for all $(t, \boldsymbol{p}) \in I \times P$. Hence, Assumption 3.3 is satisfied with $\underline{k}_i$ being any positive value. A similar argument holds for $\overline{k}_i$. □

Theorem 3.8 suggests that we can set $\underline{k}_i$ and $\overline{k}_i$ to be any positive real number, if $f_i(t, \boldsymbol{p}, \boldsymbol{x})$ does not involve x_i. This result is useful for reducing the computing effort involved in computing $\underline{k}_i$ and $\overline{k}_i$.

If the state bounds are not restricted to Harrison's method and satisfy $\dot{\boldsymbol{x}}^L(t) \leq \dot{\boldsymbol{x}}(t, \boldsymbol{p}) \leq \dot{\boldsymbol{x}}^U(t)$ for any $t \in I$ and $\boldsymbol{p} \in P$, then Conditions 2 and 3 in Assumption 3.3 are always satisfied. In other words, components $\underline{\boldsymbol{k}}$ and $\overline{\boldsymbol{k}}$ can be set to any positive value. One way to achieve this is to construct $\dot{\boldsymbol{x}}^L$ and $\dot{\boldsymbol{x}}^U$ from NIE of $\boldsymbol{f}$ without the flattening operation. However, this approach provides state bounds that are typically looser than Harrison's method [113].

3.6.2 Constructing functions $\boldsymbol{u}$ and $\boldsymbol{o}$

The method introduced in [120, Section 4.3] constructs continuous functions $\boldsymbol{u}, \boldsymbol{o}$ that describe bound-preserving dynamics and convexity-preserving dynamics for (3.1) as required by Assumption 3.2. This method is summarized here as follows. For all $(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \in I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, and each $i \in \{1, \ldots, n_x\}$, let

$$\begin{aligned} u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) &:= \tilde{u}_i(t, \boldsymbol{p}, B_i^L(\boldsymbol{\phi}, \boldsymbol{\psi})), \\ o_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) &:= \tilde{o}_i(t, \boldsymbol{p}, B_i^U(\boldsymbol{\phi}, \boldsymbol{\psi})), \end{aligned} \tag{3.38}$$

where $(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{o}})$ are obtained by relaxing $\boldsymbol{f}$ with generalized McCormick relaxations (GMR) [123]. It is readily verified that differentiable McCormick relaxations (DMR) [72, 73] and the optimization-based relaxation approach in [131] can be used to derive $(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{o}})$ from $\boldsymbol{f}$ as well, and they satisfy Assumption 3.6. In particular, DMR are differentiable, so they satisfy Assumption 3.6. To generate differentiable relaxations, we also require Assumption 3.7. If this assumption is not satisfied, we can slightly perturb the constants $(\underline{\boldsymbol{k}}, \overline{\boldsymbol{k}})$ to ensure it.

Next, we show that functions $\boldsymbol{u}, \boldsymbol{o}$ constructed in (3.38) with GMR satisfy Assumption 3.5, in which case our new state relaxations are tighter than Scott-Barton relaxations following Theorem 3.5. When $\phi_i = \hat{\phi}_i$, the flattening operation in Definition 3.4 makes $\hat{\phi}_i = \phi_i = \psi_i = \hat{\psi}_i$ and the others remain the same, that is $\hat{\boldsymbol{\phi}}_{(-i)} \leq \boldsymbol{\phi}_{(-i)} \leq \boldsymbol{\psi}_{(-i)} \leq \hat{\boldsymbol{\psi}}_{(-i)}$. Overall, $\hat{\boldsymbol{\phi}} \leq \boldsymbol{\phi} \leq \boldsymbol{\psi} \leq \hat{\boldsymbol{\psi}}$ when $\phi_i = \hat{\phi}_i$. The inclusion monotonicity of GMR [117, Theorem 2.4.32] implies that $u_i(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}) \geq u_i(t, \boldsymbol{p}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}})$. Thus, Assumption 3.5 is also satisfied.

3.6.3 Numerical error of square root evaluation

To numerically solve the new relaxation (3.15) with $\boldsymbol{v}, \boldsymbol{w}$ as in (3.13) and $\boldsymbol{\alpha}, \boldsymbol{\beta}$ as in (3.24), it is worth noting that evaluating the square root function in (3.24) might introduce numerical error. If the square root argument in (3.24) approaches zero, then a numerical ODE solver might provide a negative input to this expression. To avoid this behavior, the formulation in (3.14) can instead be adopted in the numerical implementation, in which case the max function will eliminate numerically negative values.

3.7 Examples

This section presents two examples to illustrate the computation of state relaxations with our new method. The first example uses constant state bounds and closed-form convex relaxations of the original ODE RHS function, so that the auxiliary ODEs in (3.15) may be solved analytically. In the second example, we use a numerical ODE solver to calculate the new state relaxations, which are shown to be tighter than the Scott-Barton relaxations. Moreover, partial derivatives of our new state relaxations are evaluated and illustrated.

Example 3.1. *Consider the parametric ODE:*

$$\dot{x}(t,p) = -x(t,p), \quad x(0,p) = (p+1)/2, \tag{3.39}$$

where $p \in [1,3]$ *and* $t \in [0,1]$.

The analytical solution of (3.39) is

$$x(t,p) = \frac{1}{2}(p+1)e^{-t}. \tag{3.40}$$

Observe that $x^L(t) = 0$ is a trivial lower state bound for (3.39). To generate a convex relaxation of $x(t_f,\cdot)$ on P, we take the following steps. Firstly, we determine appropriate Lipschitz constants k^L and k^t of $\dot{x}^L(\cdot)$ and $\dot{\xi}(\cdot,\boldsymbol{p})$, respectively, for any $p \in P$ following the approach in Section 3.6.1. Then, we set $\underline{k} = k^t + k^L$ according to Lemma 3.5. Since the lower state bound x^L is a constant, k^L can be set to 0. Based on the analytical solution (3.40), we choose a Lipschitz constant k^t to be the upper bound of $\dot{x}$ such that, for all $t \in [0,1]$ and $p \in [1,3]$,

$$k^t \geq |-\dot{x}(t,p)| = |x(t,p)|\,.$$

Since e^{-t} is monotonically decreasing for $t \in [0,1]$, it suffices to set $k^t \geq x(0,p) = \frac{p+1}{2}$ for all $p \in [1,3]$. Thus, we choose $k^t = 2$, and consequently $\underline{k} = 2$.

Next, we construct convex relaxations of the ODE RHS function and initial condition so that switches may happen between the two pieces of the max function in (3.14a). An affine function $u(t,p,\phi,\psi) = -\phi - \frac{3}{4}$ is selected as the convex relaxation of the RHS function in (3.39) and it satisfies Assumption 3.2. A convex relaxation of the original initial condition is generated as follows:

$$x^{cv}(0,p) = \begin{cases} 1, & p \leq 2, \\ (p-2)^2 + 1, & p > 2. \end{cases}$$

Since we are not interested in the concave state relaxation and function u does not depend on ψ, the concave relaxations of the ODE RHS and the initial condition are not needed here.

Finally, following the new formulation in (3.15), we obtain the ODE system describing the convex relaxation of (3.39) as follows:

$$\dot{x}^{cv}(t,p) = \max\left\{-x^{cv}(t,p) - \frac{3}{4}, -2\sqrt{x^{cv}(t,p)}\right\}, \tag{3.41}$$

$$x^{cv}(0,p) = \begin{cases} 1, & p \leq 2, \\ (p-2)^2 + 1, & p > 2. \end{cases}$$

When $x^{cv}(t,\boldsymbol{p}) = \frac{1}{4}$ or $\frac{9}{4}$, the two compared pieces in the max operation of (3.41) are equal. Thus, we obtain:

$$\dot{x}^{cv}(t,p) = \begin{cases} -x^{cv}(t,p) - \frac{3}{4}, & \frac{1}{4} \leq x^{cv}(t,p) \leq \frac{9}{4}, \\ -2\sqrt{x^{cv}(t,p)}, & 0 \leq x^{cv}(t,p) < \frac{1}{4} \text{ or } x^{cv}(t,p) > \frac{9}{4}. \end{cases}$$

When $p \leq 2$, $x^{cv}(0,p) = 1$ and $\dot{x}^{cv}(t,p) = -x^{cv} - \frac{3}{4}$. For $t > \ln(\frac{7}{4})$, the max operation in (3.41) switches to its $-2\sqrt{x^{cv}}$ piece. Hence, the analytical solution of the ODE system (3.41) is

$$x^{cv}(t,p) = \begin{cases} \frac{7}{4}e^{-t} - \frac{3}{4}, & t \leq \ln(\frac{7}{4}), \\ (t - \ln(\frac{7}{4}) - \frac{1}{2})^2, & t > \ln(\frac{7}{4}). \end{cases} \tag{3.42}$$

When $p > 2$, $x^{cv}(0,p) = (p-2)^2 + 1$ and $\dot{x}^{cv}(t,p) = -x^{cv} - \frac{3}{4}$. For $t > \ln((p-2)^2 + \frac{7}{4})$, the max function on the RHS of (3.41) switched to $-2\sqrt{x^{cv}}$. The analytical

solution of the ODE system (3.41) is

$$x^{cv}(t,p) = \begin{cases} ((p-2)^2 + \frac{7}{4})e^{-t} - \frac{3}{4}, & t \leq \ln((p-2)^2 + \frac{7}{4}), \\ (t - \ln((p-2)^2 + \frac{7}{4}) - \frac{1}{2})^2, & t > \ln((p-2)^2 + \frac{7}{4}). \end{cases} \tag{3.43}$$

This convex relaxation is plotted in Figure 3.1. This example shows that our new relaxation method does indeed generate a valid convex relaxation for the ODE system (3.39).

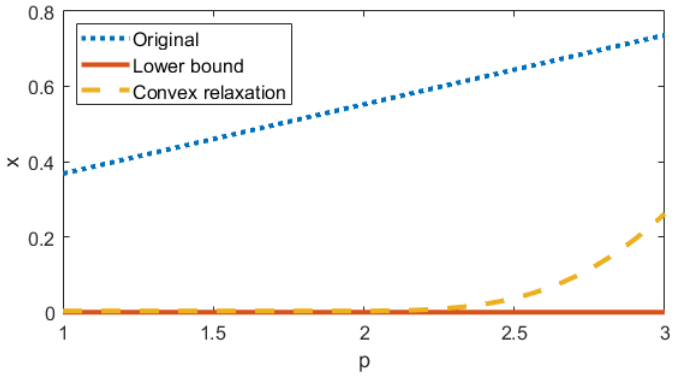


Figure 3.1: The parametric solution (3.40) of ODE (3.39) in Example 3.1, along its lower bound $x^L(t) = 0$ and convex relaxations described in (3.42) and (3.43), plotted as a function of p at $t = 1$

In a second example, we use the proposed new approach to compute state relaxations for a parametric ODE system automatically.

Example 3.2. *Consider the ODE system:*

$$\begin{aligned} \dot{x}_1(t,p) &= p\, x_2(t,p), & x_1(t_0,p) &= -1, \\ \dot{x}_2(t,p) &= -p\, x_1(t,p), & x_2(t_0,p) &= 0, \end{aligned} \tag{3.44}$$

where $p \in [0,1]$ *and* $I \equiv [t_0, t_f] = [0, \pi]$.

The analytical solution of (3.44) is

$$\begin{aligned} x_1(t,p) &= -\cos(p\,t), \\ x_2(t,p) &= \sin(p\,t). \end{aligned} \tag{3.45}$$

Since for each $p \in [0,1]$ and $t \in I$, $(p\,t) \in [0,\pi]$, constant state bounds of (3.45) may be given by $-1 \leq x_1(t,p) \leq 1$ and $0 \leq x_2(t,p) \leq 1$ for each $p \in [0,1]$ and $t \in I$. Next, we follow the approach in Section 3.6.1 to determine safe-landing constants by setting $\underline{\boldsymbol{k}} = \boldsymbol{k}^t + \boldsymbol{k}^L$ and $\overline{\boldsymbol{k}} = \boldsymbol{k}^t + \boldsymbol{k}^L$, where $\boldsymbol{k}^t$, $\boldsymbol{k}^L$, and $\boldsymbol{k}^U$ are respective Lipschitz constants of $\dot{\boldsymbol{\xi}}(\cdot, \boldsymbol{p})$, $\dot{\boldsymbol{x}}^L$, and $\dot{\boldsymbol{x}}^U$. Because the state bounds are constant, $\boldsymbol{k}^L$ and $\boldsymbol{k}^U$ can be set to $\mathbf{0}$. Next, we compute $\boldsymbol{k}^t$ using (3.34).

$$\begin{aligned} k_1^t &= [\ |p\,(-p\,x_1)|\ ]^U = 1, \\ k_2^t &= [\ |-p\,(p\,x_2)|\ ]^U = 1. \end{aligned}$$

Therefore, we set $\underline{\boldsymbol{k}} = \boldsymbol{k}^t + \boldsymbol{k}^L = \mathbf{1}$ and $\overline{\boldsymbol{k}} = \boldsymbol{k}^t + \boldsymbol{k}^U = \mathbf{1}$.

Using the implementation described in Section 3.6, state relaxations were evaluated for this system. The numerical implementation was developed in Julia [20] using DifferentialEquations.jl [104] as the ODE solver. The package McCormick.jl [134] was used to apply the GMR and DMR relaxation techniques to the ODE RHS functions. This numerical experiment was performed on a Dell Optiplex 7060 desktop with an Intel i7 CPU.

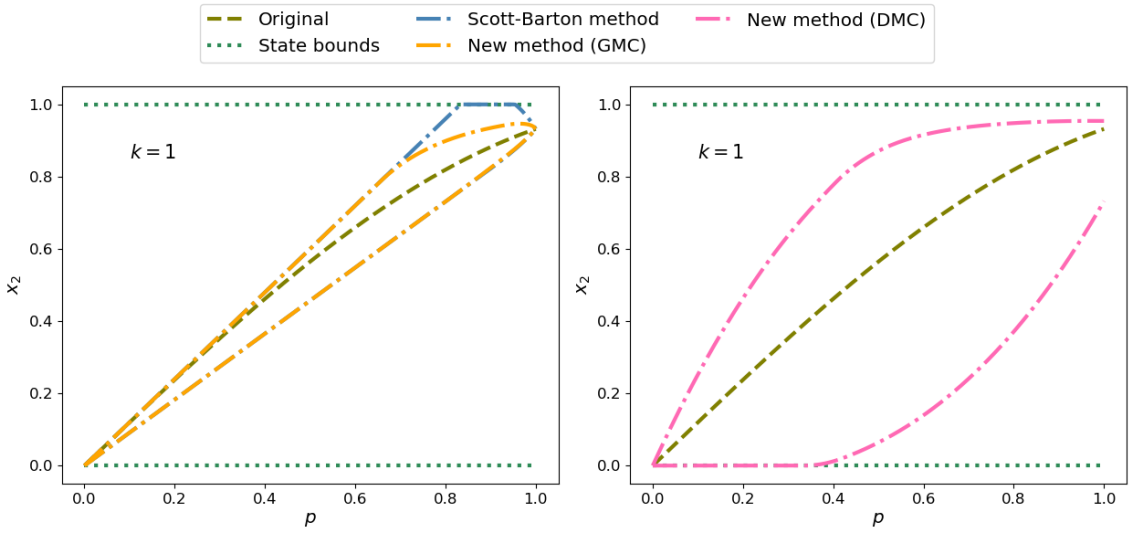


Figure 3.2: The parametric solution of x_2 in Example 3.2, along with its state bounds and state relaxations, plotted as functions of p at $t = 1.2$. (a) comparison between Scott-Barton method and new method with $\boldsymbol{u}, \boldsymbol{o}$ constructed by GMR; (b) smooth relaxations generated with new method with $\boldsymbol{u}, \boldsymbol{o}$ are constructed by DMR

Figure 3.2 depicts $x_2(t, \cdot)$, along with the corresponding state bounds and different state relaxations at $t = 1.2$. Figure 3.2(a) compares the Scott-Barton relaxations with our new relaxations, in which ODE RHS functions are relaxed using GMR. This comparison shows that the new method can generate tighter state relaxations than the Scott-Barton method as was established in Theorem 3.5. The relaxations in Figure 3.2(b) are constructed by the new method with the RHS functions relaxed by DMR, and are visibly differentiable.

In Section 3.5.8, we proposed that partial derivatives of the new state relaxations may be computed with the technique developed by [132]. Here, we demonstrate this result using Example 3.2. Figures 3.3(a) and (b) show subtangent lines constructed for the same relaxations in Example 3.2, for both nonsmooth and smooth

cases. To generate Figures 3.3(c) and (d), we increased the values of $\underline{\boldsymbol{k}}$ and $\overline{\boldsymbol{k}}$ from $\mathbf{1}$ to $\mathbf{20}$, so that part of the state relaxations overlaps with the state bonds. The putative subtangents constructed with our conjecture still appear to be valid.

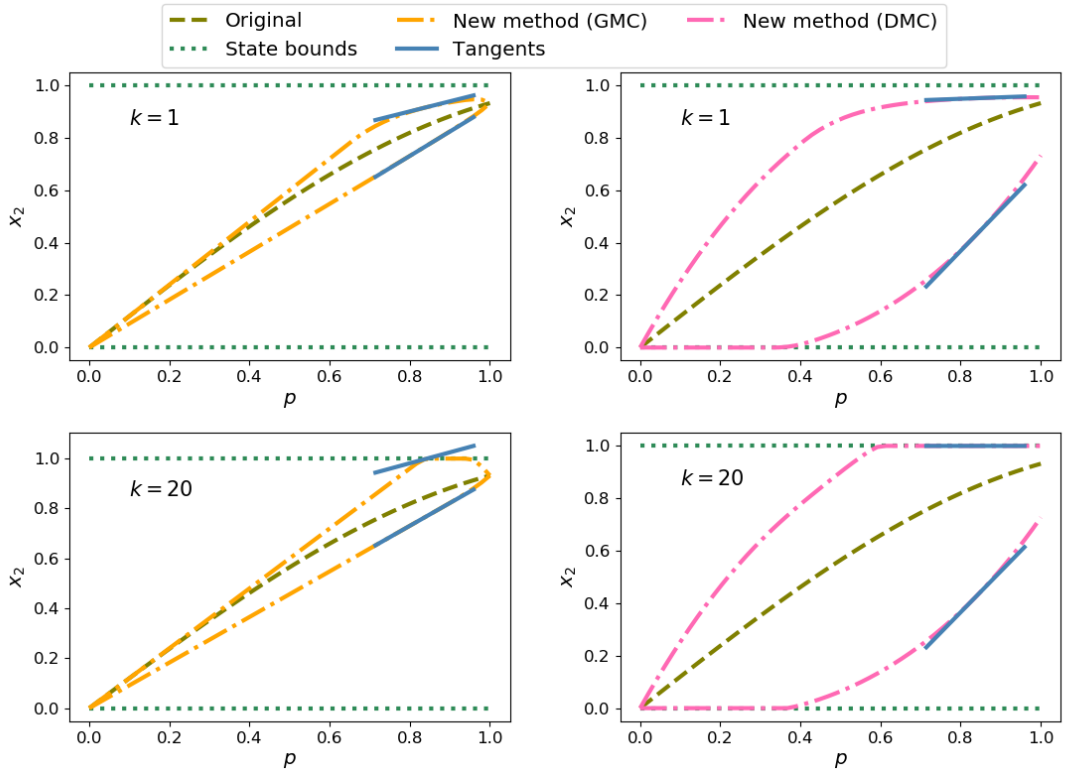


Figure 3.3: The parametric solution x_2 in Example 3.2, along with its state bounds, state relaxations, and subtangents of state relaxations, plotted as functions of p at $t = 1.2$. $k = 1$ (or 20) means that $\underline{\boldsymbol{k}}$ and $\overline{\boldsymbol{k}}$ are both set to $\mathbf{1}$ (or $\mathbf{20}$).

3.8 Conclusion

A new method was developed for enclosing the reachable set of parametric ODEs (3.1) with convex and concave relaxations described in (3.15). This new approach essentially smooths and tightens the discrete RHS jumps of an earlier relaxation

approach by [120]. Section 3.5 shows that this modification not only ensures valid state relaxations, but also provides further advantages. First of all, the auxiliary ODEs (3.15) are easier to solve numerically than the auxiliary system (3.6) from the Scott-Barton method. Secondly, the generated new state relaxations were verified in Section 3.5.7 to be as least as tight as the Scott-Barton relaxations. Such a tighter enclosure of the reachable set provides an more useful information regarding the influence of uncertainty on dynamic systems. We expect it would also improve the computational performance of branch-and-bound-based deterministic global dynamic optimization algorithms [112]. In addition to tightness, the new state relaxations were shown in Section 3.5.8 to be differentiable under additional mild assumptions. This is another useful feature for global dynamic optimization. The local optimization solvers used in a global optimization implementation typically require the functions to be differentiable. Lastly, this smoothing method permits the evaluation of partial derivatives for the state relaxations, which provide useful local sensitivity information desired by local optimization solvers. A thorough procedure for constructing the improved auxiliary system (3.15) was described in Section 3.6. A proof-of-concept numerical implementation in Julia was developed, and two examples were presented for illustration in Section 3.7.

Future work may include validating the conjecture in Section 3.5.8 about computing partial derivatives of the new state relaxations. Besides that, the current approach for determining safe-landing constants requires evaluating the original system and state bounds before constructing state relaxations. A more convenient approach for calculating safe-landing constants without this extra step is desired to make our new method easier to implement. Another possible extension of

this work is to constructing state relaxations for differential algebraic equations (DAEs).

Chapter 4

Bounding Nonconvex Optimal Control Problems using Pontryagin's Minimum Principle

This chapter is to be submitted to a journal before my anticipated thesis defense.

4.1 Introduction

This paper presents a new approach for generating guaranteed lower bounds for the solution value of a nonconvex open-loop optimal control problem (OCP) with bounded control input. Such bounding information is useful when computing the global solution of a nonconvex OCP with state-of-the-art branch-and-bound algorithms [119, 64]. The global optimization of nonconvex OCPs has been used in many engineering applications, such as the determination of optimal control inputs of batch chemical rectors [87], the nonlinear model predictive control of

continuous systems [86], and the safe landing of an autonomous spacecraft on a planet surface [2].

Branch-and-bound algorithms [42] for deterministic global optimization require the ability to compute guaranteed upper and lower bounds for the solution value of the original problem on each subset of the search space. While each upper bound may be evaluated as the objective function value at an appropriate feasible point, ideally a constrained local minimum, providing lower bounds is typically more difficult. Intuitively, since global optimization methods involve computing many such lower bounds, these lower bounds must be evaluated efficiently yet accurately. There are three main categories of approaches to generate these lower bounds for nonconvex open-loop OCPs based on how the control and state trajectories are discretized.

The first category of approaches discretize both control and state trajectories according to the discrete time mesh [143, 67]. This approach is called the *simultaneous approach*. The original OCP is then reformulated into a large-size nonlinear program (NLP), so that conventional underestimating methods in deterministic global optimization may be applied [138, 33]. If the control input values are bounded in a convex set, we may use αBB relaxations [3] or McCormick-based relaxations [92] to construct a convex relaxation problem of the resulting NLP and solve it with a local optimization solver. Due to convexity, this solution is guaranteed to be globally optimal for the convex relaxation NLP. Thus, the corresponding optimal value provides a lower bound for the optimal value of the discretized NLP that approximates the original OCP. On the other hand, if the original OCP's control input values are from a nonconvex set, we can still construct a relaxed convex

OCP for the original problem using lossless convexification [2, 22]. This approach introduces a slack variable to replace the original nonconvex set of feasible control input values with a convex set. However, there is a trade-off in this category of approaches involving how finely the control and state should be discretized. On the one hand, if a fine discretization is applied to the original problem, we will obtain an NLP with a large number of decision variables and constraints, which may be computationally expensive or impractical to solve [54]. On the other hand, if the discretization is performed over a coarse grid, then this approach will yield a poor approximation of the original continuous-time system. In this case, the optimal solution value of the convex relaxation NLP may be a poor underestimator of the original OCP.

The second category of approaches discretize only the control input trajectory into a vector of finitely many parameters, so that the original infinite-dimensional OCP is approximated as a finite-dimensional NLP with parametric ordinary differential equations (ODEs) embedded. This control-discretized dynamic optimization problem is then underestimated by applying relaxation methods for parametric ODEs. Several methods have been established in [128, 120, 131] for computing convex and concave relaxations for the parametric solutions of those embedded ODEs. However, the relaxations might be conservative when there are many parameters and state variables [64]. This relaxation conservatism may limit the applicability of a finely discretized parameterization of the control trajectory and the ability to handle dynamic systems with many state variables.

The third category underestimates an open-loop OCP without any discretization over the control or state trajectories [119]. Given an ODE with bounded control inputs, Harrison [59] proposed a method to construct componentwise interval bounds for its state variables. Then, a lower bound of the optimal solution value of the OCP can be computed by applying natural interval extension [94] to the cost function. We will call this approach *Harrison's method* throughout this article. A second approach in this category constructs a convex underestimating OCP whose optimal solution value is guaranteed to be a lower bound of the original problem's optimal solution value. In particular, Scott and Barton [119] constructed a convex underestimating OCP by relaxing the original cost function and dynamic system with generalized McCormick relaxations (GMR) [123]. They also proposed that this convex underestimating OCP can be solved to its global optimality using a gradient-based numerical method from [13]. However, this numerical method requires the functions in the OCP to be differentiable. Since GMR may generate nonsmooth convex relaxations, a more recent differentiable variant of GMR [72] might be beneficial here. To our knowledge, Scott and Barton's approach has never been implemented.

In this work, we propose a novel approach for computing lower bounds for nonconvex OCPs in the third category above, without discretizing the state or control trajectories. Our new approach essentially improves upon Scott and Barton's method [119] by constructing an underestimating OCP with a flattening operation adapted from [59, 120]. While Scott and Barton's OCP relaxation method [119] was based on their earlier relaxation approach [122] for parametric ODEs, our new approach is based on their newer superior relaxation method for ODEs [120]. These

superior ODE relaxations exhibit weakened variants of convexity compared to the prior relaxations of [122], thus introducing theoretical obstacles when extended to an optimal control setting. Hence, the theoretical development of this article is nontrivial, and proceeds quite differently to [119]. This modification leads to a lower bound that is tighter than Scott and Barton's method, which we expect will lead the branch-and-bound algorithms used in deterministic global optimization to converge faster in principle [42]. Compared with other established approaches that discretize control or state trajectories, our approach does not involve any approximation of the original OCP, and thereby avoids incurring numerical error due to discretization. It is guaranteed that the optimal solution value of our relaxed OCP is always an underestimator of the optimal solution value of the original OCP. Due to the particular structure of our relaxed OCP, Pontryagin's Minimum Principle (PMP) provides a sufficient condition [26] for determining a global optimal solution of the relaxed OCP. This enables developing numerical implementations to solve the relaxed OCP and obtain a guaranteed lower bound of the original OCP efficiently and automatically. Note that the original nonconvex OCP does not necessarily satisfy the PMP sufficient optimality conditions, so that a global lower bound cannot be obtained by applying PMP directly to the original problem.

This article is organized as follows. Section 4.2 introduces a rigorous problem formulation. The mathematical background underlying this problem is summarized in Section 4.3. Our new approach that constructs an underestimating OCP is then presented in Section 4.4. Several useful properties of this approach are established in Section 4.5, including validity and tightness. In Section 4.6, we

demonstrate that PMP provides the globally optimal solution of the underestimating OCP. Section 4.7 discusses a numerical implementation in Julia of our lower bounding approach. Several numerical examples are also presented to illustrate that our new approach constructs lower bounds for nonconvex OCPs.

The following notation conventions are used in this article. The standard Euclidean norm $\|\cdot\|$ and ℓ-infinity norm $\|\cdot\|_\infty$ are adopted for any vector space $\mathbb{R}^n$. Vectors are denoted with boldface lower-case letters (e.g. $\boldsymbol{z}$). Given vectors $\boldsymbol{z}^\dagger, \boldsymbol{z}^\ddagger \in \mathbb{R}^n$, inequalities such as $\boldsymbol{z}^\dagger < \boldsymbol{z}^\ddagger$ or $\boldsymbol{z}^\dagger \leq \boldsymbol{z}^\ddagger$, maximum operations such as $\max(\boldsymbol{z}^\dagger, \boldsymbol{z}^\ddagger)$, and minimum operations such as $\min(\boldsymbol{z}^\dagger, \boldsymbol{z}^\ddagger)$, are to be interpreted componentwise. $\boldsymbol{z}_{(-i)} \in \mathbb{R}^{n-1}$ denotes the vector $\boldsymbol{z} \in \mathbb{R}^n$ except with the ith component omitted. $\langle \boldsymbol{z}^\dagger, \boldsymbol{z}^\ddagger \rangle$ denotes the inner product of $\boldsymbol{z}^\dagger$ and $\boldsymbol{z}^\ddagger$. Throughout this article, the convexity of a vector-valued function $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ refers to convexity of all components h_i. If $\boldsymbol{h}$ is differentiable, then $\mathrm{D}_{\boldsymbol{z}}\boldsymbol{h}$ denotes the partial derivative of the function $\boldsymbol{h}$ with respect to $\boldsymbol{z}$. Dotted quantities indicate partial derivatives with respect to time t (e.g. $\dot{\boldsymbol{h}} \equiv \mathrm{D}_t\boldsymbol{h}$). The abbreviation "a.e." stands for "almost every" in the Lebesgue sense. An *interval* in $\mathbb{R}^n$ is a nonempty subset of $\mathbb{R}^n$ of the form $\{\boldsymbol{z} \in \mathbb{R}^n : \boldsymbol{z}^\mathrm{L} \leq \boldsymbol{z} \leq \boldsymbol{z}^\mathrm{U}\}$, which is denoted as an upper-case letter $Z \equiv [\boldsymbol{z}^\mathrm{L}, \boldsymbol{z}^\mathrm{U}]$. $\mathbb{IR}^n$ denotes the set of all intervals in $\mathbb{R}^n$. Lastly, $[a^i]_{i\in\{1,\ldots,n\}}$ denotes a finitely terminating sequence.

4.2 Problem Formulation

The section presents the mathematical formulation of the open-loop optimal control problem (OCP) considered in the remainder of this article. Consider scalars $t_0, t_f \in \mathbb{R}$ such that $t_0 < t_f$, and define a duration $I \equiv [t_0, t_f]$. Choose interval

domains $U \equiv [\boldsymbol{u}^{\mathrm{L}}, \boldsymbol{u}^{\mathrm{U}}] \in \mathbb{IR}^{n_u}$, and $D \in \mathbb{IR}^{n_x}$. Denote the space of all Lebesgue integrable functions that map from I into $\mathbb{R}^n$ as $\mathcal{L}^n(I)$. Let

$$\mathcal{U} \equiv \{\boldsymbol{u} \in \mathcal{L}^{n_u}(I) : \boldsymbol{u}(t) \in U,\ t \in I\}$$

be a set of admissible controls.

Assume that $\boldsymbol{f} : I \times U \times D \to \mathbb{R}^{n_x}$ and $\phi : D \to \mathbb{R}$ are *factorable* [94, 89], meaning that it is a finite composition of known simple functions such as the operations on a typical scientific calculator; see [119, Definition 2]. This structural assumption is required to generate convex relaxations for this function using McCormick relaxations [89, 123, 72]. Given an arbitrary initial state $\boldsymbol{x}_0 \in D$ and control $\boldsymbol{u} \in \mathcal{U}$, consider the following OCP in Mayer form:

$$\min_{\boldsymbol{u}\in\mathcal{U}} \quad \phi(\boldsymbol{x}(t_f, \boldsymbol{u})), \tag{4.1}$$

where $\boldsymbol{x}$ solves the following ordinary differential equation (ODE) on I:

$$\begin{aligned} \dot{\boldsymbol{x}}(t, \boldsymbol{u}) &= \boldsymbol{f}(t, \boldsymbol{u}(t), \boldsymbol{x}(t, \boldsymbol{u})), \quad t \in (t_0, t_f], \\ \boldsymbol{x}(t_0, \boldsymbol{u}) &= \boldsymbol{x}_0. \end{aligned} \tag{4.2}$$

We assume that a solution of (4.2) exists on $I \times \mathcal{U}$ and is unique. Note that the ostensibly more general Bolza OCP, formulated as (4.3) below, can be converted into Mayer form by appending an additional quadrature variable to track the running cost; see e.g. [26, Section 6.5]. The Bolza OCP is:

$$\min_{\boldsymbol{u}\in\mathcal{U}} \quad \int_{t_0}^{t_f} L(s, \boldsymbol{u}(s), \boldsymbol{x}(s, \boldsymbol{u}))\, \mathrm{d}s + \phi(\boldsymbol{x}(t_f, \boldsymbol{u})), \tag{4.3}$$

where $\boldsymbol{x}$ solves (4.2) and $L : I \times \mathcal{U} \times \mathbb{R}^{n_x} \to \mathbb{R}$ is a running cost function.

The primary goal of this work is to generate a useful lower bound for the globally optimal solution value of (4.1). We achieve this by nontrivially modifying the approach in [119] which produces a lower bound as the optimal solution value of an underestimating OCP of (4.1). Moreover, we develop and implement a new strategy that solves the underestimating OCP to global optimality using PMP.

4.3 Background

This section introduces the mathematical background underlying the methods and results in this article.

4.3.1 Convex relaxations

First, we present established terminology relating to convex relaxations of factorable functions.

Definition 4.1. *Consider an interval $S \in \mathbb{IR}^n$ and a function $\boldsymbol{h} : S \to \mathbb{R}^m$. Consider an interval-valued function $H \equiv [\boldsymbol{h}^L, \boldsymbol{h}^U] : \mathbb{IR}^n \to \mathbb{IR}^m$. Then:*

I. *H is an* inclusion function *of $\boldsymbol{h}$ on S if, for all $Z \subseteq S$, $\boldsymbol{h}(Z) \equiv \{\boldsymbol{h}(\boldsymbol{z}) : \boldsymbol{z} \in Z\} \subseteq H(Z)$.*

II. *H is* inclusion monotonic *on S if, for all $Z, Z^* \subseteq S$ such that $Z^* \subseteq Z$, $H(Z^*) \subseteq H(Z)$.*

Definition 4.2. *Consider a function $\boldsymbol{h} : \mathcal{U} \to \mathbb{R}^m$. Then:*

I. $\boldsymbol{h}^{\mathrm{L}}, \boldsymbol{h}^{\mathrm{U}} \in \mathbb{R}^m$ *are* lower *and* upper bounds *of* $\boldsymbol{h}$ *on* $\mathcal{U}$*, respectively, if* $\boldsymbol{h}^{\mathrm{L}} \leq \boldsymbol{h}(\boldsymbol{u}) \leq \boldsymbol{h}^{\mathrm{U}}$ *for all* $\boldsymbol{u} \in \mathcal{U}$*.*

II. $\boldsymbol{h}^{\mathrm{cv}}, \boldsymbol{h}^{\mathrm{cc}} : \mathcal{U} \to \mathbb{R}^m$ *are* convex *and* concave relaxations *of* $\boldsymbol{h}$ *on* $\mathcal{U}$*, respectively, if* $\boldsymbol{h}^{\mathrm{cv}}(\boldsymbol{u}) \leq \boldsymbol{h}(\boldsymbol{u}) \leq \boldsymbol{h}^{\mathrm{cc}}(\boldsymbol{u})$ *for all* $\boldsymbol{u} \in \mathcal{U}$*, and* $\boldsymbol{h}^{\mathrm{cv}}, \boldsymbol{h}^{\mathrm{cc}}$ *are respectively convex and concave on* $\mathcal{U}$*.*

Generalized McCormick relaxations (GMR) [123] and Differentiable McCormick relaxations (DMR) [72, 73] are established approaches for automatically constructing efficient convex and concave relaxations for factorable functions. In particular, DMR are continuously differentiable.

4.3.2 State relaxations

Next, we adapt two definitions from [120] that define relaxations for the open-loop ODE system (4.2) presented in Section 4.2.

Definition 4.3. *Functions* $\boldsymbol{x}^L, \boldsymbol{x}^U : I \to \mathbb{R}^{n_x}$ *are* state bounds *for the ODE* (4.2) *on* $I \times \mathcal{U}$ *if, for every* $t \in I$ *and* $\boldsymbol{u} \in \mathcal{U}$*,*

$$\boldsymbol{x}^L(t) \leq \boldsymbol{x}(t, \boldsymbol{u}) \leq \boldsymbol{x}^U(t).$$

Let $X^{\mathrm{B}} \equiv [\boldsymbol{x}^L, \boldsymbol{x}^U] : I \to \mathbb{IR}^{n_x}$ *denote the corresponding inclusion function of* $\boldsymbol{x}$ *on* $I \times \mathcal{U}$*.*

Definition 4.4. *Functions* $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc} : I \times \mathcal{U} \to \mathbb{R}^{n_x}$ *are* state relaxations *for the ODE* (4.2) *on* $I \times \mathcal{U}$*, if, for every* $t \in I$*,*

I. the mapping $\boldsymbol{u} \mapsto \boldsymbol{x}^{cv}(t, \boldsymbol{u})$ *is convex on* $\mathcal{U}$*,*

II. the mapping $\boldsymbol{u} \mapsto \boldsymbol{x}^{cc}(t,\boldsymbol{u})$ *is concave on* $\mathcal{U}$*, and*

III. $\boldsymbol{x}^{cv}(t,\boldsymbol{u}) \leq \boldsymbol{x}(t,\boldsymbol{u}) \leq \boldsymbol{x}^{cc}(t,\boldsymbol{u})$ *for all* $\boldsymbol{u} \in \mathcal{U}$*.*

Let $X^{\mathrm{R}} \equiv [\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}] : I \times \mathcal{U} \to \mathbb{IR}^{n_x}$ *denote the corresponding inclusion function of* $\boldsymbol{x}$ *on* $I \times \mathcal{U}$*.*

Scott and Barton [120] developed an approach to generate state relaxations for parametric ODEs by constructing an auxiliary system of ODEs whose right-hand side (RHS) functions are relaxations of the original ODE RHS functions. Their approach will be extended to construct state relaxations in an open-loop optimal control setting (4.2). The following results, adapted from our companion work [28], provides sufficient conditions for a system of ODEs to enclose trajectories, including solutions of parametric ODEs and ODEs with control inputs.

Definition 4.5. *Consider continuously differentiable functions* $\boldsymbol{\xi}^{\dagger}, \boldsymbol{\xi}^{\ddagger} : I \to \mathbb{R}^{n_x}$ *such that* $\boldsymbol{\xi}^{\dagger}(t) \leq \boldsymbol{\xi}^{\ddagger}(t)$ *for all* $t \in I$*. The mappings* $\boldsymbol{v}, \boldsymbol{w} : I \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ *describe* enclosing dynamics *about* $[\boldsymbol{\xi}^{\dagger}, \boldsymbol{\xi}^{\ddagger}]$ *if the following holds. For a.e.* $t \in I$*, each* $i \in \{1, \ldots, n_x\}$*, and any* $Z = [\boldsymbol{z}^L, \boldsymbol{z}^U] \in \mathbb{IR}^{n_x}$ *such that* $\boldsymbol{z}^L \leq \boldsymbol{\xi}^{\dagger}(t) \leq \boldsymbol{\xi}^{\ddagger}(t) \leq \boldsymbol{z}^U$*,*

I. If $z_i^L = \xi_i^{\dagger}(t)$*, then* $v_i(t, [\boldsymbol{z}^L, \boldsymbol{z}^U]) \leq \dot{\xi}_i^{\dagger}(t)$*.*

II. If $z_i^U = \xi_i^{\ddagger}(t)$*, then* $w_i(t, [\boldsymbol{z}^L, \boldsymbol{z}^U]) \geq \dot{\xi}_i^{\ddagger}(t)$*.*

The mappings $\boldsymbol{v}, \boldsymbol{w}$ *describe enclosing dynamics about a single trajectory* $\boldsymbol{\xi} : I \to \mathbb{R}^{n_x}$ *if the condition above holds for* $\boldsymbol{\xi}^{\dagger} \equiv \boldsymbol{\xi}^{\ddagger} \equiv \boldsymbol{\xi}$*.*

Proposition 4.1. *Consider arbitrary continuously differentiable functions* $\boldsymbol{\xi}^{\dagger}, \boldsymbol{\xi}^{\ddagger} : I \to \mathbb{R}^{n_x}$ *such that* $\boldsymbol{\xi}^{\dagger}(t) \leq \boldsymbol{\xi}^{\ddagger}(t)$ *for all* $t \in I$*. Consider quantities* $\boldsymbol{\xi}_0^{\mathrm{L}}, \boldsymbol{\xi}_0^{\mathrm{U}} \in \mathbb{R}^{n_x}$ *and continuous functions* $\boldsymbol{v}, \boldsymbol{w} : I \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$*. Let* $(\boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}})$ *solve the coupled ODE system:*

$$\begin{aligned}\dot{\boldsymbol{\xi}}^{\mathrm{L}}(t) &= \boldsymbol{v}(t,[\boldsymbol{\xi}^{\mathrm{L}}(t),\boldsymbol{\xi}^{\mathrm{U}}(t)]), \quad \boldsymbol{\xi}^{\mathrm{L}}(t_0) = \boldsymbol{\xi}_0^{\mathrm{L}},\\ \dot{\boldsymbol{\xi}}^{\mathrm{U}}(t) &= \boldsymbol{w}(t,[\boldsymbol{\xi}^{\mathrm{L}}(t),\boldsymbol{\xi}^{\mathrm{U}}(t)]), \quad \boldsymbol{\xi}^{\mathrm{U}}(t_0) = \boldsymbol{\xi}_0^{\mathrm{U}}.\end{aligned} \tag{4.4}$$

If the following conditions hold:

I. *There exists a Lipschitz constant* $k^\ell \in \mathbb{R}_{>0}$ *such that, for any* $i \in \{1,\ldots,n_x\}$, *a.e.* $t \in I$, *and any* $\boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger \in \mathbb{R}^n$ *for which* $\boldsymbol{\phi}^\ddagger \leq \boldsymbol{\phi}^\dagger \leq \boldsymbol{\psi}^\dagger \leq \boldsymbol{\psi}^\ddagger$,

$$\begin{aligned}&v_i(t,[\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger]) - v_i(t,[\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger])\\ &\quad\leq k^\ell(\|\boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger\|_\infty + \|\boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger\|_\infty),\\ &w_i(t,[\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger]) - w_i(t,[\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger])\\ &\quad\leq k^\ell(\|\boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger\|_\infty + \|\boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger\|_\infty).\end{aligned}$$

II. $\boldsymbol{v}, \boldsymbol{w}$ *describe enclosing dynamics about* $[\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger]$,

III. $\boldsymbol{\xi}_0^{\mathrm{L}} \leq \boldsymbol{\xi}^\dagger(t_0)$ *and* $\boldsymbol{\xi}^\ddagger(t_0) \leq \boldsymbol{\xi}_0^{\mathrm{U}}$,

then

$$\boldsymbol{\xi}^{\mathrm{L}}(t) \leq \boldsymbol{\xi}^\dagger(t) \leq \boldsymbol{\xi}^\ddagger(t) \leq \boldsymbol{\xi}^{\mathrm{U}}(t), \quad \forall t \in I.$$

The following definition, adapted from [120, Definition 7], describes a necessary condition for establishing convexity in the auxiliary ODEs.

Definition 4.6. *Functions* $\boldsymbol{v}, \boldsymbol{w} : I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ *describe* convexity-preserving dynamics *if, for a.e.* $t \in I$, *any* $i \in \{1,\ldots,n_x\}$, $(\lambda, \boldsymbol{p}^\dagger, \boldsymbol{p}^\ddagger) \in (0,1) \times U \times U$, $\Xi^{\mathrm{B}} \in \mathbb{IR}^{n_x}$, *and any* $\boldsymbol{\phi}^\dagger, \boldsymbol{\phi}^\ddagger, \bar{\boldsymbol{\phi}}, \boldsymbol{\psi}^\dagger, \boldsymbol{\psi}^\ddagger, \bar{\boldsymbol{\psi}} \in \Xi^{\mathrm{B}}$ *such that the following three conditions all hold:*

- $\bar{\boldsymbol{\phi}} \leq \lambda \boldsymbol{\phi}^{\dagger} + (1-\lambda)\boldsymbol{\phi}^{\ddagger}$,
- $\bar{\boldsymbol{\psi}} \geq \lambda \boldsymbol{\psi}^{\dagger} + (1-\lambda)\boldsymbol{\psi}^{\ddagger}$, *and*
- $\boldsymbol{\phi}^{\dagger} \leq \boldsymbol{\psi}^{\dagger}$, $\boldsymbol{\phi}^{\ddagger} \leq \boldsymbol{\psi}^{\ddagger}$, $\bar{\boldsymbol{\phi}} \leq \bar{\boldsymbol{\psi}}$,

$\boldsymbol{v}, \boldsymbol{w}$ *satisfy the following conditions: with* $\bar{\boldsymbol{p}} \equiv \lambda \boldsymbol{p}^{\dagger} + (1-\lambda)\boldsymbol{p}^{\ddagger}$,

I. If $\bar{\phi}_i = \lambda \phi_i^{\dagger} + (1-\lambda)\phi_i^{\ddagger}$, *then*

$$v_i(t, [\bar{\boldsymbol{p}}, \bar{\boldsymbol{p}}], [\bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}], \Xi^{\mathrm{B}}) \leq \lambda v_i(t, [\boldsymbol{p}^{\dagger}, \boldsymbol{p}^{\dagger}], [\boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}], \Xi^{\mathrm{B}}) + (1-\lambda) v_i(t, [\boldsymbol{p}^{\ddagger}, \boldsymbol{p}^{\ddagger}], [\boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}], \Xi^{\mathrm{B}}).$$

II. If $\bar{\psi}_i = \lambda \psi_i^{\dagger} + (1-\lambda)\psi_i^{\ddagger}$, *then*

$$w_i(t, [\bar{\boldsymbol{p}}, \bar{\boldsymbol{p}}], [\bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}], \Xi^{\mathrm{B}}) \geq \lambda w_i(t, [\boldsymbol{p}^{\dagger}, \boldsymbol{p}^{\dagger}], [\boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}], \Xi^{\mathrm{B}}) + (1-\lambda) w_i(t, [\boldsymbol{p}^{\ddagger}, \boldsymbol{p}^{\ddagger}], [\boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}], \Xi^{\mathrm{B}}).$$

ODE RHS functions that describe enclosing dynamics and convexity-preserving dynamics can be generated by applying *flattening operators* [120, 59] to inclusion functions of the original ODE RHS function that describe *convexity-amplifying dynamics* [120]. Such inclusion functions may be constructed with GMR [120] and DMR [72].

Definition 4.7. *For each* $i \in \{1, \ldots, n\}$, *define* flattening operators $B_i^L, B_i^U : \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ *such that*

I. $B_i^L([\boldsymbol{\phi}, \boldsymbol{\psi}]) = [\boldsymbol{\phi}, \boldsymbol{\psi}']$ *where* $\psi_i' := \phi_i$, $\boldsymbol{\psi}'_{(-i)} := \boldsymbol{\psi}_{(-i)}$,

II. $B_i^U([\boldsymbol{\phi},\boldsymbol{\psi}]) = [\boldsymbol{\phi}',\boldsymbol{\psi}]$ *where* $\phi_i' := \psi_i$, $\boldsymbol{\phi}'_{(-i)} := \boldsymbol{\phi}_{(-i)}$.

Definition 4.8. *Functions* $\boldsymbol{v},\boldsymbol{w} : I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ *describe* convexity-amplifying dynamics *if, for a.e.* $t \in I$, *any* $i \in \{1,\ldots,n_x\}$, $(\lambda,\boldsymbol{p}^\dagger,\boldsymbol{p}^\ddagger) \in (0,1) \times P \times P$, $\Xi^{\rm B} \in \mathbb{IR}^{n_x}$, *and any* $\boldsymbol{\phi}^\dagger,\boldsymbol{\phi}^\ddagger,\bar{\boldsymbol{\phi}},\boldsymbol{\psi}^\dagger,\boldsymbol{\psi}^\ddagger,\bar{\boldsymbol{\psi}} \in \Xi^{\rm B}$ *such that the following three conditions all hold:*

I. $\bar{\boldsymbol{\phi}} \leq \lambda\boldsymbol{\phi}^\dagger + (1-\lambda)\boldsymbol{\phi}^\ddagger$,

II. $\bar{\boldsymbol{\psi}} \geq \lambda\boldsymbol{\psi}^\dagger + (1-\lambda)\boldsymbol{\psi}^\ddagger$, *and*

III. $\boldsymbol{\phi}^\dagger \leq \boldsymbol{\psi}^\dagger$, $\boldsymbol{\phi}^\ddagger \leq \boldsymbol{\psi}^\ddagger$, $\bar{\boldsymbol{\phi}} \leq \bar{\boldsymbol{\psi}}$,

$\boldsymbol{v}$ *and* $\boldsymbol{w}$ *satisfy the following conditions: with* $\bar{\boldsymbol{p}} \equiv \lambda\boldsymbol{p}^\dagger + (1-\lambda)\boldsymbol{p}^\ddagger$,

$$\begin{aligned} v_i(t,[\bar{\boldsymbol{p}},\bar{\boldsymbol{p}}],[\bar{\boldsymbol{\phi}},\bar{\boldsymbol{\psi}}],\Xi^{\rm B}) &\leq \lambda v_i(t,[\boldsymbol{p}^\dagger,\boldsymbol{p}^\dagger],[\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger],\Xi^{\rm B}) \\ &\quad + (1-\lambda)v_i(t,[\boldsymbol{p}^\ddagger,\boldsymbol{p}^\ddagger],[\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger],\Xi^{\rm B}), \\ w_i(t,[\bar{\boldsymbol{p}},\bar{\boldsymbol{p}}],[\bar{\boldsymbol{\phi}},\bar{\boldsymbol{\psi}}],\Xi^{\rm B}) &\geq \lambda w_i(t,[\boldsymbol{p}^\dagger,\boldsymbol{p}^\dagger],[\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger],\Xi^{\rm B}) \\ &\quad + (1-\lambda)w_i(t,[\boldsymbol{p}^\ddagger,\boldsymbol{p}^\ddagger],[\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger],\Xi^{\rm B}). \end{aligned}$$

4.3.3 Scott and Barton's OCP relaxation method

Consider the problem formulation of Section 4.2, and assume that $\hat{X}^{\rm B} \equiv [\hat{\boldsymbol{x}}^{\rm L},\hat{\boldsymbol{x}}^{\rm U}]$ are state bounds of (4.2) on $I \times \mathcal{U}$. Consider functions $\hat{\phi}^{\rm cv} : \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{R}$ and $\hat{\boldsymbol{f}}^{\rm cv}, \hat{\boldsymbol{f}}^{\rm cc} : I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ such that the following conditions hold for any convex and concave relaxations $\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc} : \mathcal{U} \to \mathbb{R}^{n_x}$ of $\boldsymbol{u} \mapsto \boldsymbol{x}(t,\boldsymbol{u})$ on $\mathcal{U}$:

I. The mapping $\boldsymbol{u} \mapsto \hat{\phi}^{\rm cv}([\boldsymbol{\xi}^{cv}(\boldsymbol{u}),\boldsymbol{\xi}^{cc}(\boldsymbol{u})],\hat{X}^{\rm B}(t_f))$ is a convex relaxation of $\boldsymbol{u} \mapsto \phi(\boldsymbol{x}(t_f,\boldsymbol{u}))$ on $\mathcal{U}$,

II. The mappings $\boldsymbol{u} \mapsto \hat{\boldsymbol{f}}^{\mathrm{cv}}(t, [\boldsymbol{u}, \boldsymbol{u}], [\boldsymbol{\xi}^{cv}(\boldsymbol{u}), \boldsymbol{\xi}^{cc}(\boldsymbol{u})], \hat{X}^{\mathrm{B}}(t))$ and $\boldsymbol{u} \mapsto \hat{\boldsymbol{f}}^{\mathrm{cc}}(t, [\boldsymbol{u}, \boldsymbol{u}], [\boldsymbol{\xi}^{cv}(\boldsymbol{u}), \boldsymbol{\xi}^{cc}(\boldsymbol{u})], \hat{X}^{\mathrm{B}}(t))$ are convex and concave relaxations of $\boldsymbol{u} \mapsto \boldsymbol{f}(t, \boldsymbol{u}, \boldsymbol{x}(t, \boldsymbol{u})))$ on $\mathcal{U}$, respectively, for all $t \in I$.

Functions $\hat{\phi}^{\mathrm{cv}}$ and $\hat{\boldsymbol{f}}^{\mathrm{cv}}$, $\hat{\boldsymbol{f}}^{\mathrm{cc}}$ satisfying above conditions can be generated with GMR [119, Section II]. Then, Scott and Barton [119] constructed the following convex OCP whose optimal solution value underestimates the optimal solution value of the original OCP (4.1).

$$\min_{\boldsymbol{u} \in U} \quad \hat{\phi}^{\mathrm{cv}}([\hat{\boldsymbol{x}}^{\mathrm{cv}}(t_f, \boldsymbol{u}), \hat{\boldsymbol{x}}^{\mathrm{cc}}(t_f, \boldsymbol{u})], [\hat{\boldsymbol{x}}^{\mathrm{L}}(t_f), \hat{\boldsymbol{x}}^{\mathrm{U}}(t_f)]), \tag{4.5}$$

where $\hat{X}^{\mathrm{R}} \equiv [\hat{\boldsymbol{x}}^{\mathrm{cv}}, \hat{\boldsymbol{x}}^{\mathrm{cc}}]$ solve the following ODE:

$$\begin{aligned} \dot{\hat{\boldsymbol{x}}}^{\mathrm{cv}}(t, \boldsymbol{u}) &= \hat{\boldsymbol{f}}^{\mathrm{cv}}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], \hat{X}^{\mathrm{R}}(t, \boldsymbol{u}), \hat{X}^{\mathrm{B}}(t)), \quad \hat{\boldsymbol{x}}^{\mathrm{cc}}(t_0) = \boldsymbol{x}_0, \\ \dot{\hat{\boldsymbol{x}}}^{\mathrm{cc}}(t, \boldsymbol{u}) &= \hat{\boldsymbol{f}}^{\mathrm{cc}}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], \hat{X}^{\mathrm{R}}(t, \boldsymbol{u}), \hat{X}^{\mathrm{B}}(t)), \quad \hat{\boldsymbol{x}}^{\mathrm{cc}}(t_0) = \boldsymbol{x}_0. \end{aligned} \tag{4.6}$$

To solve the above convex OCP to global optimality, Scott and Barton suggested using an approach by Azhmyakov and Raisch [13] which involves gradient methods and proximal point techniques. This approach requires the mappings $\boldsymbol{u} \mapsto \hat{\boldsymbol{x}}^{\mathrm{cv}}(t_f, \boldsymbol{u})$, $\boldsymbol{u} \mapsto \hat{\boldsymbol{x}}^{\mathrm{cc}}(t_f, \boldsymbol{u})$, and $(\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}) \mapsto \hat{\phi}^{\mathrm{cv}}([\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}], [\hat{\boldsymbol{x}}^{\mathrm{L}}(t_f), \hat{\boldsymbol{x}}^{\mathrm{U}}(t_f)])$ to be differentiable [13, Theorem 6]. However, the multivariate relaxation rules in GMR are typically nonsmooth, and this differentiability requirement limits which OCPs can actually be bounded using Scott and Barton's method. Nevertheless, we can overcome this obstacle by constructing $\hat{\phi}^{\mathrm{cv}}$ and $\hat{\boldsymbol{f}}^{\mathrm{cv}}$, $\hat{\boldsymbol{f}}^{\mathrm{cc}}$ with DMR in place of GMR. Then, the previous mappings $\boldsymbol{u} \mapsto \hat{\boldsymbol{x}}^{\mathrm{cv}}(t_f, \boldsymbol{u})$, $\boldsymbol{u} \mapsto \hat{\boldsymbol{x}}^{\mathrm{cc}}(t_f, \boldsymbol{u})$, and $(\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}) \mapsto \hat{\phi}^{\mathrm{cv}}([\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}], [\hat{\boldsymbol{x}}^{\mathrm{L}}(t_f), \hat{\boldsymbol{x}}^{\mathrm{U}}(t_f)])$ become continuously differentiable and the OCP (4.5)

can then be approached using the approach in [13]. Hence, DMR will be used to construct convex relaxations for Scott and Barton's method in this work, instead of GMR.

4.3.4 Pontryagin's Minimum Principle

This subsection summarizes the well-known necessary and sufficient conditions for solving OCPs with PMP [26]. Consider a continuous function $\boldsymbol{h} : I \times U \times \mathbb{R}^{n_y} \to \mathbb{R}^{n_y}$ that is continuously differentiable with respect to $\boldsymbol{y}$. Given an initial state $\boldsymbol{y}_0 \in \mathbb{R}^{n_y}$ and a differentiable function $\psi : \mathbb{R}^{n_y} \to \mathbb{R}$, consider the following Mayer problem:

$$\begin{aligned} \min_{\boldsymbol{u}\in\mathcal{U}} \quad & \psi(\boldsymbol{y}(t_f, \boldsymbol{u})), \\ \text{s.t.} \quad & \dot{\boldsymbol{y}} = \boldsymbol{h}(t, \boldsymbol{u}(t), \boldsymbol{y}(t)), \quad t \in (t_0, t_f], \\ & \boldsymbol{y}(t_0) = \boldsymbol{y}_0. \end{aligned} \tag{4.7}$$

The following proposition is a variant of the Pontryagin's Maximum Principle, as described in [26, Theorem 6.1.1]. This variant is tailored to address the minimization problem in (4.7) following [26, Theorem 6.3.1 and Remark 6.3].

Proposition 4.2. *Let $\boldsymbol{u}^*$ be an optimal solution of* (4.7)*, and let $\boldsymbol{y}^*$ be the corresponding optimal trajectory of the ODE embedded in* (4.7)*. Let $\boldsymbol{\mu} : I \to (\mathbb{R}^{n_y})^\top$ be the row-vector-valued solution of the adjoint equation*

$$\begin{aligned} \dot{\boldsymbol{\mu}}(t) &= -\boldsymbol{\mu}(t)\, \mathrm{D}_{\boldsymbol{y}} \boldsymbol{h}(t, \boldsymbol{u}^*(t), \boldsymbol{y}(t, \boldsymbol{u}^*)), \\ \boldsymbol{\mu}(t_f) &= \mathrm{D}\psi(\boldsymbol{y}(t_f, \boldsymbol{u}^*)), \end{aligned} \tag{4.8}$$

where $\boldsymbol{y}, \boldsymbol{h}$ are column vectors. Then, for a.e. $t \in I$,

$$\boldsymbol{u}^*(t) \in \underset{\boldsymbol{\omega} \in U}{\arg\min} \langle \boldsymbol{\mu}(t),\, \boldsymbol{h}(t, \boldsymbol{\omega}, \boldsymbol{y}^*) \rangle. \tag{4.9}$$

The necessary optimality condition (4.9) could also be satisfied by suboptimal local solutions of the OCP (4.7). However, the following proposition, adapted from [26, Theorem 7.2.1], presents a sufficient condition for global optimality of (4.7).

Proposition 4.3. *Suppose that $\mathrm{D}_{\boldsymbol{u}}\boldsymbol{h}$ is continuous. Suppose that the mapping $\boldsymbol{u} \mapsto \psi(\boldsymbol{y}(t_f, \boldsymbol{u}))$ from $\mathcal{U}$ into $\mathbb{R}$ is convex. Then, any mapping $\boldsymbol{u}^*$ that satisfies* (4.9) *in Proposition 4.2 is a globally optimal solution of the Mayer problem* (4.7)*.*

4.4 New optimal control relaxation

Consider the problem formulation in Section 4.2, this section constructs a novel relaxed OCP of the original OCP (4.1).

Assumption 4.1. *Suppose that $\boldsymbol{x}$ solves* (4.2)*. Suppose that $\boldsymbol{x}^L, \boldsymbol{x}^U$ are state bounds of $\boldsymbol{x}$ on $I \times \mathcal{U}$, and $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$ are state relaxations of $\boldsymbol{x}$ on $I \times \mathcal{U}$. Let $X^{\mathrm{B}} \equiv [\boldsymbol{x}^L, \boldsymbol{x}^U]$ and $X^{\mathrm{R}} \equiv [\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}]$ be interval-valued functions. Suppose that a function $\phi^{\mathrm{cv}} : \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{R}$ satisfies the following conditions:*

I. The mapping $\boldsymbol{u} \mapsto \phi^{\mathrm{cv}}([\boldsymbol{x}^{cv}(t_f, \boldsymbol{u}), \boldsymbol{x}^{cc}(t_f, \boldsymbol{u})], [\boldsymbol{x}^L(t_f), \boldsymbol{x}^U(t_f)])$ is a convex relaxation of $\boldsymbol{u} \mapsto \phi(\boldsymbol{x}(t_f, \boldsymbol{u}))$ on $\mathcal{U}$,

II. The mapping $(\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}) \mapsto \phi^{\mathrm{cv}}([\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}], \Xi^{\mathrm{B}})$ is continuously differentiable for all $\Xi^{\mathrm{B}} \in \mathbb{IR}^{n_x}$.

Note that DMR [72] will construct a function ϕ^{cv} that satisfies this assumption.

Assumption 4.2. *Assume that functions $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc} : I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ satisfy the following conditions for all $\Xi^{\text{B}} \in \mathbb{IR}^{n_x}$:*

I. There exist a Lipschitz constant $k^r \in \mathbb{R}_{>0}$ such that, for any $(t, \boldsymbol{p}) \in I \times U$ and any $\Xi^{\dagger} \equiv [\boldsymbol{\xi}^{L\dagger}, \boldsymbol{\xi}^{U\dagger}], \Xi^{\ddagger} \equiv [\boldsymbol{\xi}^{L\ddagger}, \boldsymbol{\xi}^{U\ddagger}], \bar{\Xi}^{\dagger} \equiv [\bar{\boldsymbol{\xi}}^{L\dagger}, \bar{\boldsymbol{\xi}}^{U\dagger}], \bar{\Xi}^{\ddagger} \equiv [\bar{\boldsymbol{\xi}}^{L\ddagger}, \bar{\boldsymbol{\xi}}^{U\ddagger}] \in \mathbb{IR}^{n_x}$,

$$\begin{aligned}
&\| \boldsymbol{f}^{cv}(t, [\boldsymbol{p}, \boldsymbol{p}], \Xi^{\dagger}, \Xi^{\ddagger}) - \boldsymbol{f}^{cv}(t, [\boldsymbol{p}, \boldsymbol{p}], \bar{\Xi}^{\dagger}, \bar{\Xi}^{\ddagger}) \|_{\infty} \\
&\quad + \| \boldsymbol{f}^{cc}(t, [\boldsymbol{p}, \boldsymbol{p}], \Xi^{\dagger}, \Xi^{\ddagger}) - \boldsymbol{f}^{cc}(t, [\boldsymbol{p}, \boldsymbol{p}], \bar{\Xi}^{\dagger}, \bar{\Xi}^{\ddagger}) \|_{\infty} \\
&\quad \leq k^r \left(\| \boldsymbol{\xi}^{L\dagger} - \bar{\boldsymbol{\xi}}^{L\dagger} \|_{\infty} + \| \boldsymbol{\xi}^{U\dagger} - \bar{\boldsymbol{\xi}}^{U\dagger} \|_{\infty} + \| \boldsymbol{\xi}^{L\ddagger} - \bar{\boldsymbol{\xi}}^{L\ddagger} \|_{\infty} + \| \boldsymbol{\xi}^{U\ddagger} - \bar{\boldsymbol{\xi}}^{U\ddagger} \|_{\infty} \right),
\end{aligned}$$

II. The mappings $(\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}) \mapsto \boldsymbol{f}^{cv}(t, [\boldsymbol{p}, \boldsymbol{p}], [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}], \Xi^{\text{B}})$ and $(\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}) \mapsto \boldsymbol{f}^{cc}(t, [\boldsymbol{p}, \boldsymbol{p}], [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}], \Xi^{\text{B}}$ are continuously differentiable on $\Xi^{\text{B}} \times \Xi^{\text{B}}$ for all $(t, \boldsymbol{p}) \in I \times U$,

III. The mappings $(t, [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}]) \mapsto \boldsymbol{f}^{cv}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}], \Xi^{\text{B}})$ and $(t, [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}]) \mapsto \boldsymbol{f}^{cc}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}], \Xi^{\text{B}})$ describe enclosing dynamics about $t \mapsto \boldsymbol{x}(t, \boldsymbol{u})$ for all $\boldsymbol{u} \in \mathcal{U}$,

IV. $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ describe convexity-preserving dynamics,

V. For a.e. $t \in I$, any $P \subseteq \widehat{P} \subseteq U$ and any $Z \equiv [\boldsymbol{z}^L, \boldsymbol{z}^U] \subseteq \widehat{Z} \equiv [\widehat{\boldsymbol{z}}^{\text{L}}, \widehat{\boldsymbol{z}}^{\text{U}}] \in \Xi^{\text{B}}$,

(a) If $z_i^{\text{L}} = \widehat{z}_i^{\text{L}}$, then $f_i^{cv}(t, \widehat{P}, \widehat{Z}, \Xi^{\text{B}}) \leq f_i^{cv}(t, P, Z, \Xi^{\text{B}})$,

(b) If $z_i^{\text{U}} = \widehat{z}_i^{\text{U}}$, then $f_i^{cc}(t, \widehat{P}, \widehat{Z}, \Xi^{\text{B}}) \geq f_i^{cc}(t, P, Z, \Xi^{\text{B}})$.

Under Assumptions 4.1 and 4.2, our new relaxed OCP of the original OCP (4.1)

is as follows:

$$\min_{\boldsymbol{u}\in U} \quad \phi^{\mathrm{cv}}([\boldsymbol{x}^{cv}(t_f,\boldsymbol{u}),\boldsymbol{x}^{cv}(t_f,\boldsymbol{u})],[\boldsymbol{x}^{L}(t_f),\boldsymbol{x}^{U}(t_f)]), \tag{4.10}$$

where $(\boldsymbol{x}^{cv},\boldsymbol{x}^{cv},\boldsymbol{x}^{L},\boldsymbol{x}^{U})$ solves the coupled ODE system:

$$\begin{aligned} \dot{\boldsymbol{x}}^{L}(t) &= \boldsymbol{f}^{cv}(t,U,X^{\mathrm{B}}(t),X^{\mathrm{B}}(t)), & \boldsymbol{x}^{L}(t_0) &= \boldsymbol{x}_0, \\ \dot{\boldsymbol{x}}^{U}(t) &= \boldsymbol{f}^{cc}(t,U,X^{\mathrm{B}}(t),X^{\mathrm{B}}(t)), & \boldsymbol{x}^{U}(t_0) &= \boldsymbol{x}_0, \end{aligned} \tag{4.11a}$$

$$\begin{aligned} \dot{\boldsymbol{x}}^{cv}(t,\boldsymbol{u}) &= \boldsymbol{f}^{cv}(t,[\boldsymbol{u}(t),\boldsymbol{u}(t)],X^{\mathrm{R}}(t,\boldsymbol{u}),X^{\mathrm{B}}(t)), \quad \boldsymbol{x}^{cv}(t_0) = \boldsymbol{x}_0 \\ \dot{\boldsymbol{x}}^{cc}(t,\boldsymbol{u}) &= \boldsymbol{f}^{cc}(t,[\boldsymbol{u}(t),\boldsymbol{u}(t)],X^{\mathrm{R}}(t,\boldsymbol{u}),X^{\mathrm{B}}(t)), \quad \boldsymbol{x}^{cc}(t_0) = \boldsymbol{x}_0. \end{aligned} \tag{4.11b}$$

The new relaxed OCP in (4.10) is the main contribution of this article. Its useful properties, including validity and tightness, are illustrated in Section 4.5. This new formulation differs from Scott and Barton's formulation (4.5) in three aspects.

First, our new approach described above places a weaker requirement on the auxiliary ODE RHS functions than Scott and Barton's method. Specifically, $\hat{\boldsymbol{f}}^{\mathrm{cv}}$, $\hat{\boldsymbol{f}}^{\mathrm{cc}}$ in (4.6) need to be convex and concave relaxations of the original ODE RHS function $\boldsymbol{f}$ on $I\times\mathcal{U}\times\hat{X}^{\mathrm{B}}(t)$ for all $t\in I$, while our new formulation only requires $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ in (4.11) to describe enclosing dynamics and convexity-preserving dynamics. This permits constructing $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ with tighter but nonconvex relaxations of $\boldsymbol{f}$, and one valid technique is introduced Section 4.4.1. Theorem 4.2 in Section 4.5.5 illustrates that tighter auxiliary ODE RHS functions lead to tighter state relaxations, and then provide tighter lower bounds for the original OCP (4.1). Furthermore, even though $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ are not necessarily convex, the auxiliary ODE solutions of (4.11) have been shown to be convex after nontrivial theoretical development in

Section 4.5.3. This implies that our new relaxed OCP (4.10) can be solved to its global optimality using PMP as illustrated in Section 4.6. Note that the original OCP does not satisfy the PMP sufficient optimality conditions, so that PMP cannot be applied to (4.1) directly to obtain a global lower bound.

Second, while Scott and Barton's method depends on state bounds that are known in advance, our new approach constructs state bounds and state relaxations simultaneously. It is also illustrated in Section 4.5.2 that the state relaxations generated in our new approach are always at least as tight as state bounds. This property is fundamental to constructing auxiliary ODE RHS function using GMR [123] and DMR [72]. Scott and Barton did not address this in (4.6), but their later work [120] may be adapted to deal with it.

4.4.1 Constructing ODE RHS functions

This section presents an approach to construct the auxiliary ODE RHS functions $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ in (4.11) by applying the flatten operators in Definition 4.7 to DMR. As discussed above, these generated functions are tighter than DMR but are not convex. It will be verified that they satisfy Assumption 4.2.

Assumption 4.3. *Assume that an interval-valued function* $\bar{F} \equiv [\bar{\boldsymbol{f}}^{cv}, \bar{\boldsymbol{f}}^{cc}] : I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ *satisfies the following conditions for all* $\Xi^{\mathrm{B}} \in \mathbb{IR}^{n_x}$*:*

I. There exist a Lipschitz constant $\bar{k}^r \in \mathbb{R}_{>0}$ *such that, for any* $(t, \boldsymbol{p}) \in I \times U$ *and any*

$$\Xi^\dagger \equiv [\boldsymbol{\xi}^{L\dagger}, \boldsymbol{\xi}^{U\dagger}], \Xi^\ddagger \equiv [\boldsymbol{\xi}^{L\ddagger}, \boldsymbol{\xi}^{U\ddagger}], \bar{\Xi}^\dagger \equiv [\bar{\boldsymbol{\xi}}^{L\dagger}, \bar{\boldsymbol{\xi}}^{U\dagger}], \bar{\Xi}^\ddagger \equiv [\bar{\boldsymbol{\xi}}^{L\ddagger}, \bar{\boldsymbol{\xi}}^{U\ddagger}] \in \mathbb{IR}^{n_x},$$

$$\begin{aligned}
&\|\bar{\boldsymbol{f}}^{cv}(t,[\boldsymbol{p},\boldsymbol{p}],\Xi^\dagger,\Xi^\ddagger) - \bar{\boldsymbol{f}}^{cv}(t,[\boldsymbol{p},\boldsymbol{p}],\bar{\Xi}^\dagger,\bar{\Xi}^\ddagger)\|_\infty \\
&\quad + \|\bar{\boldsymbol{f}}^{cc}(t,[\boldsymbol{p},\boldsymbol{p}],\Xi^\dagger,\Xi^\ddagger) - \bar{\boldsymbol{f}}^{cc}(t,[\boldsymbol{p},\boldsymbol{p}],\bar{\Xi}^\dagger,\bar{\Xi}^\ddagger)\|_\infty \\
&\quad \leq \bar{k}^r \left(\|\boldsymbol{\xi}^{L\dagger} - \bar{\boldsymbol{\xi}}^{L\dagger}\|_\infty + \|\boldsymbol{\xi}^{U\dagger} - \bar{\boldsymbol{\xi}}^{U\dagger}\|_\infty + \|\boldsymbol{\xi}^{L\ddagger} - \bar{\boldsymbol{\xi}}^{L\ddagger}\|_\infty + \|\boldsymbol{\xi}^{U\ddagger} - \bar{\boldsymbol{\xi}}^{U\ddagger}\|_\infty \right),
\end{aligned}$$

II. $(\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}) \mapsto \bar{\boldsymbol{f}}^{cv}(t,[\boldsymbol{p},\boldsymbol{p}],[\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc}],\Xi^{\mathrm{B}})$ *and* $(\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}) \mapsto \bar{\boldsymbol{f}}^{cc}(t,[\boldsymbol{p},\boldsymbol{p}],[\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc}],\Xi^{\mathrm{B}})$ *are continuously differentiable on* Ξ^{B} *for all* $(t,\boldsymbol{p}) \in I \times U$,

III. $([\boldsymbol{p}^{\mathrm{L}}, \boldsymbol{p}^{\mathrm{U}}], [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}]) \mapsto \bar{F}(t,[\boldsymbol{p}^{\mathrm{L}}, \boldsymbol{p}^{\mathrm{U}}],[\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc}],\Xi^{\mathrm{B}})$ *is an inclusion function of* $(\boldsymbol{p},\boldsymbol{\xi}) \mapsto \boldsymbol{f}(t,\boldsymbol{p},\boldsymbol{\xi})$ *on* $U \times \Xi^{\mathrm{B}}$ *for a.e.* $t \in I$,

IV. $\bar{\boldsymbol{f}}^{cv}, \bar{\boldsymbol{f}}^{cc}$ *describe convexity-amplifying dynamics,*

V. $([\boldsymbol{p}^{\mathrm{L}}, \boldsymbol{p}^{\mathrm{U}}], [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}]) \mapsto \bar{F}(t,[\boldsymbol{p}^{\mathrm{L}}, \boldsymbol{p}^{\mathrm{U}}],[\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc}],\Xi^{\mathrm{B}})$ *is inclusion monotonic on* $U \times \Xi^{\mathrm{B}}$ *for a.e.* $t \in I$.

Under this assumption, consider $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ such that, for each $t \in I$, $i \in \{1,\ldots,n_x\}$, $P \in \mathbb{IR}^{n_u}$, and $\Xi^{\mathrm{R}}, \Xi^{\mathrm{B}} \in \mathbb{IR}^{n_x}$,

$$\begin{aligned}
f_i^{cv}(t,P,\Xi^{\mathrm{R}},\Xi^{\mathrm{B}}) &= \bar{f}_i^{\mathrm{cv}}(t,P,B_i^L(\Xi^{\mathrm{R}}),\Xi^{\mathrm{B}}), \\
f_i^{cc}(t,P,\Xi^{\mathrm{R}},\Xi^{\mathrm{B}}) &= \bar{f}_i^{\mathrm{cc}}(t,P,B_i^U(\Xi^{\mathrm{R}}),\Xi^{\mathrm{B}}).
\end{aligned} \tag{4.12}$$

The following result confirms that applying the flattening operators B_i^L, B_i^U in Definition 4.7 to DMR provides a valid approach to construct the auxiliary ODE RHS functions in (4.11).

Lemma 4.1. *Under Assumption 4.3, the functions* $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ *defined in* (4.12) *satisfy Assumption 4.2.*

Proof. We will shown that $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ satisfy all conditions in Assumption 4.2. Consider a.e. $t \in I$, any $i \in \{1, \ldots, n_x\}$, $\boldsymbol{u} \in \mathcal{U}$, and any $\Xi^{\mathrm{B}}, \Xi^{\mathrm{R}} \in \mathbb{IR}^{n_x}$ such that $\boldsymbol{x}(t, \boldsymbol{u}) \in \Xi^{\mathrm{R}} \equiv [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}] \subseteq \Xi^{\mathrm{B}}$.

Since the flattening operators B_i^L, B_i^U are linear, Conditions I. and II. in Assumption 4.3 guarantee Conditions I. and II. of Assumption 4.2.

Next, we verify the enclosing dynamics in Condition III. of Assumption 4.2. According to Definition 4.5, this is equivalent to showing that,

1. If $\xi_i^{cv} = \dot{x}_i(t, \boldsymbol{u})$, then $f_i^{cv}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], \Xi^{\mathrm{R}}, \Xi^{\mathrm{B}}) \leq \dot{x}_i(t, \boldsymbol{u})$,
2. If $\xi_i^{cc} = \dot{x}_i(t, \boldsymbol{u})$, then $f_i^{cc}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], \Xi^{\mathrm{R}}, \Xi^{\mathrm{B}}) \geq \dot{x}_i(t, \boldsymbol{u})$.

It will be shown that the first requirement holds; verifying the second is analogous. If $\xi_i^{cv} = \dot{x}_i(t, \boldsymbol{u})$, then the flattening operator B_i^L from Definition 4.7 ensures that $\boldsymbol{x}(t, \boldsymbol{u}) \in B_i^L(\Xi^{\mathrm{R}})$. Condition III. in Assumption 4.3 shows that

$$\begin{aligned} \bar{f}_i^{\mathrm{cv}}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], B_i^L(\Xi^{\mathrm{R}}), \Xi^{\mathrm{B}}) &\leq f_i(t, \boldsymbol{u}(t), \boldsymbol{x}(t, \boldsymbol{u})) \\ &= \dot{x}_i(t, \boldsymbol{u}), \end{aligned}$$

which ensures the first requirement. Hence, Condition III. in Assumption 4.2 is verified.

Lemma 11 from [120] shows that, if $\bar{\boldsymbol{f}}^{\mathrm{cv}}, \bar{\boldsymbol{f}}^{\mathrm{cc}}$ describe convexity-amplifying dynamics, then $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ describe convexity-preserving dynamics. Therefore, Condition IV. ensures Condition IV. in Assumption 4.2.

Lastly, we verify Condition V. in Assumption 4.2. Consider any $P \subseteq \widehat{P} \in \mathbb{IR}^{n_u}$ and $Z \equiv [\boldsymbol{z}^L, \boldsymbol{z}^U] \subseteq \widehat{Z} \equiv [\widehat{\boldsymbol{z}}^{\mathrm{L}}, \widehat{\boldsymbol{z}}^{\mathrm{U}}] \in \Xi^{\mathrm{B}}$. According to Definition 4.5, this is equivalent to showing that,

1. If $z_i^{\mathrm{L}} = \widehat{z}_i^{\mathrm{L}}$, then $f_i^{cv}(t, \widehat{P}, \widehat{Z}, \Xi^{\mathrm{B}}) \leq f_i^{cv}(t, P, Z, \Xi^{\mathrm{B}})$,

2. If $z_i^{\mathrm{U}} = \widehat{z}_i^{\mathrm{U}}$, then $f_i^{cc}(t, \widehat{P}, \widehat{Z}, \Xi^{\mathrm{B}}) \geq f_i^{cc}(t, P, Z, \Xi^{\mathrm{B}})$.

We will show that the first requirement holds; showing the second is analogous. When $z_i^{\mathrm{L}} = \widehat{z}_i^{\mathrm{L}}$ and $Z \subseteq \widehat{Z}$, the flattening operation ensures that $B_i^L(Z) \subseteq B_i^L(\widehat{Z})$. Combined with $P \subseteq \widehat{P}$, Condition V. in Assumption 4.3 shows that

$$\bar{f}_i^{\mathrm{cv}}(t, \widehat{P}, B_i^L(\widehat{Z}), \Xi^{\mathrm{B}}) \leq \bar{f}_i^{\mathrm{cv}}(t, P, B_i^L(Z), \Xi^{\mathrm{B}}).$$

which is equivalent to the inequality in the first requirement. Condition V. in Assumption 4.2 is verified.

Thus, all conditions in Assumption 4.2 are indeed satisfied. □

Lemma 4.2. *Functions $\bar{\boldsymbol{f}}^{cv}$, $\bar{\boldsymbol{f}}^{cc}$ satisfying Assumption 4.3 can be generated using DMR.*

Proof. We will show that all conditions in Assumption 4.3 hold.

First, since DMR are continuously differentiable with respect to their convex and concave relaxation inputs [72], Condition II. holds. Next, the interval bounds used in DMR are typically computed with natural interval extension [94], which is locally Lipschitz continuous [113]. Since the composition of locally Lipschitz continuous functions is also locally Lipschitz continuous [117], DMR is locally Lipschitz continuous with respect to its relaxation inputs and interval bound inputs. The global Lipschitz continuity required in Condition I. is satisfied with appropriate Lipschitz extensions [120, 131]. Next, the inclusion properties required in Conditions III. and V. can be verified using similar arguments as in [117, Section 2.4]. Lastly, the convexity-amplifying dynamics required in Condition IV. can be confirmed by adapting [120, Lemma 9].

Thus, $\bar{\boldsymbol{f}}^{cv}, \bar{\boldsymbol{f}}^{cc}$ generated using DMR satisfy all conditions in Assumption 4.3.

□

4.5 Theoretical Development

This section develops the useful theoretical properties of our new formulations in (4.10)-(4.11), including existence and uniqueness, bounding properties, convexity, as well as tightness in comparison to Scott and Barton's method [119]. Note that this development process is completely different from Scott and Barton's proof [119] showing that (4.5) provides valid lower bounds for the original OCP (4.1). Scott and Barton's proof used successive approximations (also known as Picard Iteration), which is a standard construction in ODE theory [43]. But this technique cannot be applied here. The reason is that, unlike $\hat{\boldsymbol{f}}^{\text{cv}}, \hat{\boldsymbol{f}}^{\text{cc}}$ in (4.6), $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ in our new formulation (4.11) are not convex and concave relaxations of the original RHS functions $\boldsymbol{f}$. Instead, our proofs in this section heavily relies on the theory of differential inequalities [148, 120], which is a totally different technique from successive approximations.

4.5.1 Existence and uniqueness of a solution

Lemma 4.3. *Under Assumption 4.2,* (4.11) *has a unique solution.*

Proof. Condition I. in Assumption 4.2 ensures that $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ are Lipschitz continuous with respect to $\boldsymbol{x}^L, \boldsymbol{x}^U$ in (4.11a) and with respect to $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}$ in (4.11b). Therefore, (4.11) has a unique solution following the Picard-Lindelöf Theorem as summarized in [60, Theorem 1.1, Chapter II]. □

4.5.2 Bounding properties

Lemma 4.4. *Under Assumption 4.2, let* $(\boldsymbol{x}^L, \boldsymbol{x}^U, \boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ *be a solution of* (4.11) *on* $I \times \mathcal{U}$*. Then, the following results are true:*

I. $\boldsymbol{x}^L, \boldsymbol{x}^U$ *are state bounds of* (4.2) *on* $I \times \mathcal{U}$*,*

II. $\boldsymbol{x}^{cv}(t, \boldsymbol{u}) \leq \boldsymbol{x}(t, \boldsymbol{u}) \leq \boldsymbol{x}^{cc}(t, \boldsymbol{u})$ *for all* $(t, \boldsymbol{u}) \in I \times \mathcal{U}$*,*

III. $\boldsymbol{x}^L(t) \leq \boldsymbol{x}^{cv}(t, \boldsymbol{u}) \leq \boldsymbol{x}^{cc}(t, \boldsymbol{u}) \leq \boldsymbol{x}^U(t)$ *for all* $(t, \boldsymbol{u}) \in I \times \mathcal{U}$*.*

Proof. Result I. has been verified in [28, Theorem 3]. Consider any $\boldsymbol{u} \in \mathcal{U}$. Result II. can be verified by showing that all three requirements in Proposition 4.1 are satisfied with $(t \mapsto \boldsymbol{x}(t, \boldsymbol{u}), t \mapsto \boldsymbol{x}(t, \boldsymbol{u}))$ in place of $(\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger)$, $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ in place of $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U)$, and $((t, \Xi^{\mathrm{R}}) \mapsto \boldsymbol{f}^{cv}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], \Xi^{\mathrm{R}}, X^{\mathrm{B}}(t)), (t, \Xi^{\mathrm{R}}) \mapsto \boldsymbol{f}^{cc}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], \Xi^{\mathrm{R}}, X^{\mathrm{B}}(t)))$ in place of $(\boldsymbol{v}, \boldsymbol{w})$. Condition I. in Assumption 4.2 ensures Requirement I. of Proposition 4.1. Condition III. in Assumption 4.2 guarantees Requirement II. of Proposition 4.1. Requirement III. of Proposition 4.1 is ensured by the construction of the initial conditions in (4.11). Hence, all three requirements in Proposition 4.1 are satisfied.

Similarly, Result III. will be demonstrated by showing that all three requirements in Proposition 4.1 are satisfied with $(t \mapsto \boldsymbol{x}^{cv}(t, \boldsymbol{u}), t \mapsto \boldsymbol{x}^{cc}(t, \boldsymbol{u}))$ in place of $(\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger)$, $(\boldsymbol{x}^L, \boldsymbol{x}^U)$ in place of $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U)$, and $((t, \Xi^{\mathrm{B}}) \mapsto \boldsymbol{f}^{cv}(t, U, \Xi^{\mathrm{B}}, \Xi^{\mathrm{B}}), (t, \Xi^{\mathrm{B}}) \mapsto \boldsymbol{f}^{cc}(t, U, \Xi^{\mathrm{B}}, \Xi^{\mathrm{B}}))$ in place of $(\boldsymbol{v}, \boldsymbol{w})$. Condition I. in Assumption 4.2 ensures Requirement I. of Proposition 4.1. Requirement III. of Proposition 4.1 is guaranteed by the construction of initial conditions in (4.11). Next, we verify that the mappings $(t, \Xi^{\mathrm{B}}) \mapsto \boldsymbol{f}^{cv}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], \Xi^{\mathrm{B}}, \Xi^{\mathrm{B}})$ and $(t, \Xi^{\mathrm{B}}) \mapsto \boldsymbol{f}^{cc}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], \Xi^{\mathrm{B}}, \Xi^{\mathrm{B}})$ describe enclosing dynamics about $[t \mapsto \boldsymbol{x}^{cv}(t, \boldsymbol{u}), t \mapsto \boldsymbol{x}^{cc}(t, \boldsymbol{u})]$ for all $\boldsymbol{u} \in \mathcal{U}$, so that

Requirement II. of Proposition 4.1 is satisfied. It suffices to show that, for a.e. $t \in I$, any $\boldsymbol{u} \in \mathcal{U}$, and $Z = [\boldsymbol{z}^L, \boldsymbol{z}^U] \in \mathbb{IR}^{n_x}$ such that $X^{\mathrm{R}}(t, \boldsymbol{u}) \equiv [\boldsymbol{x}^{cv}(t, \boldsymbol{u}), \boldsymbol{x}^{cc}(t, \boldsymbol{u})] \subseteq Z$,

i. If $z_i^L = x_i^{cv}(t, \boldsymbol{u})$, then $f_i^{cv}(t, U, Z, Z) \leq \dot{x}_i^{cv}(t, \boldsymbol{u})$.

ii. If $z_i^U = x_i^{cc}(t, \boldsymbol{u})$, then $f_i^{cc}(t, U, Z, Z) \geq \dot{x}_i^{cc}(t, \boldsymbol{u})$.

It will be shown that the first requirement holds; verifying the second is analogous. Since $z_i^L = x_i^{cv}(t, \boldsymbol{u})$, $X^{\mathrm{R}}(t, \boldsymbol{u}) \subseteq Z$, and $[\boldsymbol{u}(t), \boldsymbol{u}(t)] \subseteq U$, Condition V. in Assumption 4.2 shows that,

$$\begin{aligned} f_i^{cv}(t, U, Z, Z) &\leq f_i^{cv}(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], X^{\mathrm{R}}(t, \boldsymbol{u}), Z) \\ &= \dot{x}_i^{cv}(t, \boldsymbol{u}), \end{aligned}$$

which is the desired inequality. □

4.5.3 Convexity

Lemma 4.5. *Under Assumption 4.2, suppose that* $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}, \boldsymbol{x}^L, \boldsymbol{x}^U)$ *solves the ODE* (4.11) *on* $I \times \mathcal{U}$. *Then,* $\boldsymbol{u} \mapsto \boldsymbol{x}^{cv}(t, \boldsymbol{u})$ *and* $\boldsymbol{u} \mapsto \boldsymbol{x}^{cc}(t, \boldsymbol{u})$ *are convex and concave, respectively, on* $\mathcal{U}$ *for every* $t \in I$.

Proof. We proceed very similarly to the proof of [120, Theorem 3]. Choose any fixed $\boldsymbol{u}^\dagger, \boldsymbol{u}^\ddagger \in \mathcal{U}$ and $\lambda \in (0, 1)$. For all $t \in I$, define

$$\begin{aligned} \bar{\boldsymbol{u}}(t) &\equiv \lambda \boldsymbol{u}^\dagger(t) + (1 - \lambda)\boldsymbol{u}^\ddagger(t), \\ \bar{\boldsymbol{x}}^{cv}(t) &\equiv \lambda \boldsymbol{x}^{cv}(t, \boldsymbol{u}^\dagger) + (1 - \lambda)\boldsymbol{x}^{cv}(t, \boldsymbol{u}^\ddagger), \\ \bar{\boldsymbol{x}}^{cc}(t) &\equiv \lambda \boldsymbol{x}^{cc}(t, \boldsymbol{u}^\dagger) + (1 - \lambda)\boldsymbol{x}^{cc}(t, \boldsymbol{u}^\ddagger). \end{aligned}$$

To achieve a contradiction, assume that there exists $\hat{t} \in I$ such that either $x_i^{\mathrm{cv}}(\hat{t}, \bar{\boldsymbol{u}}) > \bar{x}_i^{\mathrm{cv}}(\hat{t})$ or $x_i^{\mathrm{cc}}(\hat{t}, \bar{\boldsymbol{u}}) < \bar{x}_i^{\mathrm{cc}}(\hat{t})$ for at least one index $i \in \{1, \ldots, n_x\}$. Define $\boldsymbol{\delta} : I \to \mathbb{R}^{2n_x}$ by

$$\boldsymbol{\delta}(t) \equiv (\boldsymbol{x}^{cv}(t, \bar{\boldsymbol{u}}) - \bar{\boldsymbol{x}}^{cv}(t),\ \bar{\boldsymbol{x}}^{cc}(t) - \boldsymbol{x}^{cc}(t, \bar{\boldsymbol{u}})), \quad \forall t \in I.$$

Then, there is $\delta_i(\hat{t}) > 0$ or $\delta_{i+n_x}(\hat{t}) > 0$. Let $k^r \in \mathbb{R}_{>0}$ be the Lipschitz constant in Condition I. in Assumption 4.2. According to Lemma 3 in [120], there exist $j \in \{1, \ldots, 2n_x\}$, $t_1, t_2 \in I$ with $t_1 < t_2$, and a continuously differentiable function $\rho : I \to \mathbb{R}$ satisfying

$$0 < \rho(t) \quad \text{and} \quad \dot{\rho}(t) > (2k^r)\rho(t), \quad \forall t \in I.$$

Suppose that $j \leq n_x$; the case in which $j > n_x$ is analogous. The following inequalities then hold.

$$\boldsymbol{x}^{cv}(t, \bar{\boldsymbol{u}}) \leq \bar{\boldsymbol{x}}^{cv}(t) + \mathbf{1}\rho(t), \quad \forall t \in [t_1, t_2), \tag{4.13a}$$

$$\boldsymbol{x}^{cc}(t, \bar{\boldsymbol{u}}) \geq \bar{\boldsymbol{x}}^{cc}(t) - \mathbf{1}\rho(t), \quad \forall t \in [t_1, t_2), \tag{4.13b}$$

$$\bar{x}_j^{\mathrm{cv}}(t) < x_j^{\mathrm{cv}}(t, \bar{\boldsymbol{u}}) < \bar{x}_j^{\mathrm{cv}}(t) + \rho(t), \quad \forall t \in (t_1, t_2), \tag{4.13c}$$

$$x_j^{\mathrm{cv}}(t_2, \bar{\boldsymbol{u}}) = \bar{x}_j^{\mathrm{cv}}(t_2) + \rho(t_2), \tag{4.13d}$$

$$x_j^{\mathrm{cv}}(t_1, \bar{\boldsymbol{u}}) = \bar{x}_j^{\mathrm{cv}}(t_1), \tag{4.13e}$$

where $\mathbf{1}$ is a vector whose elements are 1.

Define $\boldsymbol{x}^{cv\dagger}(t) \equiv \min(\boldsymbol{x}^{cv}(t, \bar{\boldsymbol{u}}),\ \bar{\boldsymbol{x}}^{cv}(t))$ and $\boldsymbol{x}^{cc\dagger}(t) \equiv \max(\boldsymbol{x}^{cc}(t, \bar{\boldsymbol{u}}),\ \bar{\boldsymbol{x}}^{cc}(t))$ for

all $t \in [t_1, t_2]$. The Lipschitz continuity of $\boldsymbol{f}^{cv}$ and $\boldsymbol{f}^{cc}$ implies that

$$\begin{aligned}\dot{x}_i^{\mathrm{cv}}(t, \bar{\boldsymbol{u}}) =& f_j^{\mathrm{cv}}(t, [\bar{\boldsymbol{u}}(t), \bar{\boldsymbol{u}}(t)], [\boldsymbol{x}^{\mathrm{cv}}(t, \bar{\boldsymbol{u}}), \boldsymbol{x}^{\mathrm{cc}}(t, \bar{\boldsymbol{u}})], X^{\mathrm{B}}(t)) \\ \leq& f_j^{\mathrm{cv}}(t, [\bar{\boldsymbol{u}}(t), \bar{\boldsymbol{u}}(t)], [\boldsymbol{x}^{cv\dagger}(t), \boldsymbol{x}^{cc\dagger}(t)], X^{\mathrm{B}}(t)) \\ &+ k^r(\|\boldsymbol{x}^{cv}(t, \bar{\boldsymbol{u}}) - \boldsymbol{x}^{cv\dagger}(t)\|_\infty \\ &+ \|\boldsymbol{x}^{cc}(t, \bar{\boldsymbol{u}}) - \boldsymbol{x}^{cc\dagger}(t)\|_\infty),\end{aligned}$$

for a.e. $t \in [t_1, t_2]$. By the inequalities in (4.13a) and (4.13b), it follows that

$$\begin{aligned}\dot{x}_i^{\mathrm{cv}}(t, \bar{\boldsymbol{u}}) &\leq f_j^{\mathrm{cv}}(t, [\bar{\boldsymbol{u}}(t), \bar{\boldsymbol{u}}(t)], [\boldsymbol{x}^{cv\dagger}(t), \boldsymbol{x}^{cc\dagger}(t)], X^{\mathrm{B}}(t)) + 2k^r\rho(t) \\ &\leq f_j^{\mathrm{cv}}(t, [\bar{\boldsymbol{u}}(t), \bar{\boldsymbol{u}}(t)], [\boldsymbol{x}^{cv\dagger}(t), \boldsymbol{x}^{cc\dagger}(t)], X^{\mathrm{B}}(t)) + \dot{\rho}(t),\end{aligned}$$

for a.e. $t \in [t_1, t_2]$.

Next, following Definition 4.6, we use the fact that $\boldsymbol{f}^{cv}$ and $\boldsymbol{f}^{cc}$ describe convexity-preserving dynamics to show that, for a.e. $t \in [t_1, t_2]$,

$$\begin{aligned}&\dot{x}_j^{\mathrm{cv}}(t, \bar{\boldsymbol{u}}) \\ &\quad\leq f_j^{\mathrm{cv}}(t, [\bar{\boldsymbol{u}}(t), \bar{\boldsymbol{u}}(t)], [\boldsymbol{x}^{cv\dagger}(t), \boldsymbol{x}^{cc\dagger}(t)], X^{\mathrm{B}}(t)) + \dot{\rho}(t) \\ &\quad\leq \lambda f_j^{\mathrm{cv}}(t, [\boldsymbol{u}^\dagger(t), \boldsymbol{u}^\dagger(t)], [\boldsymbol{x}^{cv}(t, \boldsymbol{u}^\dagger), \boldsymbol{x}^{cc}(t, \boldsymbol{u}^\dagger)], X^{\mathrm{B}}(t)) \\ &\qquad+ (1-\lambda) f_j^{\mathrm{cv}}(t, [\boldsymbol{u}^\ddagger(t), \boldsymbol{u}^\ddagger(t)], [\boldsymbol{x}^{cv}(t, \boldsymbol{u}^\ddagger), \boldsymbol{x}^{cc}(t, \boldsymbol{u}^\ddagger)], X^{\mathrm{B}}(t)) \\ &\qquad+ \dot{\rho}(t) \\ &\quad= \lambda \dot{x}_j^{\mathrm{cv}}(t, \boldsymbol{u}^\dagger) + (1-\lambda)\dot{x}_j^{\mathrm{cv}}(t, \boldsymbol{u}^\ddagger) + \dot{\rho}(t). \qquad (4.14)\end{aligned}$$

First, it is assured by Lemma 4.4 that, for a.e. $t \in [t_1, t_2]$,

$$\begin{aligned} \boldsymbol{x}^{cv}(t, \boldsymbol{u}^\dagger) &\leq \boldsymbol{x}^{cc}(t, \boldsymbol{u}^\dagger), \\ \boldsymbol{x}^{cv}(t, \boldsymbol{u}^\ddagger) &\leq \boldsymbol{x}^{cc}(t, \boldsymbol{u}^\ddagger), \\ \boldsymbol{x}^{cv\dagger}(t) \leq \boldsymbol{x}^{cv}(t, \bar{\boldsymbol{u}}) &\leq \boldsymbol{x}^{cc}(t, \bar{\boldsymbol{u}}) \leq \boldsymbol{x}^{cc\dagger}(t). \end{aligned}$$

Moreover, by definition,

$$\begin{aligned} \boldsymbol{x}^{cv\dagger}(t) \leq \bar{\boldsymbol{x}}^{cv}(t) &= \lambda \boldsymbol{x}^{cv}(t, \boldsymbol{u}^\dagger) + (1-\lambda)\boldsymbol{x}^{cv}(t, \boldsymbol{u}^\ddagger), \\ \boldsymbol{x}^{cc\dagger}(t) \geq \bar{\boldsymbol{x}}^{cc}(t) &= \lambda \boldsymbol{x}^{cc}(t, \boldsymbol{u}^\dagger) + (1-\lambda)\boldsymbol{x}^{cc}(t, \boldsymbol{u}^\ddagger), \end{aligned}$$

and the inequality in (4.13c) shows that, for a.e. $t \in [t_1, t_2]$,

$$x_j^{cv\dagger}(t) = \bar{x}_j^{\mathrm{cv}}(t) = \lambda x_j^{\mathrm{cv}}(t, \boldsymbol{u}^\dagger) + (1-\lambda) x_j^{\mathrm{cv}}(t, \boldsymbol{u}^\ddagger).$$

Thus, Definition 4.6 ensures (4.14) with the following substitutions:

$$\begin{aligned} &\boldsymbol{p}^\dagger \equiv \boldsymbol{u}^\dagger(t) && \boldsymbol{p}^\ddagger \equiv \boldsymbol{u}^\ddagger(t) && \bar{\boldsymbol{p}} \equiv \bar{\boldsymbol{u}}(t), \\ &\boldsymbol{\phi}^\dagger \equiv \boldsymbol{x}^{cv}(t, \boldsymbol{u}^\dagger), && \boldsymbol{\phi}^\ddagger \equiv \boldsymbol{x}^{cv}(t, \boldsymbol{u}^\ddagger), && \bar{\boldsymbol{\phi}} \equiv \boldsymbol{x}^{cv\dagger}(t), \\ &\boldsymbol{\psi}^\dagger \equiv \boldsymbol{x}^{cc}(t, \boldsymbol{u}^\dagger), && \boldsymbol{\psi}^\ddagger \equiv \boldsymbol{x}^{cc}(t, \boldsymbol{u}^\ddagger), && \bar{\boldsymbol{\psi}} \equiv \boldsymbol{x}^{cc\dagger}(t). \end{aligned}$$

According to Theorem 1 in [120], (4.14) implies that $\dot{x}_j^{\mathrm{cv}}(t, \bar{\boldsymbol{u}}) - \dot{\bar{x}}_j^{\mathrm{cv}}(t) - \dot{\rho}(t)$ is non-increasing on $[t_1, t_2]$. So,

$$x_j^{\mathrm{cv}}(t_2, \bar{\boldsymbol{u}}) - \bar{x}_j^{\mathrm{cv}}(t_2) - \rho(t_2) \leq x_j^{\mathrm{cv}}(t_1, \bar{\boldsymbol{u}}) - \bar{x}_j^{\mathrm{cv}}(t_1) - \rho(t_1).$$

The inequalities in (4.13d) and (4.13e) suggest that $0 \geq \rho(t_1)$, which is a contradiction. □

4.5.4 Underestimating the original OCP

This section uses the results in Sections 4.5.1-4.5.3 to illustrate that, the optimal solution value of OCP (4.10) is a guaranteed lower bound of the optimal solution value of the original OCP (4.1).

Theorem 4.1. *Under Assumptions 4.1 and 4.2, the following results hold:*

I. The following mapping is convex:

$$\boldsymbol{u} \mapsto \phi^{\mathrm{cv}}([\boldsymbol{x}^{cv}(t_f, \boldsymbol{u}), \boldsymbol{x}^{cc}(t_f, \boldsymbol{u})], [\boldsymbol{x}^{L}(t_f), \boldsymbol{x}^{U}(t_f)]),$$

II. The optimal solution value of OCP (4.10) *is a lower bound of the optimal solution value of OCP* (4.1).

Proof. Under Assumption 4.2, Lemmas 4.3, 4.4, and 4.5 ensure that $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}, \boldsymbol{x}^{L}, \boldsymbol{x}^{U})$ is a unique solution of (4.11) and provides state relaxations and state bounds of (4.2). Assumption 4.1 shows that

$$\boldsymbol{u} \mapsto \phi^{\mathrm{cv}}([\boldsymbol{x}^{cv}(t_f, \boldsymbol{u}), \boldsymbol{x}^{cc}(t_f, \boldsymbol{u})], [\boldsymbol{x}^{L}(t_f), \boldsymbol{x}^{U}(t_f)]) \tag{4.15}$$

is a convex relaxation of

$$\boldsymbol{u} \mapsto \phi(\boldsymbol{x}(t_f, \boldsymbol{u})) \tag{4.16}$$

on $\mathcal{U}$. Thus, Result I. holds.

Next, let $\boldsymbol{u}^* \in \mathcal{U}$ be an optimal solution of (4.1) and $\boldsymbol{u}^\dagger \in \mathcal{U}$ be an optimal solution of (4.10). Since $\boldsymbol{u}^\dagger$ minimizes (4.10) and since (4.15) is a convex relaxation of (4.16),

$$\begin{aligned}
&\phi^{\mathrm{cv}}([\boldsymbol{x}^{cv}(t_f, \boldsymbol{u}^\dagger), \boldsymbol{x}^{cv}(t_f, \boldsymbol{u}^\dagger)], [\boldsymbol{x}^L(t_f), \boldsymbol{x}^U(t_f)]) \\
&\quad \leq \phi^{\mathrm{cv}}([\boldsymbol{x}^{cv}(t_f, \boldsymbol{u}^*), \boldsymbol{x}^{cv}(t_f, \boldsymbol{u}^*)], [\boldsymbol{x}^L(t_f), \boldsymbol{x}^U(t_f)]) \\
&\quad \leq \phi(\boldsymbol{x}(t_f, \boldsymbol{u}^*)),
\end{aligned}$$

which ensures the Result II.. □

4.5.5 Tightness

In this section, we demonstrate that if Scott and Barton's relaxed OCP (4.5) uses the state bounds in (4.11a) and if the RHS functions of (4.11b) are constructed as in (4.12), then (4.10) generates a lower bound that is tighter than Scott and Barton's lower bound (4.5).

Theorem 4.2. *Suppose that* ϕ^{cv} *in* (4.10), $\hat{\phi}^{\mathrm{cv}}$ *in* (4.5), $\hat{\boldsymbol{f}}^{\mathrm{cv}}, \hat{\boldsymbol{f}}^{\mathrm{cc}}$ *in* (4.6), *and* $\bar{\boldsymbol{f}}^{cv}, \bar{\boldsymbol{f}}^{cc}$ *in* (4.12) *are constructed with DMR. Let* $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ *in* (4.11) *be defined as in* (4.12). *Let* $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}, \boldsymbol{x}^L, \boldsymbol{x}^U)$ *be a solution of* (4.11). *Moreover, let* $\boldsymbol{x}^L, \boldsymbol{x}^U$ *be the state bounds used in* (4.6) *and let* $(\hat{\boldsymbol{x}}^{\mathrm{cv}}, \hat{\boldsymbol{x}}^{\mathrm{cc}})$ *be a solution of* (4.6). *Assume that* $\boldsymbol{u}^\dagger \in \mathcal{U}$ *minimizes* (4.5) *and* $\boldsymbol{u}^\ddagger \in \mathcal{U}$ *minimizes* (4.10). *Then,*

$$\begin{aligned}
&\phi^{\mathrm{cv}}([\boldsymbol{x}^{\mathrm{cv}}(t_f, \boldsymbol{u}^\ddagger), \boldsymbol{x}^{\mathrm{cc}}(t_f, \boldsymbol{u}^\ddagger)], [\boldsymbol{x}^L(t_f), \boldsymbol{x}^U(t_f)]) \\
&\quad \geq \hat{\phi}^{\mathrm{cv}}([\hat{\boldsymbol{x}}^{\mathrm{cv}}(t_f, \boldsymbol{u}^\dagger), \hat{\boldsymbol{x}}^{\mathrm{cc}}(t_f, \boldsymbol{u}^\dagger)], [\boldsymbol{x}^L(t_f), \boldsymbol{x}^U(t_f)]).
\end{aligned}$$

Proof. First, we demonstrate that

$$[\boldsymbol{x}^{\mathrm{cv}}(t_f, \boldsymbol{u}^{\ddagger}), \boldsymbol{x}^{\mathrm{cc}}(t_f, \boldsymbol{u}^{\ddagger})] \subseteq [\hat{\boldsymbol{x}}^{\mathrm{cv}}(t_f, \boldsymbol{u}^{\ddagger}), \hat{\boldsymbol{x}}^{\mathrm{cc}}(t_f, \boldsymbol{u}^{\ddagger})] \tag{4.17}$$

using Proposition 4.1. It is desired to verify that the three conditions in Proposition 4.1 hold with $t \mapsto \hat{\boldsymbol{x}}^{\mathrm{cv}}(t, \boldsymbol{u}^{\ddagger}), t \mapsto \hat{\boldsymbol{x}}^{\mathrm{cc}}(t, \boldsymbol{u}^{\ddagger})$ in place of $\boldsymbol{\xi}^{\mathrm{L}}, \boldsymbol{\xi}^{\mathrm{U}}$ and $t \mapsto \boldsymbol{x}^{\mathrm{cv}}(t, \boldsymbol{u}^{\ddagger}), t \mapsto \boldsymbol{x}^{\mathrm{cc}}(t, \boldsymbol{u}^{\ddagger})$ in place of $\boldsymbol{\xi}^{\dagger}, \boldsymbol{\xi}^{\ddagger}$. Since $\hat{\boldsymbol{f}}^{\mathrm{cv}} \equiv \bar{\boldsymbol{f}}^{cv}$ and $\hat{\boldsymbol{f}}^{\mathrm{cc}} \equiv \bar{\boldsymbol{f}}^{cc}$ are constructed with DMR, $\hat{\boldsymbol{f}}^{\mathrm{cv}}, \hat{\boldsymbol{f}}^{\mathrm{cc}}$ satisfy Assumption 4.3. Condition I. in Assumption 4.3 ensures the first condition in Proposition 4.1. Next, we investigate the enclosing dynamics in the second condition of Proposition 4.1 following Definition 4.5. It suffices to show that, for a.e. $t \in I$ and any $\Xi^{\mathrm{R}} \in \mathbb{IR}^{n_x}$ such that $[\boldsymbol{x}^{cv}(t, \boldsymbol{u}^{\ddagger}), \boldsymbol{x}^{cc}(t, \boldsymbol{u}^{\ddagger})] \subseteq \Xi^{\mathrm{R}} \equiv [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}] \subseteq X^{\mathrm{B}}(t) \equiv [\boldsymbol{x}^{L}(t), \boldsymbol{x}^{U}(t)]$,

1. If $\xi_i^{cv} = x_i^{cv}(t, \boldsymbol{u}^{\ddagger})$, then $\hat{f}_i^{\mathrm{cv}}(t, [\boldsymbol{u}^{\ddagger}(t), \boldsymbol{u}^{\ddagger}(t)], \Xi^{\mathrm{R}}, X^{\mathrm{B}}(t)) \leq \dot{x}_i^{cv}(t, \boldsymbol{u}^{\ddagger})$.
2. If $\xi_i^{cc} = x_i^{cc}(t, \boldsymbol{u}^{\ddagger})$, then $\hat{f}_i^{\mathrm{cc}}(t, [\boldsymbol{u}^{\ddagger}(t), \boldsymbol{u}^{\ddagger}(t)], \Xi^{\mathrm{R}}, X^{\mathrm{B}}(t)) \geq \dot{x}_i^{cc}(t, \boldsymbol{u}^{\ddagger})$.

It will be shown that the first requirement holds; verifying the second is analogous. Since $[\boldsymbol{x}^{cv}(t, \boldsymbol{u}^{\ddagger}), \boldsymbol{x}^{cc}(t, \boldsymbol{u}^{\ddagger})] \subseteq \Xi^{\mathrm{R}}$, $B_i^L([\boldsymbol{x}^{cv}(t, \boldsymbol{u}^{\ddagger}), \boldsymbol{x}^{cc}(t, \boldsymbol{u}^{\ddagger})]) \subseteq \Xi^{\mathrm{R}}$. The inclusion monotonicity of $\hat{\boldsymbol{f}}^{\mathrm{cv}}, \hat{\boldsymbol{f}}^{\mathrm{cc}}$ ensures that

$$\begin{aligned}
&\hat{f}_i^{\mathrm{cv}}(t, [\boldsymbol{u}^{\ddagger}(t), \boldsymbol{u}^{\ddagger}(t)], \Xi^{\mathrm{R}}, X^{\mathrm{B}}(t)) \\
&\quad \leq \hat{f}_i^{\mathrm{cv}}(t, [\boldsymbol{u}^{\ddagger}(t), \boldsymbol{u}^{\ddagger}(t)], B_i^L([\boldsymbol{x}^{cv}(t, \boldsymbol{u}^{\ddagger}), \boldsymbol{x}^{cc}(t, \boldsymbol{u}^{\ddagger})]), X^{\mathrm{B}}(t)) \\
&\quad = \bar{f}_i^{\mathrm{cv}}(t, [\boldsymbol{u}^{\ddagger}(t), \boldsymbol{u}^{\ddagger}(t)], B_i^L([\boldsymbol{x}^{cv}(t, \boldsymbol{u}^{\ddagger}), \boldsymbol{x}^{cc}(t, \boldsymbol{u}^{\ddagger})]), X^{\mathrm{B}}(t)) \\
&\quad = f_i^{\mathrm{cv}}(t, [\boldsymbol{u}^{\ddagger}(t), \boldsymbol{u}^{\ddagger}(t)], [\boldsymbol{x}^{cv}(t, \boldsymbol{u}^{\ddagger}), \boldsymbol{x}^{cc}(t, \boldsymbol{u}^{\ddagger})], X^{\mathrm{B}}(t)) \\
&\quad = \dot{x}_i^{cv}(t, \boldsymbol{u}^{\ddagger}),
\end{aligned}$$

which is the desired inequality in the first requirement.

Lastly, since ϕ^{cv} is generated with DMR and since (4.17) holds, the inclusion monotonicity of DMR [72] ensures that

$$\begin{aligned}&\phi^{\text{cv}}([\boldsymbol{x}^{\text{cv}}(t_f,\boldsymbol{u}^{\ddagger}),\boldsymbol{x}^{\text{cc}}(t_f,\boldsymbol{u}^{\ddagger})],[\boldsymbol{x}^{L}(t_f),\boldsymbol{x}^{U}(t_f)])\\&\quad\geq \phi^{\text{cv}}([\hat{\boldsymbol{x}}^{\text{cv}}(t_f,\boldsymbol{u}^{\ddagger}),\hat{\boldsymbol{x}}^{\text{cc}}(t_f,\boldsymbol{u}^{\ddagger})],[\boldsymbol{x}^{L}(t_f),\boldsymbol{x}^{U}(t_f)]).\end{aligned}$$

Since $\hat{\phi}^{\text{cv}} \equiv \phi^{\text{cv}}$ and since $\boldsymbol{u}^{\dagger} \in \mathcal{U}$ minimizes (4.5),

$$\begin{aligned}&\phi^{\text{cv}}([\boldsymbol{x}^{\text{cv}}(t_f,\boldsymbol{u}^{\ddagger}),\boldsymbol{x}^{\text{cc}}(t_f,\boldsymbol{u}^{\ddagger})],[\boldsymbol{x}^{L}(t_f),\boldsymbol{x}^{U}(t_f)])\\&\quad\geq \hat{\phi}^{\text{cv}}([\hat{\boldsymbol{x}}^{\text{cv}}(t_f,\boldsymbol{u}^{\ddagger}),\hat{\boldsymbol{x}}^{\text{cc}}(t_f,\boldsymbol{u}^{\ddagger})],[\boldsymbol{x}^{L}(t_f),\boldsymbol{x}^{U}(t_f)])\\&\quad\geq \hat{\phi}^{\text{cv}}([\hat{\boldsymbol{x}}^{\text{cv}}(t_f,\boldsymbol{u}^{\dagger}),\hat{\boldsymbol{x}}^{\text{cc}}(t_f,\boldsymbol{u}^{\dagger})],[\boldsymbol{x}^{L}(t_f),\boldsymbol{x}^{U}(t_f)]),\end{aligned}$$

which is the desired inequality. □

4.6 Solving the Underestimating Problem

This section presents an approach to solve our new relaxed OCP (4.10) to global optimality. Since RHS functions $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ in ODE (4.11) are not convex, the OCP (4.10) is not a convex OCP in the sense of [13]. Thus, the gradient-based numerical method in [13], which was proposed to solve Scott and Barton's OCP (4.5) [119], is not applicable here. Instead, we will solve OCP (4.10) using PMP. Because the mapping $\boldsymbol{u} \mapsto \phi^{\text{cv}}([\boldsymbol{x}^{cv}(t_f,\boldsymbol{u}),\boldsymbol{x}^{cc}(t_f,\boldsymbol{u})],[\boldsymbol{x}^{L}(t_f),\boldsymbol{x}^{U}(t_f)])$ is convex on $\mathcal{U}$ as demonstrated in Theorem 4.1, the PMP conditions in (4.8)-(4.9) actually provide a globally optimal solution of (4.10) according to Proposition 4.3 in Section 4.3.

For simplicity, define $\boldsymbol{y}^{\mathrm{B}}, \boldsymbol{y}^{\mathrm{R}}, \boldsymbol{y}_0 \in \mathbb{R}^{2n_x}$ and $\boldsymbol{f}^{\mathrm{R}} : I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{2n_x}$ such that $\boldsymbol{y}^{\mathrm{B}} \equiv (\boldsymbol{x}^L, \boldsymbol{x}^U)$, $\boldsymbol{y}^{\mathrm{R}} \equiv (\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$, and $\boldsymbol{f}^{\mathrm{R}} \equiv (\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc})$. Then, (4.10) becomes:

$$\min_{\boldsymbol{u}\in U} \quad \phi^{\mathrm{cv}}(\boldsymbol{y}^R(t_f, \boldsymbol{u}), \boldsymbol{y}^B(t_f)), \tag{4.18}$$

where $(\boldsymbol{y}^R, \boldsymbol{y}^B)$ solves the following ODE:

$$\begin{aligned} \dot{\boldsymbol{y}}^B(t) &= \boldsymbol{f}^R(t, U, \boldsymbol{y}^B(t), \boldsymbol{y}^B(t)), & \boldsymbol{y}^B(t_0) &= \boldsymbol{y}_0, \\ \dot{\boldsymbol{y}}^R(t, \boldsymbol{u}) &= \boldsymbol{f}^R(t, [\boldsymbol{u}(t), \boldsymbol{u}(t)], \boldsymbol{y}^R(t, \boldsymbol{u}), \boldsymbol{y}^B(t)), & \boldsymbol{y}^R(t_0, \boldsymbol{u}) &= \boldsymbol{y}_0. \end{aligned}$$

For this OCP, the PMP conditions (4.8) and (4.9) are stated as follows, in terms of variables $\boldsymbol{y}^{\mathrm{R}}$, $\boldsymbol{y}^{\mathrm{B}}$, and $\boldsymbol{\lambda}$:

$$\begin{aligned} \dot{\boldsymbol{y}}^B(t) &= \boldsymbol{f}^B(t, U, \boldsymbol{y}^B(t), \boldsymbol{y}^B(t)), \\ \dot{\boldsymbol{y}}^R(t, \boldsymbol{u}) &= \boldsymbol{f}^R(t, [\boldsymbol{u}^*(t), \boldsymbol{u}^*(t)], \boldsymbol{y}^R(t, \boldsymbol{u}^*), \boldsymbol{y}^B(t)), \\ \dot{\boldsymbol{\lambda}}(t) &= -\boldsymbol{\lambda}(t)\, \mathrm{D}_{\boldsymbol{y}^R} \boldsymbol{f}^R(t, [\boldsymbol{u}^*(t), \boldsymbol{u}^*(t)], \boldsymbol{y}^{\mathrm{R}}(t, \boldsymbol{u}^*), \boldsymbol{y}^B(t)), \\ \boldsymbol{y}^B(t_0) &= \boldsymbol{y}_0, \quad \boldsymbol{y}^R(t_0) = \boldsymbol{y}_0, \\ \boldsymbol{\lambda}(t_f) &= \mathrm{D}_{\boldsymbol{y}^R} \phi^{\mathrm{cv}}(\boldsymbol{y}^R(t_f, \boldsymbol{u}^*), \boldsymbol{y}^B(t_f)), \end{aligned} \tag{4.19}$$

where for a.e. $t \in I$,

$$\boldsymbol{u}^*(t) = \arg\min_{\boldsymbol{\omega}\in U} \langle \boldsymbol{\lambda}(t),\ \boldsymbol{f}^R(t, [\boldsymbol{\omega}, \boldsymbol{\omega}], \boldsymbol{y}^R(t, \boldsymbol{u}^*), \boldsymbol{y}^B(t)) \rangle. \tag{4.20}$$

Theorem 4.3. *Under Assumptions 4.1 and 4.2, let* $(\boldsymbol{y}^R, \boldsymbol{y}^B, \boldsymbol{\lambda})$ *be a trajectory satisfying* (4.19) *and let* $\boldsymbol{u}^*$ *be the corresponding control. Then,* $\boldsymbol{u}^*$ *is a globally optimal control input for* (4.18)*. Moreover,* $\phi^{\mathrm{cv}}(\boldsymbol{y}^R(t_f, \boldsymbol{u}^*), \boldsymbol{y}^B(t_f))$ *is a lower bound of the optimal solution*

value of the original OCP (4.1).

Proof. $\boldsymbol{f}^{\mathrm{R}}$ is continuously differentiable with respect to $\boldsymbol{y}^{\mathrm{R}}$ according to Assumption 4.2, so $\mathrm{D}_{\boldsymbol{y}^{\mathrm{R}}}\boldsymbol{f}^{R}$ is continuous. Result I. in Theorem 4.1 shows that the mapping $\boldsymbol{u} \mapsto \phi^{\mathrm{cv}}(\boldsymbol{y}^{\mathrm{R}}(t_f, \boldsymbol{u}), \boldsymbol{y}^{\mathrm{B}}(t_f))$ is convex. Thus, Proposition 4.3 ensures that $\boldsymbol{u}$ is a globally optimal solution of (4.18). Moreover, since $\phi^{\mathrm{cv}}(\boldsymbol{y}^{R}(t_f, \boldsymbol{u}^*), \boldsymbol{y}^{B}(t_f))$ is a globally optimal solution value of (4.10), Result II. in Theorem 4.1 guarantees that it is a lower bound of the optimal solution value of (4.1). □

Observe that the ODE in (4.19) is a two-point boundary-value problem with both initial and terminal conditions. It may be solved numerically by standard collocation methods or shooting methods. Furthermore, the optimization problem in (4.20) is trivial to solve if $\boldsymbol{\omega} \mapsto \boldsymbol{f}^{R}(t, [\boldsymbol{\omega}, \boldsymbol{\omega}], \boldsymbol{y}^{R}(t, \boldsymbol{u}), \boldsymbol{y}^{B}(t))$ is affine. In this case, the original OCP (4.1) is a control-affine OCP, and has been widely studied [85, 13, 26].

Corollary 4.1. *Consider a nonlinear function* $\boldsymbol{g} : I \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ *and a matrix-valued function* $B : I \to \mathbb{R}^{n_x \times n_u}$. *Consider a control-affine instance of* (4.2) *where the ODE RHS function* $\boldsymbol{f}$ *is given by*

$$\boldsymbol{f}(t, \boldsymbol{u}, \boldsymbol{x}) \equiv \boldsymbol{g}(t, \boldsymbol{x}(t, \boldsymbol{u})) + B(t)\boldsymbol{u}(t). \tag{4.21}$$

Suppose that the ODE RHS functions $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ *are constructed following Section 4.4.1 using DMR. Then, the control* $\boldsymbol{u}^*$ *in the following provides a closed-form globally optimal*

solution for (4.20)*: for each* $j \in \{1, \ldots, n_u\}$,

$$u_j^*(t) \equiv \begin{cases} u_j^{\mathrm{L}}, & \text{if } (\boldsymbol{\lambda}^{\mathrm{cv}}(t) + \boldsymbol{\lambda}^{\mathrm{cc}}(t))B_{(j)}(t) \geq 0, \\ u_j^{\mathrm{U}}, & \text{otherwise}, \end{cases} \tag{4.22}$$

where $B_{(j)}$ *is the jth column of* B.

Proof. Consider functions $\bar{\boldsymbol{g}}^{\mathrm{cv}}, \bar{\boldsymbol{g}}^{\mathrm{cc}} : I \times \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ such that, for any convex and concave relaxations $\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc} : \mathcal{U} \to \mathbb{R}^{n_x}$ of $\boldsymbol{u} \mapsto \boldsymbol{x}(t, \boldsymbol{u})$ on $\mathcal{U}$, the mappings $\boldsymbol{u} \mapsto \bar{\boldsymbol{g}}^{\mathrm{cv}}(t, [\boldsymbol{\xi}^{cv}(\boldsymbol{u}), \boldsymbol{\xi}^{cc}(\boldsymbol{u})], X^{\mathrm{B}}(t))$ and $\boldsymbol{u} \mapsto \bar{\boldsymbol{g}}^{\mathrm{cc}}(t, [\boldsymbol{\xi}^{cv}(\boldsymbol{u}), \boldsymbol{\xi}^{cc}(\boldsymbol{u})], X^{\mathrm{B}}(t))$ are convex and concave relaxations of $\boldsymbol{u} \mapsto \boldsymbol{g}(t, \boldsymbol{x}(t, \boldsymbol{u})))$ on $\mathcal{U}$, respectively, for all $t \in I$. In this case, functions $\bar{\boldsymbol{g}}^{\mathrm{cv}}, \bar{\boldsymbol{g}}^{\mathrm{cc}}$ are constructed from $\boldsymbol{g}$ using DMR. Define $\boldsymbol{g}^{\mathrm{cv}}, \boldsymbol{g}^{\mathrm{cc}} : I \times \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ such that

$$\begin{aligned} \boldsymbol{g}^{\mathrm{cv}}(t, \Xi^{\mathrm{R}}, \Xi^{\mathrm{B}}) &\equiv \bar{\boldsymbol{g}}^{\mathrm{cv}}(t, B_i^L(\Xi^{\mathrm{R}}), \Xi^{\mathrm{B}}), \\ \boldsymbol{g}^{\mathrm{cc}}(t, \Xi^{\mathrm{R}}, \Xi^{\mathrm{B}}) &\equiv \bar{\boldsymbol{g}}^{\mathrm{cc}}(t, B_i^U(\Xi^{\mathrm{R}}), \Xi^{\mathrm{B}}), \end{aligned}$$

and let $\boldsymbol{g}^{\mathrm{R}} \equiv (\boldsymbol{g}^{\mathrm{cv}}, \boldsymbol{g}^{\mathrm{cc}})$. Then, we construct $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ as follows:

$$\begin{aligned} \boldsymbol{f}^{cv}(t, [\boldsymbol{p}^{\mathrm{L}}, \boldsymbol{p}^{\mathrm{U}}], \Xi^{\mathrm{R}}, \Xi^{\mathrm{B}}) &\equiv \boldsymbol{g}^{\mathrm{cv}}(t, \Xi^{\mathrm{R}}, \Xi^{\mathrm{B}}) + B(t)\boldsymbol{p}^{\mathrm{L}}, \\ \boldsymbol{f}^{cc}(t, [\boldsymbol{p}^{\mathrm{L}}, \boldsymbol{p}^{\mathrm{U}}], \Xi^{\mathrm{R}}, \Xi^{\mathrm{B}}) &\equiv \boldsymbol{g}^{\mathrm{cc}}(t, \Xi^{\mathrm{R}}, \Xi^{\mathrm{B}}) + B(t)\boldsymbol{p}^{\mathrm{U}}. \end{aligned}$$

Representing $\boldsymbol{\lambda}$ in (4.19) as $(\boldsymbol{\lambda}^{\text{cv}}, \boldsymbol{\lambda}^{\text{cc}})$, in this case (4.20) becomes

$$\begin{aligned}\boldsymbol{u}^*(t) \in \underset{\boldsymbol{\omega}\in U}{\arg\min}\{&\langle \boldsymbol{\lambda}^{\text{cv}}(t),\, \boldsymbol{g}^{\text{cv}}(t, \boldsymbol{y}^R(t, \boldsymbol{u}^*), \boldsymbol{y}^B(t)) + B(t)\boldsymbol{\omega}\rangle \\ &+ \langle \boldsymbol{\lambda}^{\text{cc}}(t),\, \boldsymbol{g}^{\text{cc}}(t, \boldsymbol{y}^R(t, \boldsymbol{u}^*), \boldsymbol{y}^B(t)) + B(t)\boldsymbol{\omega}\rangle\} \\ = \underset{\boldsymbol{\omega}\in U}{\arg\min}\{&\langle \boldsymbol{\lambda}(t),\, \boldsymbol{g}^R(t, \boldsymbol{y}^R(t, \boldsymbol{u}^*), \boldsymbol{y}^B(t)\rangle \\ &+ \langle \boldsymbol{\lambda}^{\text{cv}}(t) + \boldsymbol{\lambda}^{\text{cc}}(t),\, B(t)\boldsymbol{\omega}\rangle\}. \end{aligned} \tag{4.23}$$

Therefore, (4.22) is a globally optimal solution of (4.20) by inspection. □

Corollary 4.1 demonstrates that, when the original OCP (4.1) is a control-affine instance, our new relaxed OCP (4.10) can be solved trivially using PMP and a closed-form solution is available. This property is particularly useful in deterministic global optimization, because branch-and-bound algorithms typically require many evaluations of a guaranteed global lower bound.

4.7 Numerical Examples

This section presents numerical examples to demonstrate bounding the optimal solution value of nonconvex OCPs using the new underestimating OCP in (4.10). To compute valid lower bounds, PMP was applied as in Theorem 4.3. A proof-of-concept implementation that solves the two-point boundary-value problem in (4.19) was developed in Julia v1.5.3 [20] using the package DifferentiableEquations.jl [104] as the numerical integrator. Natural interval extension was constructed using the package IntervalArithmetic.jl. DMR was generated with the package McCormick.jl in EAGO.jl [152]. All numerical experiments were performed on a

Windows 10 machine with a 3.6 GHz Ryzen 5 2600X CPU and 8 GB memory.

The following example is modified from [19, Example 3.2.1] by adding a nonlinear nonconvex term to the RHS function of the ODE (4.25).

Example 4.1. *Consider the following instance of the OCP* (4.1)*:*

$$\min_{u\in\mathcal{U}} \quad \phi(x(t_f,u)) = \frac{1}{2}(x(t_f,u))^2, \tag{4.24}$$

where x solves the ODE:

$$\begin{aligned} \dot{x}(t,u) &= -\frac{1}{4}((x(t,u))^3 - (x(t,u))^2) + u(t), \\ x(t_0,u) &= -\frac{1}{2}, \end{aligned} \tag{4.25}$$

and $U \equiv [-1,1]$, $u \in \mathcal{U}$, $I \equiv [t_0,t_f] = [0,1]$.

To generate a lower bound for (4.24), we first construct the relaxed OCP in (4.10). Since $\phi(\xi) = \frac{1}{2}\xi^2$ is a convex quadratic function, we may set $\phi^{\mathrm{cv}} = \phi$, which satisfies Assumption 4.2. Next, we applied the results in Section 4.6 to solve the relaxed OCP using the PMP conditions described in (4.19) and (4.20). Observe that the ODE RHS function in (4.25) is consistent with the formulation in (4.21) with $g(t,x) = -\frac{1}{4}(x^3 - x^2)$ and $B = 1$. So, we constructed convex and concave relaxations of ODE RHS function with DMR following Section 4.4.1 and determined the optimal control inputs using Corollary 4.1.

The relaxed trajectory $t \mapsto \phi^{\mathrm{cv}}(x(t,u^*))$ is plotted in Figure 4.1, along with the system trajectories $t \mapsto \phi(x(t,u))$ where $u(t) = \sin(2(t+\pi p))$ and $p \in [0,1]$. As shown in Figure 4.1, the optimal solution value of the underestimating problem is a lower bound for $\phi(x(t,u))$ with various control inputs u at the terminal time $t_f =$

1. Observe that the trajectory $t \mapsto \phi^{\text{cv}}(x(t, u^*))$ does not underestimate trajectories $t \mapsto \phi(x(t, u))$ for most earlier times $t < t_f$ in this example.

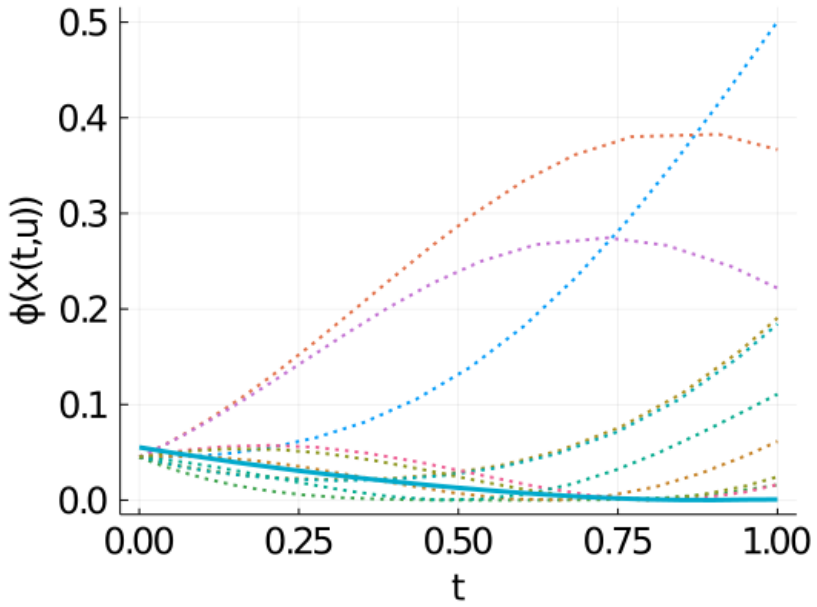


Figure 4.1: The system trajectories and the relaxed trajectory in Example 4.1: trajectories $t \mapsto \phi(x(t, u))$ (dotted) where u is a suboptimal control and trajectory $t \mapsto \phi^{\text{cv}}(x(t, u^*))$ (solid) where u^* is a globally optimal control

Next, we consider a second example adapted from Example 6.3 in [26], The cost function in this example is nonlinear and nonconvex, so the terminal condition of the adjoint equations is not known *a priori*.

Example 4.2. *Consider the instance of the OCP* (4.1)*:*

$$\min_{u \in \mathcal{U}} \quad \phi(\boldsymbol{x}(t_f, u)) = x_1(t_f, u) - (x_2(t_f, u))^2, \tag{4.26}$$

where $\boldsymbol{x}$ solves the following ODE:

$$\begin{aligned} \dot{x}_1(t,u) &= (x_2(t,u))^3, \quad & x_1(t_0,u) &= 0.2, \\ \dot{x}_2(t,u) &= u(t), & x_2(t_0,u) &= 0.1, \end{aligned}$$

and $U \equiv [-1,1]$, $u \in \mathcal{U}$, $I \equiv [t_0,t_f] = [0,1]$.

An underestimating OCP of (4.26) was constructed and solved similarly to Example 4.1, except that ϕ^{cv} was generated with DMR. The objective trajectory $t \mapsto \phi^{\text{cv}}(\boldsymbol{x}(t,u^*))$ is plotted in Figure 4.2, along with trajectories $t \mapsto \phi(\boldsymbol{x}(t,u))$ for various suboptimal control trajectories u. Observe that the optimal solution value of the underestimating OCP, $\phi^{\text{cv}}(\boldsymbol{x}(t_f,u^*))$, is an underestimator of $\phi(\boldsymbol{x}(t_f,u))$ at the final time $t_f = 1$.

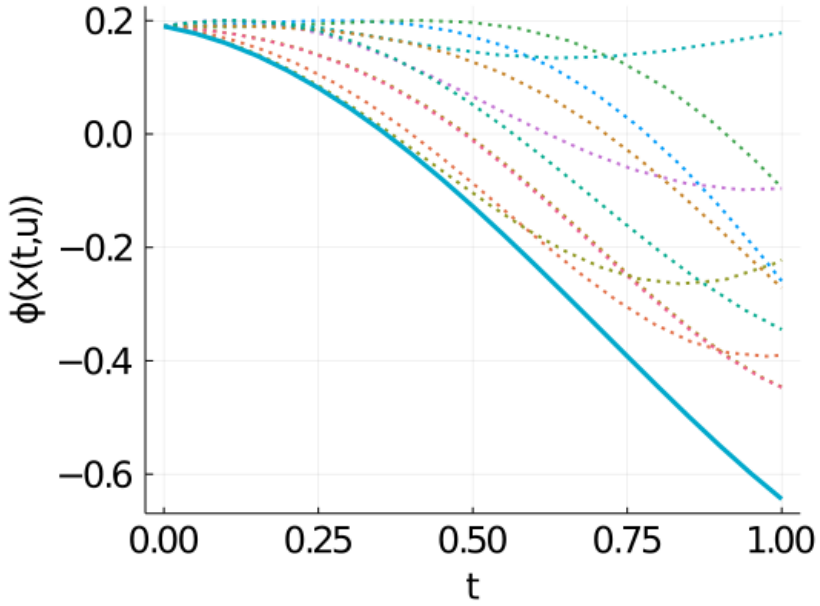


Figure 4.2: The system trajectories and the relaxed trajectory in Example 4.2: trajectories $t \mapsto \phi(\boldsymbol{x}(t,u))$ (dotted) where u is a suboptimal control and trajectory $t \mapsto \phi^{\text{cv}}(\boldsymbol{x}(t,u^*))$ (solid) where u^* is a globally optimal control

Furthermore, we compare the new underestimating method with the Harrison's method [59] as described in Section 4.1. Using Harrison's method, we obtained an underestimator for (4.26) of -1.174, which is less than the optimal solution value of the new underestimating OCP $\phi^{\text{cv}}(x(t_f, u^*)) = -0.644$. Therefore, our new method constructed a tighter lower bound than Harrison's method for this example.

Example 4.3. *Consider the instance of the OCP* (4.1)*:*

$$\min_{u \in \mathcal{U}} \quad \phi(x(t_f, u)) = -\frac{1}{2}(x(t_f, u))^2, \tag{4.27}$$

where x solves the ODE:

$$\begin{aligned} \dot{x}(t, u) &= -\frac{1}{3}(x(t, u))^3 - \frac{1}{2}(x(t, u))^2 + u(t), \\ x(t_0, u) &= -0.1, \end{aligned} \tag{4.28}$$

and $U \equiv [-1, 1]$, $u \in \mathcal{U}$, $I \equiv [t_0, t_f] = [0, 1]$.

In this example, we compare our new approach with Scott and Barton's method [119], as well as the simultaneous approach described in Section 4.1. Our new approach was implemented similarly to Example 4.2. For Scott and Barton's method, the functions $\hat{\phi}^{\text{cv}}$ in (4.5) and $\hat{\boldsymbol{f}}^{\text{cv}}$, $\hat{\boldsymbol{f}}^{\text{cc}}$ in (4.6) were generated with DMR as discussed in Section 4.3.3, and thereby satisfy Assumptions 4.1 and 4.2, respectively. Therefore, when we implemented Scott and Barton's method, the globally optimal solution of the convex OCP (4.5) was determined using PMP following Theorem 4.3. Denote this optimal control as $\hat{u}^*$.

To implement the simultaneous approach, we first discretized both control and

state evenly over n time-intervals of duration and each time step $\Delta t = \frac{t_f - t_0}{n}$. Then, the original OCP (4.1) was approximated as the following NLP where the ODE solution is approximated using the implicit Euler method.

$$\begin{aligned} \min_{\substack{\boldsymbol{v}^i \in U,\, \boldsymbol{\xi}^i \in \Xi^i, \\ i \in \{1,\ldots,n\}}} \quad & \phi(\boldsymbol{\xi}^n), \\ \text{subject to} \quad & \boldsymbol{\xi}^i - \boldsymbol{\xi}^{i-1} - \Delta t\, \boldsymbol{f}(i\Delta t, \boldsymbol{v}^i, \boldsymbol{\xi}^i) = \boldsymbol{0}, \quad i \in \{1,\ldots,n\}, \\ & \boldsymbol{\xi}^0 = \boldsymbol{x}_0. \end{aligned} \tag{4.29}$$

The interval bound of the discretized state $\boldsymbol{\xi}^i$, Ξ^i, was generated using the forward Euler method and natural interval extension:

$$\begin{aligned} \Xi^i &= \Xi^{i-1} + \Delta t\, F(i\Delta t, U, \Xi^{i-1}), \quad \forall i \in \{1,\ldots,n\}, \\ \Xi^0 &= [\boldsymbol{x}_0, \boldsymbol{x}_0], \end{aligned}$$

where $F : I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ is the natural interval extension of $\boldsymbol{f}$. Finally, we used DMR to construct a convex relaxation problem of the NLP (4.29) with same numbers of variables and constraints, and solved this to global optimality. This approach has been used to underestimate nonconvex NLPs in deterministic global optimization [138]. Denote this convex relaxation problem's global solution as $(\tilde{\boldsymbol{v}}^i, \tilde{\boldsymbol{\xi}}^i)$ and its optimal objective value as $\tilde{\phi}^{\text{cv}}(\tilde{\boldsymbol{\xi}}^n)$ where $\tilde{\phi}^{\text{cv}}$ is the convex relaxation of ϕ generated with DMR. Then, $\tilde{\phi}^{\text{cv}}(\tilde{\boldsymbol{\xi}}^n)$ is a guaranteed lower bound of the global optimal solution value of (4.29). Furthermore, the sequence $[\tilde{\phi}^{\text{cv}}(\tilde{\boldsymbol{\xi}}^i)]_{i \in \{1,\ldots,n\}}$ represents the value of the relaxed objective function $\tilde{\phi}^{\text{cv}}$ at each mesh point. It describes how the value of the relaxed cost function evolves with time, to mimic the

trajectory $t \mapsto \phi^{\mathrm{cv}}(x(t, u^*))$ in our continuous-time approach and form a comparison.

Figure 4.3 shows the trajectory $t \mapsto \phi^{\mathrm{cv}}(x(t, u^*))$ generated using our new approach, the trajectory $t \mapsto \hat{\phi}^{\mathrm{cv}}(x(t, \hat{u}^*))$ generated using Scott and Barton's method, as well as the sequence $[\tilde{\phi}^{\mathrm{cv}}(\tilde{\xi}^i)]_{i \in \{1,\ldots,n\}}$ generated using the discretization approach with two different numbers of discretization points: $n = 10$ and $n = 30$. For this example, our new approach generated a tighter lower bound for the globally optimal solution of the original OCP (4.1) than these two established approaches.

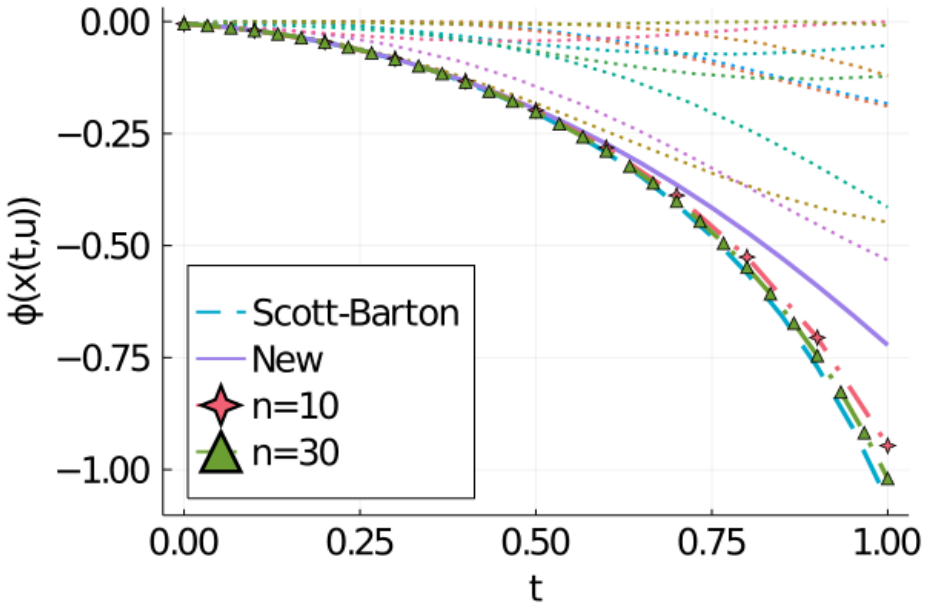


Figure 4.3: The system trajectories and the relaxed trajectories in Example 4.3: our new trajectory $t \mapsto \phi^{\mathrm{cv}}(x(t, u^*))$ (solid), the Scott-Barton trajectory $t \mapsto \hat{\phi}^{\mathrm{cv}}(x(t, \hat{u}^*))$ (dashed), and the sequence $[\tilde{\phi}^{\mathrm{cv}}(\tilde{\xi}^i)]_{i \in \{1,\ldots,n\}}$ with $n = 10$ (star) and $n = 30$ (triangle), and trajectories $t \mapsto \phi(x(t, u))$ (dotted) for various suboptimal control inputs u

4.8 Conclusion

This article presented a novel approach for underestimating a nonconvex open-loop OCP without discretizing the control or state trajectories. We constructed a relaxed OCP in (4.10) using the state relaxations and state bounds described in (4.11). Our new approach improved Scott and Barton's relaxed OCP (4.5) [119] by tightening the ODE RHS functions with the flattening operators in Definition 4.7. This modification weakened the convexity of the relaxed OCP and made the theoretical results in [119] no longer applicable. Instead, we underwent a completely different theoretical development process using the theory of differential inequalities [148, 120]. Theorem 4.1 demonstrated that the optimal solution value of the relaxed OCP is a guaranteed lower bound for the optimal solution value of the original OCP. It was also verified in Theorem 4.2 that this lower bound is always at least as tight as the lower bound generated with Scott and Barton's method. Such a tighter lower bound can in principle hasten the convergence of the branch-and-bound algorithms used in deterministic global optimization [42].

Furthermore, Theorem 4.3 illustrated that the relaxed OCP (4.10) can be solved to its global optimality using PMP. A two-point boundary-value problem (4.19) was developed to describe the globally optimal solution, which can be solved with a standard numerical integrator. Unlike Scott and Barton's method [119], the solution strategy of our approach is straightforward to implement and can be automated easily. In particular, when the original problem (4.1) is a control-affine OCP, the closed-form globally optimal solution of our relaxed OCP (4.10) can be determined trivially following Corollary 4.1. A proof-of-concept implementation of our new approach was developed in Julia. Numerical examples were presented

in Section 4.7 to illustrate that the new method generates valid lower bounds for nonconvex OCPs. For these examples, our new bounds are tighter than established methods [119, 59].

Future work may include constructing the ODE RHS functions in (4.11) using GMR [123], which is tighter than DMR but nonsmooth. Achieving this would likely require nonsmooth variants of the PMP formulations considered in this article, along the lines of [41]. Moreover, Section 4.6 focuses on control-affine OCPs. This is to ensure that the ODE RHS functions $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ in the underestimating OCP are also affine with respect to the control input, so that the lower-bounding problem (4.20) may ultimately be solved easily. It may be worthwhile to develop other techniques to construct functions $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ instead of using GMR or DMR.

Chapter 5

An Optimization-Based Framework for Enclosing Reachable Sets with Differential Inequalities

This chapter represents a manuscript in preparation for submission to a journal.

5.1 Introduction

The reachable set of a dynamic system is the set of possible states that the system may attain, given bounded uncertain controls, parameters, and initial conditions. This article presents a novel framework for enclosing the reachable set of nonlinear ordinary differential equations (ODEs) using differential inequalities. Such enclosures provide critical bounding information for algorithms in deterministic global dynamic optimization [101, 129, 81] and global optimal control [64]. They are also widely used in applications like safety verification [66], fault detection [82], and

state estimation [69]. While a number of different strategies have been established to compute such enclosures [91, 8, 78, 96, 76], this work particularly focuses on an extensively studied category of methods that are based on differentiable inequalities [59, 128, 118, 127, 62].

The theories of differential inequalities can be found in the book by Walter [148]. To generate time-varying bounds for a dynamic system using differential inequalities, an auxiliary system of ODEs need to constructed and solved numerically. This strategy requires rigorous lower and upper bounding information of the original system's right-hand side (RHS) functions, which has been central to the research in differential inequality-based methods. In established approaches, various bounding techniques have been developed. Harrison [59] proposed to calculate interval bounds of the RHS function automatically with natural interval extension (NIE) [94]. A flattening technique was implemented in Harrison's method to reduce the overestimation generated by NIE. Singer and Barton [128] provided bounding information by minimizing affine relaxations of the original RHS function. These affine relaxations are generated by linearizing classic McCormick-based relaxations using subgradients [92]. Neverthelss, the constructed auxiliary system of ODEs is not guaranteed to have a unique solution. Another method described by Harwood, et al. [62] involves embedding linear programs (LPs) into the RHS of the auxiliary system. To ensure the existence and uniqueness of a solution, a special relaxation technique is required to construct those LPs. A formulation of generalized differential inequalities was proposed by Villanueva, et al. [146], which may serve as a framework for established differential inequalities and ellipsoidal bounding approach [79]. It also supports propagating nonconvex enclosures for reachable

sets using Taylor models [88]. But the authors didn't report any improved convex enclosing methods developed using their framework.

Summarizing the differential inequality-based methods reviewed above, one thing in common can be found. They provide bounding information for the original system by optimizing various relaxations of the original ODE's RHS function. This commonality is the intuition of our new framework. In addition to covering these established approaches, this novel framework also permits the application of various other relaxation techniques for the original ODE's RHS functions. Therefore, it is necessary to provide a brief overview of available convex and concave relaxation methods.

McCormick [89] proposed a method for deriving (nonsmooth) convex and concave relaxation pairs for functions that are *factorable*. A function being factorable means that it is a composition of finite simple operations such as addition, multiplication, etc. A method for propagating the subgradients of McCormick's relaxations was later developed by Mitsos, et al. [92]. McCormick's relaxation method, along with Mitsos, et al.'s subgradient propagation rule, will be referred as *McCormick relaxations* (MC) in this article. Note that affine relaxations and piecewise-affine relaxations can be constructed by linearizing the nonlinear McCormick relaxations at fixed points using subgradients. Compared with nonlinear relaxations, these affine or piecewise-affine relaxations are computationally cheaper to optimize [50, 33]. Scott et al. [123] described a generalized formulation of MC and named it as *generalized McCormick relaxations* (GMC). It has one important property of taking previously known convex relaxations as arguments for further calculation, which is useful in the applications like differential inequalities and iterative algorithms.

Khan and his coworkers developed two smooth variants of GMC, termed as *differentiable McCormick relaxations* (DMC), to eliminate the theoretical and computational obstacles caused the non-smoothness in MC and GMC. The DMC introduced in [74] yields convex relaxations that are twice-continuously differentiable ($\mathscr{C}^2$), and will be referred as $\mathscr{C}^2$-DMC. The other version of DMC [72, 73], $\mathscr{C}^1$-DMC, constructs tighter but continuously differentiable ($\mathscr{C}^1$) relaxations. It is also worth mention that $\mathscr{C}^1$-DMC was developed based on the generalization of multivariate McCormick relaxations by Tsoukalas and Mitsos (T-M) [144]. Their method reformulates McCormick's composition theorem with an optimization problem to generate tighter convex relaxations than standard McCormick relaxations (i.e. MC and GMC). Similar to GMC and DMC, T-M relaxations can also take convex relaxations that are known *a priori* for further computation.

All the methods introduced above originate from McCormick's work and will be referred as *McCormick-type relaxations* here. Those generalized methods that can take known convex relaxations for further computation, i.e. GMC, DMC and T-M relaxations, will be referred as *generalized McCormick-type relaxations* in this work. The hierarchical relationship among these methods is described in Figure 5.1.

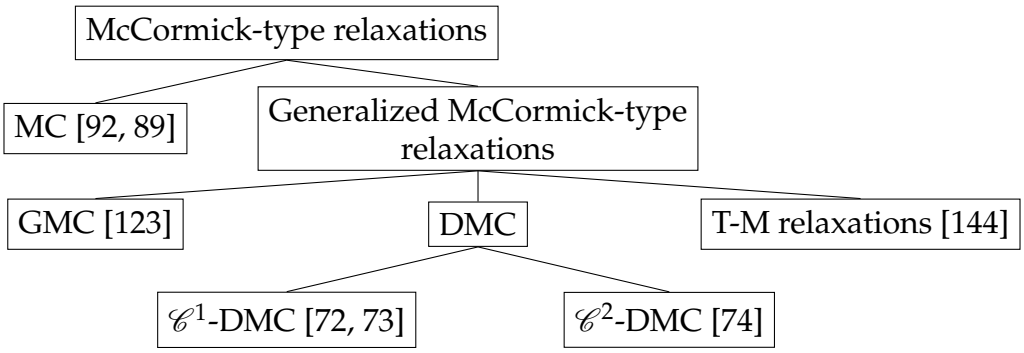


Figure 5.1: An illustration of the hierarchical relationship of McCormick-type relaxations

Moreover, we propose a novel usage of generalized McCormick-type relaxations in this work. If GMC, DMC, or T-M relaxations are fed with known interval bounds without parameter dependence for further calculation, then the computed convex relaxations will continue to be parameter-independent interval bounds. Hence, similar to NIE, GMC, DMC, and T-M relaxations are also able to construct interval extensions. Such interval extensions will be called *McCormick interval extensions* (MIE). They can be used to replace NIE in some applications, such as Harrison's method.

In addition to McCormick-type relaxations, αBB relaxation is another established method for generating convex relaxations [3]. It constructs convex underestimators for nonconvex $\mathscr{C}^2$ functions by adding a negative convex quadratic term to the original function. Another available method for bounding nonconvex functions is edge-concave relaxations [63]. Compared with previously described convex relaxation methods, edge-concave relaxation is relatively unorthodox. It provides bounding information for the original function via generating concave underestimators.

Besides permitting new methods for bounding the original system's RHS function, our new framework also support the strategy of generating tighter enclosures of reachable sets using known constraints of the original system, including the non-negativity of the states, physical bounds of the dynamic system, and conservation laws. Such constraints are typically expressed as *a priori* enclosures of the reachable set, and have been employed to construct tighter interval bounds for the original system [121, 118, 62, 127]. A generalized formulation of this strategy was developed in [124] to support for a wider range of *a priori* knowledge of

the original system, such as nonlinear constraints and constraints depending on time-varying inputs. Moreover, given an unconstrained nonlinear dynamic system, [126] proposed an strategy to deliberately construct useful constraints via introducing redundant state variables in order to produce tighter enclosures for the reachable set.

The remainder of this article is organized as follows. Section 5.2 introduces the notation convention and necessary definitions. The problem of interest is formally formulated in 5.3. Our new framework is presented in Section 5.4. Section 5.5 describes four use cases of this framework, as well as various methods for relaxing the original ODE's RHS function. Lastly, numerical examples are presented in Section 5.7 to illustrate the interval bounds constructed using this new framework for enclosing reachable sets.

5.2 Preliminaries

This section introduces the mathematical background underlying the methods and results in this article. The following notation conventions are used. The set of positive real numbers is represented as $\mathbb{R}_{>0}$, and $\mathbb{R}_{\geq 0}$ stands for the set of non-negative real numbers. The standard Euclidean norm $\|\cdot\|$ is adopted for any vector space $\mathbb{R}^n$, and $\|\cdot\|_\infty$ represents infinity norm. Vectors are denoted with boldface lower-case letters (e.g. $\boldsymbol{z}$). Given vectors $\boldsymbol{z}^\dagger, \boldsymbol{z}^\ddagger \in \mathbb{R}^n$, inequalities such as $\boldsymbol{z}^\dagger < \boldsymbol{z}^\ddagger$ or $\boldsymbol{z}^\dagger \leq \boldsymbol{z}^\ddagger$ are to be interpreted component-wise. $\boldsymbol{z}_{-i} \in \mathbb{R}^{n-1}$ stands for the vector $\boldsymbol{z}$ with the ith component excluded. Throughout this article, the convexity of a vector-valued function $\boldsymbol{h}$ refers to convexity of all components h_i. Dotted quantities indicate time-derivatives (e.g. $\dot{\boldsymbol{z}} \equiv \frac{\partial \boldsymbol{z}}{\partial t}$). The abbreviation "a.e." stands

for "almost every" in the Lebesgue sense. $\mathscr{C}^1$ stands for continuously differentiable and $\mathscr{C}^2$ stands for twice-continuously differentiable.

5.2.1 Intervals and interval functions

Definition 5.1 (Interval)**.** *For any $\boldsymbol{z}^L, \boldsymbol{z}^U \in \mathbb{R}^n$ such that $\boldsymbol{z}^L \leq \boldsymbol{z}^U$, define interval $Z \equiv [\boldsymbol{z}^L, \boldsymbol{z}^U]$ as the nonempty compact connected set of $\{\boldsymbol{z} \in \mathbb{R}^n : \boldsymbol{z}^L \leq \boldsymbol{z} \leq \boldsymbol{z}^U\}$. The set of all interval subsets of $D \subset \mathbb{R}^n$ is denoted as $\mathbb{I}D$, and $\mathbb{IR}^n$ denotes the set of all interval subsets of $\mathbb{R}^n$.*

Definition 5.2 (Interval function)**.** *Let $S \in \mathbb{IR}^n$ and $\boldsymbol{h} : S \to \mathbb{R}^m$.*

1. *An interval function $H \equiv [\boldsymbol{h}^L, \boldsymbol{h}^U] : \mathbb{IR}^n \to \mathbb{IR}^m$ is a* inclusion function *of $\boldsymbol{h}$ on S if for all $Z \subseteq S$,*

$$\{\boldsymbol{h}(\boldsymbol{z}) : \boldsymbol{z} \in Z\} \subseteq H(Z) \equiv [\boldsymbol{h}^L(Z), \boldsymbol{h}^U(Z)].$$

2. *Let $H^\dagger, H^\ddagger : \mathbb{IR}^n \to \mathbb{IR}^m$ be interval functions. $H^\dagger$ is* tighter *than $H^\ddagger$ on S if*

$$H^\dagger(Z) \subseteq H^\ddagger(Z), \quad \forall Z \subseteq S.$$

3. *H is* inclusion monotonic *on S if for all $Z^\dagger, Z^\ddagger \in S$ such that $Z^\dagger \subseteq Z^\ddagger$,*

$$H(Z^\dagger) \subseteq H(Z^\ddagger).$$

Definition 5.3 (Convex relaxation)**.** *Let $Z \in \mathbb{IR}^n$ and $\boldsymbol{h} : Z \to \mathbb{R}^m$.*

1. $\boldsymbol{h}^{cv} : Z \to \mathbb{R}^m$ *is a* convex relaxation *of* $\boldsymbol{h}$ *on* Z *if* $\boldsymbol{h}^{cv}(\boldsymbol{z}) \leq \boldsymbol{h}(\boldsymbol{z})$ *for all* $\boldsymbol{z} \in Z$ *and* $\boldsymbol{h}^{cv}$ *is convex on* Z.

2. $\boldsymbol{h}^{cc} : Z \to \mathbb{R}^m$ *is a* concave relaxation *of* $\boldsymbol{h}$ *on* Z *if* $\boldsymbol{h}^{cc}(\boldsymbol{z}) \geq \boldsymbol{h}(\boldsymbol{z})$ *for all* $\boldsymbol{z} \in Z$ *and* $\boldsymbol{h}^{cc}$ *is concave on* Z.

3. *The interval function* $H \equiv [\boldsymbol{h}^{cv}, \boldsymbol{h}^{cc}]$ *is a* convex inclusion function *of* $\boldsymbol{h}$ *on* Z.

Definition 5.4 (Flattening operators [118])**.** *For each* $i \in \{1, \ldots, n\}$, *define flattening operators* $B_i^L, B_i^U : \mathbb{IR}^n \to \mathbb{IR}^n$ *such that,*

1. $B_i^L([\boldsymbol{\phi}, \boldsymbol{\psi}]) = [\boldsymbol{\phi}, \boldsymbol{\psi}']$, *where* $\psi_i' = \phi_i$, *and* $\boldsymbol{\psi}'_{-i} = \boldsymbol{\psi}_{-i}$,

2. $B_i^U([\boldsymbol{\phi}, \boldsymbol{\psi}]) = [\boldsymbol{\phi}', \boldsymbol{\psi}]$, *where* $\phi_i' = \psi_i$, *and* $\boldsymbol{\phi}'_{-i} = \boldsymbol{\phi}_{-i}$.

5.2.2 Enclosing trajectories

Based on the bounding methods in [120], [113], and [130], we develop a new general result for bounding trajectories using ODEs. While those established results provide sufficient conditions for bounding ODEs only, our new result can be used to bound any continuous differentiable trajectories such as solutions of ODEs and DAEs, or system trajectories of optimal control problems. Moreover, weaker assumptions are required in our new result.

Definition 5.5 (Enclosing dynamics)**.** *Consider* $t_0, t_f \in \mathbb{R}$ *with* $t_0 < t_f$, *and define* $I := [t_0, t_f]$. *Consider arbitrary continuously differentiable functions* $\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger : I \to \mathbb{R}^n$ *such that* $\boldsymbol{\xi}\dagger(t) \leq \boldsymbol{\xi}^\ddagger(t)$ *for all* $t \in I$. *Functions* $\boldsymbol{h}^L, \boldsymbol{h}^U : I \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ *describe* enclosing dynamics *about* $[\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger]$ *if the following holds: For a.e.* $t \in I$, $\boldsymbol{z}^L, \boldsymbol{z}^U \in \mathbb{R}^n$ *such that* $\boldsymbol{z}^L \leq \boldsymbol{\xi}^\dagger(t) \leq \boldsymbol{\xi}^\ddagger(t) \leq \boldsymbol{z}^U$,

1. *If* $z_i^L = \xi_i^\dagger(t)$, *then* $h_i^L(t, \boldsymbol{z}^L, \boldsymbol{z}^U) \leq \dot{\xi}_i^\dagger(t)$,

2. *If* $z_i^U = \xi_i^\ddagger(t)$, *then* $h_i^U(t, \boldsymbol{z}^L, \boldsymbol{z}^U) \geq \dot{\xi}_i^\ddagger(t)$.

Definition 5.5 is modified from [120, Definition 6], [118, Theorem 2] and [130, Theorem 2]. We use time derivatives $\dot{\xi}_i^\dagger, \dot{\xi}_i^\ddagger$ to replace the ODE RHS functions used in the original definitions and theorems. This modification not only reduces a monotonicity requirement on the ODE RHS functions, but also makes this result applicable to trajectories other than solution of ODEs.

Theorem 5.1. *Consider arbitrary continuously differentiable functions* $\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger : I \to \mathbb{R}^n$ *such that* $\boldsymbol{\xi}\dagger(t) \leq \boldsymbol{\xi}^\ddagger(t)$ *for all* $t \in I$. *Define* $\boldsymbol{\xi}_0^L, \boldsymbol{\xi}_0^U \in \mathbb{R}^n$ *and continuous functions* $\boldsymbol{h}^L, \boldsymbol{h}^U : I \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$. *Let* $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U)$ *solves the following ODEs:*

$$\begin{aligned} \dot{\boldsymbol{\xi}}^L(t) &= \boldsymbol{h}^L(t, \boldsymbol{\xi}^L(t), \boldsymbol{\xi}^U(t)), \quad \boldsymbol{\xi}^L(t_0) = \boldsymbol{\xi}_0^L, \\ \dot{\boldsymbol{\xi}}^U(t) &= \boldsymbol{h}^U(t, \boldsymbol{\xi}^L(t), \boldsymbol{\xi}^U(t)), \quad \boldsymbol{\xi}^U(t_0) = \boldsymbol{\xi}_0^U. \end{aligned} \tag{5.1}$$

If the following holds:

1. *There exists* $k^L \in \mathbb{R}_{>0}$ *such that, for any* $i \in \{1, \ldots, n\}$, *a.e.* $t \in I$, *and any* $\boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger \in \mathbb{R}^n$ *for which* $\boldsymbol{\phi}^\ddagger \leq \boldsymbol{\phi}^\dagger \leq \boldsymbol{\psi}^\dagger \leq \boldsymbol{\psi}^\ddagger$,

$$\begin{aligned} h_i^L(t, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) - h_i^L(t, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) &\leq k^L(\|\boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger\|_\infty + \|\boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger\|_\infty), \\ h_i^U(t, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) - h_i^U(t, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) &\leq k^L(\|\boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger\|_\infty + \|\boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger\|_\infty). \end{aligned}$$

2. $\boldsymbol{h}^L, \boldsymbol{h}^U$ *describe enclosing dynamics about* $[\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger]$,

3. $\boldsymbol{\xi}_0^L \leq \boldsymbol{\xi}^\dagger(t_0)$ *and* $\boldsymbol{\xi}^\ddagger(t_0) \leq \boldsymbol{\xi}_0^U$,

then

$$\boldsymbol{\xi}^L(t) \leq \boldsymbol{\xi}^\dagger(t) \leq \boldsymbol{\xi}^\ddagger(t) \leq \boldsymbol{\xi}^U(t), \quad \forall t \in I.$$

Proof. Since $\boldsymbol{h}^L, \boldsymbol{h}^U$ are continuous, (5.1) is guaranteed to have a solution by the Peano's existence theorem. The one-sided Lipschitz continuity of $\boldsymbol{h}^L, \boldsymbol{h}^U$ ensures the uniqueness of such a solution $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U)$ on I [148, p. 88].

Next, we proceed with a similar strategy used in the proves of [118, Theorem 2] and [130, Theorem 1]. Since $\boldsymbol{\xi}^\dagger \leq \boldsymbol{\xi}^\ddagger$ for all $t \in I$, it suffices to show that for all $t \in I$,

$$\boldsymbol{\xi}^L(t) \leq \boldsymbol{\xi}^\dagger(t) \quad \text{and} \quad \boldsymbol{\xi}^\ddagger(t) \leq \boldsymbol{\xi}^U(t). \tag{5.2}$$

By Condition 3, (5.2) holds at $t := t_0$. We will verify that (5.2) holds for all $t \in (t_0, t_f]$. To arrive at a contradiction, suppose there exist a $\tilde{t} \in (t_0, t_f]$ and $i \in \{1, \ldots, n\}$ such that

$$\xi_i^L(\tilde{t}) > \xi_i^\dagger(\tilde{t}) \quad \text{or} \quad \xi_i^\ddagger(\tilde{t}) < \xi_i^U(\tilde{t}).$$

Define

$$t_1 := \inf\{t \in (t_0, t_f] : \exists i \in \{1, \ldots, n\} \text{ such that } \xi_i^L(t) > \xi_i^\dagger(t) \text{ or } \xi_i^\ddagger(t) < \xi_i^U(t)\}$$

Define a function $\boldsymbol{\delta} : I \to \mathbb{R}^{2n}$ such that for each $t \in I$,

$$\boldsymbol{\delta}(t) := (\boldsymbol{\xi}^L(t) - \boldsymbol{\xi}^\dagger(t), \boldsymbol{\xi}^\ddagger - \boldsymbol{\xi}^U(t)).$$

Since $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U)$ solves (5.1) on I, $\boldsymbol{\xi}^L, \boldsymbol{\xi}^U$ are absolutely continuous. Thus, $\boldsymbol{\delta}$ is also absolutely continuous on I. A contradiction will be developed following Lemma 2 and 3 in [120].

Define functions $\phi, \psi : I \to \mathbb{R}^n$ such that for each $t \in I$ and $i \in \{1, \ldots, n\}$,

$$\phi_i(t) := \min\{\xi_i^\dagger(t), \xi_i^L(t)\}, \quad \text{and} \quad \psi_i(t) := \max\{\xi_i^\ddagger(t), \xi_i^U(t)\}.$$

Let $\mathbf{1} \in \mathbb{R}^n$ be a constant vector whose components are 1. It holds that $t_1 < t_f$, and for any $t_4 \in (t_1, t_f]$, there exist $j \in \{1, \ldots, n\}$, $\epsilon \in \mathbb{R}_{>0}$, and 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

$$0 < \rho(t) \leq \epsilon, \quad \forall t \in [t_1, t_4], \tag{5.3}$$

$$\dot{\rho}(t) > k^L \rho(t), \quad \text{a.e. } t \in [t_1, t_4], \tag{5.4}$$

$$\boldsymbol{\xi}^L(t) - \rho(t)\mathbf{1} < \boldsymbol{\xi}^\dagger(t), \quad \boldsymbol{\xi}^\ddagger(t) < \boldsymbol{\xi}^U(t) + \rho(t)\mathbf{1}, \quad \forall t \in [t_2, t_3), \tag{5.5}$$

and

$$\xi_j^\dagger(t_2) = \xi_j^L(t_2), \tag{5.6}$$

$$\xi_j^\dagger(t_3) = \xi_j^L(t_3) - \rho(t_3), \tag{5.7}$$

$$\xi_j^\dagger(t) < \xi_j^L(t), \quad \forall t \in (t_2, t_3). \tag{5.8}$$

or

$$\xi_j^\ddagger(t_2) = \xi_j^U(t_2),$$

$$\xi_j^\ddagger(t_3) = \xi_j^U(t_3) + \rho(t_3),$$

$$\xi_j^\ddagger(t) > \xi_j^U(t), \quad \forall t \in (t_2, t_3).$$

We proceed by assuming that (5.6)-(5.8) hold; the proof is analogous if the conditions on the right-hand side hold.

(5.6)-(5.8) show that $\xi_j^\dagger(t) \leq \xi_j^L(t)$ for all $t \in [t_2, t_3)$, and thus

$$\phi_j(t) = \xi_j^\dagger(t), \quad \forall t \in [t_2, t_3). \tag{5.9}$$

By the definition of ϕ and ψ,

$$\phi(t) \leq \boldsymbol{\xi}^\dagger(t) \leq \boldsymbol{\xi}^\ddagger(t) \leq \psi(t), \quad \forall t \in [t_2, t_3). \tag{5.10}$$

According to Condition 2, (5.9) and (5.10) show that

$$\dot{\xi}_i^\dagger(t) \geq h_j^L(t, \phi(t), \psi(t)), \quad \text{a.e. } t \in [t_2, t_3). \tag{5.11}$$

For each $k \in \{1, \ldots, n\}$ and each $t \in [t_2, t_3)$, one of the following cases will occur:

1. If $\xi_k^\dagger(t) \geq \xi_k^L(t)$ and $\xi_k^\ddagger(t) \leq \xi_k^U(t)$, then $\phi_k(t) = \xi_k^L(t)$ and $\psi_k(t) = \xi_k^U(t)$, which is

$$\phi_k(t) - \xi_k^L(t) = 0 \quad \text{and} \quad \psi_k(t) - \xi_k^U(t) = 0.$$

2. If $\xi_k^\dagger(t) < \xi_k^L(t)$ and $\xi_k^\ddagger(t) \leq \xi_k^U(t)$, then $\phi_k(t) = \xi_k^\dagger(t)$ and $\psi_k(t) = \xi_k^U(t)$. Combining with (5.5),

$$0 < \xi_k^L(t) - \phi_k(t) < \rho(t) \quad \text{and} \quad \psi_k(t) - \xi_k^U(t) = 0.$$

3. If $\xi_k^\dagger(t) \geq \xi_k^L(t)$ and $\xi_k^\ddagger(t) < \xi_k^U(t)$, then $\phi_k(t) = \xi_k^L(t)$ and $\psi_k(t) = \xi_k^\ddagger(t)$.

Combining with (5.5),

$$\phi_k(t) - \xi_k^L(t) = 0 \quad \text{and} \quad 0 < \psi_k(t) - \xi_k^U(t) < \rho(t).$$

4. If $\xi_k^\dagger(t) < \xi_k^L(t)$ and $\xi_k^\ddagger(t) < \xi_k^U(t)$, then $\phi_k(t) = \xi_k^\dagger(t)$ and $\psi_k(t) = \xi_k^\ddagger(t)$. Combining with (5.5),

$$0 < \xi_k^L(t) - \phi_k(t) < \rho(t) \quad \text{and} \quad 0 < \psi_k(t) - \xi_k^U(t) < \rho(t).$$

The above four cases ensure that

$$(\|\boldsymbol{\phi}(t) - \boldsymbol{\xi}^L(t)\|_\infty + \|\boldsymbol{\psi}(t) - \boldsymbol{\xi}^U(t)\|_\infty) < \rho(t), \quad \forall t \in [t_2, t_3). \tag{5.12}$$

Condition 1 and (5.12) show that

$$\begin{aligned} h_j^L(t, \boldsymbol{\xi}^L(t), \boldsymbol{\xi}^U(t)) &\leq h_j^L(t, \boldsymbol{\phi}(t), \boldsymbol{\psi}(t)) + k^L(\|\boldsymbol{\phi}(t) - \boldsymbol{\xi}^L(t)\|_\infty + \|\boldsymbol{\psi}(t) - \boldsymbol{\xi}^U(t)\|_\infty) \\ &< h_j^L(t, \boldsymbol{\phi}(t), \boldsymbol{\psi}(t)) + k^L\rho(t), \quad \forall t \in [t_2, t_3). \end{aligned} \tag{5.13}$$

Combining (5.11) and (5.13),

$$h_j^L(t, \boldsymbol{\xi}^L(t), \boldsymbol{\xi}^U(t)) < \dot{\xi}_i^\dagger(t) + k^L\rho(t), \quad \text{a.e. } t \in [t_2, t_3). \tag{5.14}$$

(5.4) shows that

$$\dot{\rho}(t) > k^L\rho(t), \quad \text{a.e. } t \in [t_2, t_3].$$

(5.14) becomes

$$h_j^L(t, \boldsymbol{\xi}^L(t), \boldsymbol{\xi}^U(t)) - \dot{\xi}_i^\dagger(t) - \dot{\rho}(t) < 0, \quad \text{a.e. } t \in [t_2, t_3]. \tag{5.15}$$

According to Theorem 3.1 in [137], function $(\xi_j^L(t) - \xi_j^\dagger(t) - \rho(t))$ is decreasing with respect to t on $[t_2, t_3]$. Thus,

$$\xi_j^L(t_3) - \xi_j^\dagger(t_3) - \rho(t_3) < \xi_j^L(t_2) - \xi_j^\dagger(t_2) - \rho(t_2). \tag{5.16}$$

However, (5.6) and (5.7) show that

$$\begin{aligned} \xi_j^L(t_2) - \xi_j^\dagger(t_2) &= 0, \\ \xi_j^L(t_3) - \xi_j^\dagger(t_3) - \rho(t_3) &= 0. \end{aligned}$$

Substituting into (5.16) yields $0 > \rho(t_2)$, which contradicts (5.3). □

Remark 5.1. *A single trajectory $\boldsymbol{\xi} : I \to \mathbb{R}^n$ can be bounded by* (5.1) *if we let $\boldsymbol{\xi}^\dagger \equiv \boldsymbol{\xi}^\ddagger \equiv \boldsymbol{\xi}$. For simplicity, we extend Definition 5.5 so that $\boldsymbol{h}^L, \boldsymbol{h}^U$ describing enclosing dynamics about $\boldsymbol{\xi}$ is interpreted as $\boldsymbol{h}^L, \boldsymbol{h}^U$ describing enclosing dynamics about $[\boldsymbol{\xi}, \boldsymbol{\xi}]$.*

5.3 Problem Statement

Let $U := [\boldsymbol{u}^L, \boldsymbol{u}^U] \subset \mathbb{R}^{n_u}$ be an interval, and $D \subset \mathbb{R}^{n_x}$ be open. Denote the space of all Lebesgue integrable functions $\boldsymbol{h} : I \to \mathbb{R}^n$ as $L^n(I)$. Let $\mathcal{U} := \{\boldsymbol{u} \in L^{n_u}(I) : \boldsymbol{u}(t) \in U, t \in I\}$ be a set of admissible controls, and $X_0 \equiv [\boldsymbol{x}_0^L, \boldsymbol{x}_0^U] \in D$ be a set of admissible initial conditions. Given a locally Lipschitz continuous function

$\boldsymbol{f} : I \times U \times D \to \mathbb{R}^{n_x}$, consider an initial-value problem

$$\begin{aligned} \dot{\boldsymbol{x}}(t, \boldsymbol{u}, \boldsymbol{x}_0) &= \boldsymbol{f}(t, \boldsymbol{u}(t), \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0)), \quad \forall t \in (t_0, t_f], \\ \boldsymbol{x}(t_0, \boldsymbol{u}, \boldsymbol{x}_0) &= \boldsymbol{x}_0, \end{aligned} \tag{5.17}$$

where $(\boldsymbol{u}, \boldsymbol{x}_0) \in \mathcal{U} \times X_0$. The local Lipschitz continuity of $\boldsymbol{f}$ implies that the ODE (5.17) is guaranteed to have a unique solution by the Picard-Lindelöf Theorem (Theorem 1.1, Chapter II in [60]). Moreover, the ODE solution $\boldsymbol{x}$ is continuously differentiable on I [148, p. 39].

This work concerns about enclosing the reachable set of (5.17) with *state bounds*, which are time-varying interval bounds of the state variables.

Definition 5.6 (State bounds [120, 113])**.** *Functions $\boldsymbol{x}^L, \boldsymbol{x}^U : I \to \mathbb{R}^{n_x}$ are* state bounds *for* (5.17) *if, for each $t \in I$, $\boldsymbol{u} \in \mathcal{U}$, and $\boldsymbol{x}_0 \in X_0$,*

$$\boldsymbol{x}^L(t) \leq \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \leq \boldsymbol{x}^U(t).$$

Let $X^B : I \to \mathbb{IR}^{n_x}$ denote the corresponding inclusion function: $X^B(t) \equiv [\boldsymbol{x}^L(t), \boldsymbol{x}^U(t)]$ for each $t \in I$.

5.4 New Framework for Enclosing Reachable Sets

This section introduces a novel framework for constructing state bounds of (5.17).

Assumption 5.1. *Assume that functions $\boldsymbol{f}^L, \boldsymbol{f}^U : I \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ satisfy the following conditions:*

1. *$\boldsymbol{f}^L$ and $\boldsymbol{f}^U$ are continuous,*

2. $\boldsymbol{f}^L(t,\boldsymbol{p},\cdot,\cdot)$ *and* $\boldsymbol{f}^U(t,\boldsymbol{p},\cdot,\cdot)$ *are Lipschitz continuous on* $\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$*, uniformly in* $(t,\boldsymbol{p})$*,*

3. $(t,\boldsymbol{\xi}^L,\boldsymbol{\xi}^U)\mapsto \boldsymbol{f}^L(t,\boldsymbol{u}(t),\boldsymbol{\xi}^L,\boldsymbol{\xi}^U)$ *and* $(t,\boldsymbol{\xi}^L,\boldsymbol{\xi}^U)\mapsto \boldsymbol{f}^U(t,\boldsymbol{u}(t),\boldsymbol{\xi}^L,\boldsymbol{\xi}^U)$ *describe enclosing dynamics about* $t\mapsto \boldsymbol{x}(t,\boldsymbol{u},\boldsymbol{x}_0)$ *for all* $\boldsymbol{u}\in\mathcal{U}$ *and* $\boldsymbol{x}_0\in X_0$*.*

Note that an inclusion function of $\boldsymbol{f}(t,\boldsymbol{p},\cdot)$ for all $(t,\boldsymbol{p})\in I\times U$ describes enclosing dynamics about $t\mapsto \boldsymbol{x}(t,\boldsymbol{u},\boldsymbol{x}_0)$. However, to construct tighter state bounds, the flattening operation in Definition 5.4 are usually applied to the inclusion function [59, 118]. The generated new function also describes enclosing dynamics about $t\mapsto \boldsymbol{x}(t,\boldsymbol{u},\boldsymbol{x}_0)$. Numerous methods for deriving $\boldsymbol{f}^L,\boldsymbol{f}^U$ are introduced in Section 5.5, which typically involves the flattening operation.

Under Assumption 5.1, consider the following auxiliary ODE system: for each $i\in\{1,\ldots,n_x\}$,

$$\begin{aligned}
\dot{x}_i^L(t) &= \min_{\boldsymbol{p}\in U} f_i^L(t,\boldsymbol{p},\boldsymbol{x}^L(t),\boldsymbol{x}^U(t)), \quad x_i^L(t_0)=x_{0,i}^L,\\
\dot{x}_i^U(t) &= \max_{\boldsymbol{p}\in U} f_i^U(t,\boldsymbol{p},\boldsymbol{x}^L(t),\boldsymbol{x}^U(t)), \quad x_i^U(t_0)=x_{0,i}^U.
\end{aligned} \tag{5.18}$$

It will be shown in the remainder of this section that the solution of (5.18) provides valid state bounds of (5.17).

5.4.1 Existence and uniqueness

First, we verify that the ODE system (5.18) has exactly one solution under Assumption 5.1.

Lemma 5.1. *Under Assumption 5.1, define function* $\boldsymbol{g}^L,\boldsymbol{g}^U : I\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}\to\mathbb{R}^{n_x}$

such that, for each $i \in \{1, \ldots, n_x\}$,

$$g_i^L(t, \boldsymbol{\phi}, \boldsymbol{\psi}) = \min_{\boldsymbol{p} \in U} f_i^L(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}),$$
$$g_i^U(t, \boldsymbol{\phi}, \boldsymbol{\psi}) = \max_{\boldsymbol{p} \in U} f_i^U(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi}).$$

Then, $\boldsymbol{g}^L(t, \boldsymbol{p}, \cdot, \cdot), \boldsymbol{g}^U(t, \boldsymbol{p}, \cdot, \cdot)$ *are Lipschitz continuous on* $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ *for all* $(t, \boldsymbol{p}) \in I \times U$.

Proof. Consider any $i \in \{1, \ldots, n_x\}$, $t \in I$, $\boldsymbol{p} \in U$, and compact set $S \subset \mathbb{R}^{n_x}$. Condition 2 in Assumption 5.1 ensures that, for any $\boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger, \boldsymbol{\phi}^\ddagger, \boldsymbol{\phi}^\ddagger \in S$, there exists $k^S \in \mathbb{R}_{>0}$ such that

$$f_i^L(t, \boldsymbol{p}, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) - f_i^L(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) \leq k^S \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\|_\infty + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\|_\infty \right).$$
$$f_i^U(t, \boldsymbol{p}, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) - f_i^U(t, \boldsymbol{p}, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) \leq k^S \left(\left\| \boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger \right\|_\infty + \left\| \boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger \right\|_\infty \right).$$

Define set-valued mappings $\omega^L, \omega^U : I \times S \times S \rightrightarrows U$ such that

$$\omega^L(t, \boldsymbol{\phi}, \boldsymbol{\psi}) := \{\boldsymbol{p} \in U : g_i^L(t, \boldsymbol{\phi}, \boldsymbol{\psi}) = f_i^L(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi})\},$$
$$\omega^U(t, \boldsymbol{\phi}, \boldsymbol{\psi}) := \{\boldsymbol{p} \in U : g_i^U(t, \boldsymbol{\phi}, \boldsymbol{\psi}) = f_i^U(t, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\psi})\}.$$

Since $\boldsymbol{f}^L$, $\boldsymbol{f}^U$ are continuous and U is compact, $\omega^L(t, \boldsymbol{\phi}, \boldsymbol{\psi}), \omega^U(t, \boldsymbol{\phi}, \boldsymbol{\psi})$ are nonempty

for all $t \in I$, and $\boldsymbol{\phi}, \boldsymbol{\psi} \in S$. Consider any $\boldsymbol{p}^{\ddagger} \in \omega^L(t, \boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger})$ and $\boldsymbol{p}^{\dagger} \in \omega^U(t, \boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger})$.

$$\begin{aligned} g_i^L(t, \boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}) &= \min_{\boldsymbol{p} \in U} f_i^L(t, \boldsymbol{p}, \boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}) \\ &\leq f_i^L(t, \boldsymbol{p}^{\ddagger}, \boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}) \\ &\leq f_i^L(t, \boldsymbol{p}^{\ddagger}, \boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}) + k^S \left(\left\| \boldsymbol{\phi}^{\dagger} - \boldsymbol{\phi}^{\ddagger} \right\|_{\infty} + \left\| \boldsymbol{\psi}^{\dagger} - \boldsymbol{\psi}^{\ddagger} \right\|_{\infty} \right) \\ &= g_i^L(t, \boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}) + k^S \left(\left\| \boldsymbol{\phi}^{\dagger} - \boldsymbol{\phi}^{\ddagger} \right\|_{\infty} + \left\| \boldsymbol{\psi}^{\dagger} - \boldsymbol{\psi}^{\ddagger} \right\|_{\infty} \right). \end{aligned}$$

Similarly,

$$\begin{aligned} g_i^U(t, \boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}) &= f_i^U(t, \boldsymbol{p}^{\dagger}, \boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}) \\ &\leq f_i^U(t, \boldsymbol{p}^{\dagger}, \boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}) + k^S \left(\left\| \boldsymbol{\phi}^{\dagger} - \boldsymbol{\phi}^{\ddagger} \right\|_{\infty} + \left\| \boldsymbol{\psi}^{\dagger} - \boldsymbol{\psi}^{\ddagger} \right\|_{\infty} \right) \\ &\leq \max_{\boldsymbol{p} \in U} f_i^U(t, \boldsymbol{p}, \boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}) + k^S \left(\left\| \boldsymbol{\phi}^{\dagger} - \boldsymbol{\phi}^{\ddagger} \right\|_{\infty} + \left\| \boldsymbol{\psi}^{\dagger} - \boldsymbol{\psi}^{\ddagger} \right\|_{\infty} \right) \\ &= g_i^U(t, \boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}) + k^S \left(\left\| \boldsymbol{\phi}^{\dagger} - \boldsymbol{\phi}^{\ddagger} \right\|_{\infty} + \left\| \boldsymbol{\psi}^{\dagger} - \boldsymbol{\psi}^{\ddagger} \right\|_{\infty} \right). \end{aligned}$$

Since i is arbitrarily selected,

$$\begin{aligned} \left\| \boldsymbol{g}^L(t, \boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}) - \boldsymbol{g}^L(t, \boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}) \right\|_{\infty} &\leq k^S \left(\left\| \boldsymbol{\phi}^{\dagger} - \boldsymbol{\phi}^{\ddagger} \right\|_{\infty} + \left\| \boldsymbol{\psi}^{\dagger} - \boldsymbol{\psi}^{\ddagger} \right\|_{\infty} \right), \\ \left\| \boldsymbol{g}^U(t, \boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}) - \boldsymbol{g}^U(t, \boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}) \right\|_{\infty} &\leq k^S \left(\left\| \boldsymbol{\phi}^{\dagger} - \boldsymbol{\phi}^{\ddagger} \right\|_{\infty} + \left\| \boldsymbol{\psi}^{\dagger} - \boldsymbol{\psi}^{\ddagger} \right\|_{\infty} \right). \end{aligned}$$

Thus, $\boldsymbol{g}^L, \boldsymbol{g}^U$ are Lipschitz continuous with respect to $\boldsymbol{\phi}, \boldsymbol{\psi}$. Since t and $\boldsymbol{p}$ were chosen arbitrarily, the desired result holds. □

Theorem 5.2. *Under Assumption 5.1,* (5.18) *has a unique solution.*

Proof. Condition 1 of Assumption 5.1 shows that $\boldsymbol{f}^L, \boldsymbol{f}^U$ are continuous. The Maximum Theorem [65, Theorem 3.4] ensures that the RHS functions of (5.18) are continuous. Thus, (5.18) has at least one solution according to Peano's existence theorem as summarized in [60, Theorem 2.1].

Lemma 5.1 ensures that the RHS functions of (5.1) are Lipschitz continuous with respect to state variables. According to the uniqueness theorem in [148, Chapter II, Section 10], (5.18) has at most one solution.

Combining the two results above, (5.18) has a unique solution. □

5.4.2 Bounding the original system

Theorem 5.3. *Under Assumptions 5.1, let* $(\boldsymbol{x}^L, \boldsymbol{x}^U)$ *be a solution of* (5.18)*. Then,* $\boldsymbol{x}^L, \boldsymbol{x}^U$ *are state bounds of* (5.17)*.*

Proof. It suffices to show that the requirements in Theorem 5.1 are satisfied with $\boldsymbol{x}^L, \boldsymbol{x}^U$ in place of $\boldsymbol{\xi}^L, \boldsymbol{\xi}^U$ and $t \mapsto \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0)$ in place of $\boldsymbol{\xi}^\dagger$ and $\boldsymbol{\xi}^\ddagger$, respectively. Assumption 5.1 and Lemma 5.1 ensure the RHS functions of (5.18) are Lipschitz continuous with respect to state variables, which is sufficient for the first requirement. Since $\boldsymbol{x}_0 \in X_0$, third requirement is satisfied.

Next, we verify the enclosing dynamics in the second requirement. It suffices to show that, for a.e. $t \in I$, any $i \in \{1, \ldots, n_x\}$, $\boldsymbol{u} \in \mathcal{U}$, $\boldsymbol{x}_0 \in X_0$, and $\boldsymbol{z}^L, \boldsymbol{z}^U \in \mathbb{R}^{n_x}$ such that $\boldsymbol{z}^L \leq \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \leq \boldsymbol{z}^U$,

1. If $z_i^L = x_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$, then $\min_{\boldsymbol{p} \in U} f_i^L(t, \boldsymbol{p}, \boldsymbol{z}^L, \boldsymbol{z}^U) \leq \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$.

2. If $z_i^U = x_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$, then $\max_{\boldsymbol{p} \in U} f_i^U(t, \boldsymbol{p}, \boldsymbol{z}^L, \boldsymbol{z}^U) \geq \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$.

We will show that the first condition holds. It is analogous to verify the second. Condition 3 of Assumption 5.1 ensures that, if $z_i^L = x_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$, then

$$f_i^L(t, \boldsymbol{u}(t), \boldsymbol{z}^L, \boldsymbol{z}^U) \leq \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0).$$

Thus,

$$\min_{\boldsymbol{p} \in U} f_i^L(t, \boldsymbol{p}, \boldsymbol{z}^L, \boldsymbol{z}^U) \leq f_i^L(t, \boldsymbol{u}(t), \boldsymbol{z}^L, \boldsymbol{z}^U) \leq \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0),$$

which ensures the first condition.

Hence, all three requirements in Theorem 5.1 are satisfied. □

Note that since $\boldsymbol{x}^L, \boldsymbol{x}^U$ are state bounds of (5.17), it is ensured that $\boldsymbol{x}^L(t) \leq \boldsymbol{x}^U(t)$ for all $t \in I$. Following Definition 5.6, they form an interval function $X^B(t) \equiv [\boldsymbol{x}^L(t), \boldsymbol{x}^U(t)]$. This result will be used implicitly in the remainder of this article.

5.4.3 Comparison of tightness

This subsection describes a general result showing that if the ODE system (5.18) has tighter RHS functions, then it generates tighter state bounds. It is developed based on [130, Theorem 2], but requires weaker assumptions.

Assumption 5.2. *Assume that functions $\boldsymbol{f}^{L,\dagger}, \boldsymbol{f}^{U,\dagger}$ and $\boldsymbol{f}^{L,\ddagger}, \boldsymbol{f}^{U,\ddagger} : I \times \mathbb{R}^{n_u} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ satisfy Assumption 5.1 and the following conditions: for any $\boldsymbol{p} \in U$, $t \in I$, $\Xi^\dagger \equiv [\boldsymbol{\xi}^{L,\dagger}, \boldsymbol{\xi}^{U,\dagger}] \subseteq \Xi^\ddagger \equiv [\boldsymbol{\xi}^{L,\ddagger}, \boldsymbol{\xi}^{U,\ddagger}] \in \mathbb{IR}^{n_x}$, and $i \in \{1, \ldots, n_x\}$,*

1. *if $\xi_i^{L,\dagger} = \xi_i^{L,\ddagger}$, then $f_i^{L,\ddagger}(t, \boldsymbol{p}, \Xi^\ddagger) \leq f_i^{L,\dagger}(t, \boldsymbol{p}, \Xi^\dagger)$,*
2. *if $\xi_i^{U,\dagger} = \xi_i^{U,\ddagger}$, then $f_i^{U,\ddagger}(t, \boldsymbol{p}, \Xi^\ddagger) \geq f_i^{U,\dagger}(t, \boldsymbol{p}, \Xi^\dagger)$.*

Consider the following two ODE systems that are similar to (5.18):

$$\begin{aligned}
\dot{\boldsymbol{x}}^{L,\dagger}(t) &= \min_{\boldsymbol{p}\in U} \boldsymbol{f}^{L,\dagger}(t,\boldsymbol{p},X^{B,\dagger}(t)), \quad \boldsymbol{x}^{L,\dagger}(t_0) = \boldsymbol{x}_0^L, \\
\dot{\boldsymbol{x}}^{U,\dagger}(t) &= \max_{\boldsymbol{p}\in U} \boldsymbol{f}^{U,\dagger}(t,\boldsymbol{p},X^{B,\dagger}(t)), \quad \boldsymbol{x}^{U,\dagger}(t_0) = \boldsymbol{x}_0^U,
\end{aligned} \tag{5.19}$$

and

$$\begin{aligned}
\dot{\boldsymbol{x}}^{L,\ddagger}(t) &= \min_{\boldsymbol{p}\in U} \boldsymbol{f}^{L,\ddagger}(t,\boldsymbol{p},X^{B,\ddagger}(t)), \quad \boldsymbol{x}^{L,\ddagger}(t_0) = \boldsymbol{x}_0^L, \\
\dot{\boldsymbol{x}}^{U,\ddagger}(t) &= \max_{\boldsymbol{p}\in U} \boldsymbol{f}^{U,\ddagger}(t,\boldsymbol{p},X^{B,\ddagger}(t)), \quad \boldsymbol{x}^{U,\ddagger}(t_0) = \boldsymbol{x}_0^U.
\end{aligned} \tag{5.20}$$

Theorem 5.4. *Under Assumption 5.2, let* $(\boldsymbol{x}^{L,\dagger},\boldsymbol{x}^{U,\dagger})$ *and* $(\boldsymbol{x}^{L,\ddagger},\boldsymbol{x}^{U,\ddagger})$ *be solutions of* (5.19) *and* (5.20) *on* $I\times U$*, respectively. Then,* $X^{B,\dagger}(t)\equiv[\boldsymbol{x}^{L,\dagger}(t),\boldsymbol{x}^{U,\dagger}(t)]\subseteq X^{B,\ddagger}(t)\equiv[\boldsymbol{x}^{L,\ddagger}(t),\boldsymbol{x}^{U,\ddagger}(t)]$ *for all* $t\in I$*.*

Proof. This theorem can be proved by showing that all three requirements in Theorem 5.1 are satisfied with $\boldsymbol{x}^{L,\dagger},\boldsymbol{x}^{U,\dagger}$ in place of $\boldsymbol{\xi}^\dagger,\boldsymbol{\xi}^\dagger$ and $\boldsymbol{x}^{L,\ddagger},\boldsymbol{x}^{U,\ddagger}$ in place of $\boldsymbol{\xi}^L,\boldsymbol{\xi}^U$, respectively. Assumption 5.1 and Lemma 5.1 guarantee that the RHS functions of (5.19) and (5.20) are Lipschitz continuous with respect to state variables, which ensures the first requirement. Since (5.19) and (5.20) share the same initial condition, the third requirement is satisfied.

Next, we verify the second requirement by showing that $\boldsymbol{f}^{L,\ddagger},\boldsymbol{f}^{U,\ddagger}$ describe enclosing dynamics about $[\boldsymbol{x}^{L,\dagger},\boldsymbol{x}^{U,\dagger}]$. For a.e. $t\in I$, any $i\in\{1,\ldots,n_x\}$, and $\boldsymbol{z}^L,\boldsymbol{z}^U\in\mathbb{R}^{n_x}$ such that $\boldsymbol{z}^L\leq\boldsymbol{x}^{L,\dagger}\leq\boldsymbol{x}^{U,\dagger}\leq\boldsymbol{z}^U$, it suffices to show that,

1. If $z_i^L = x_i^{L,\dagger}(t)$, then $\min_{\boldsymbol{p}\in U} f_i^{L,\ddagger}(t,\boldsymbol{p},\boldsymbol{z}^L,\boldsymbol{z}^U)\leq\dot{x}_i^{L,\dagger}(t)$.
2. If $z_i^U = x_i^{U,\dagger}(t)$, then $\max_{\boldsymbol{p}\in U} f_i^{U,\ddagger}(t,\boldsymbol{p},\boldsymbol{z}^L,\boldsymbol{z}^U)\geq\dot{x}_i^{U,\dagger}(t)$.

It will be shown that the first condition holds; showing the second is analogous. Define $\boldsymbol{p}^* \in U$ such that

$$\boldsymbol{p}^* := \arg\min_{\boldsymbol{p}\in U} f_i^{L,\dagger}(t, \boldsymbol{p}, \boldsymbol{x}^{L,\dagger}(t), \boldsymbol{x}^{U,\dagger}(t)).$$

According to Assumption 5.2, if $z_i^L = x_i^{L,\dagger}(t)$, then

$$f_i^{L,\ddagger}(t, \boldsymbol{p}^*, \boldsymbol{z}^L, \boldsymbol{z}^U) \leq f_i^{L,\dagger}(t, \boldsymbol{p}^*, \boldsymbol{x}^{L,\dagger}(t), \boldsymbol{x}^{U,\dagger}(t)).$$

Thus, if $z_i^L = x_i^{L,\dagger}(t)$,

$$\begin{aligned}
\min_{\boldsymbol{p}\in U} f_i^{L,\ddagger}(t, \boldsymbol{p}, \boldsymbol{z}^L, \boldsymbol{z}^U) &\leq f_i^{L,\ddagger}(t, \boldsymbol{p}^*, \boldsymbol{z}^L, \boldsymbol{z}^U) \\
&\leq f_i^{L,\dagger}(t, \boldsymbol{p}^*, \boldsymbol{x}^{L,\dagger}(t), \boldsymbol{x}^{U,\dagger}(t)) \\
&= \min_{\boldsymbol{p}\in U} f_i^{L,\dagger}(t, \boldsymbol{p}, \boldsymbol{x}^{L,\dagger}(t), \boldsymbol{x}^{U,\dagger}(t)) \\
&= \dot{x}_i^{L,\dagger}(t),
\end{aligned}$$

which verifies the first condition.

Thus, all three requirements in Theorem 5.1 are satisfied. □

5.5 Use Cases

In this section, four use cases of our new framework (5.18) are presented. Each of them represents a strategy for constructing $\boldsymbol{f}^L, \boldsymbol{f}^U$ that satisfy Assumption 5.1. In every use case, multiple practical methods are introduced for generating functions $\boldsymbol{f}^L, \boldsymbol{f}^U$ from $\boldsymbol{f}$. Some of these methods have been established by other researchers,

but fall within our new framework; the rest are novel approaches discovered with this new framework. A preview of these use cases is presented in Table 5.1. It briefly describes the embedded optimization problem size in the RHS of (5.18) along with expected tightness of the generated state bounds. Note that some of the embedded optimization problems may be solved trivially without a numerical optimization solver. For example, if $\boldsymbol{f}^L, \boldsymbol{f}^U$ are constructed using interval extensions, then finding their minimum or maximum only involves choosing the lower or upper bound, respectively.

Table 5.1: Preview of use cases introduced in Section 5.5

Use case	Section	Opt variables	Computing effort	Tightness
Interval extension	5.5.1	0	Low	Loose
Optimize states	5.5.2	n_x	Medium	Moderate
Optimize parameters	5.5.3	n_u	Medium	Moderate
Optimize states and parameters	5.5.4	$n_x + n_u$	High	Tight

In addition to the flattening operation introduced previously, a methodology studied in [118, 62, 124] is also considered here to reduce the overestimation of state bounds. It utilizes the information from an *a priori* enclosure of the original system (5.17) to refine state bounds. Such an enclosure can be obtained from physical boundaries, conservative laws, or constraints in optimization problems [121, 62, 124].

Definition 5.7. *A set $G \subset I \times U \times \mathbb{R}^{n_x}$ is considered as an* a priori *enclosure of* (5.17) *if for all $(t, \boldsymbol{u}, \boldsymbol{x}_0) \in I \times \mathcal{U} \times X_0$, $(t, \boldsymbol{u}(t), \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0)) \in G$.*

To include such *a priori* knowledge into auxiliary system (5.18), interval operators $\Pi_i^L, \Pi_i^U : I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ are introduced for each $i \in \{1, \ldots, n_x\}$, and they satisfy the following assumption.

Assumption 5.3. *Assume that for every $i \in \{1, \ldots, n_x\}$, Π_i^L, Π_i^U satisfy the following:*

1. *For any $\Xi \equiv [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U] \in \mathbb{IR}^{n_x}$ and $(t, \boldsymbol{u}, \boldsymbol{x}_0) \in I \times \mathcal{U} \times X_0$ such that $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in \Xi$:*
   - *If $x_i(t, \boldsymbol{u}, \boldsymbol{x}_0) = \xi_i^L$, then $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in \Pi_i^L(t, U, \Xi)$,*
   - *If $x_i(t, \boldsymbol{u}, \boldsymbol{x}_0) = \xi_i^U$, then $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in \Pi_i^U(t, U, \Xi)$,*
2. *$(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \Pi_i^L(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ and $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \Pi_i^U(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ are Lipschitz continuous on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly continuous in t.*

Remark 5.2. *If Π_i^L, Π_i^U are independent of t and U, then we recovered the operators $\Omega_i^L, \Omega_i^U : \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ proposed in [118]. Ω_i^L and Ω_i^U are refinement operators that incorporate time-invariant and input-invariant* a priori *enclosures to construct tighter state bounds. Another pair of similar operators $\Phi_i^L, \Phi_i^U : \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_u} \to \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_u}$ were developed in [127]. They were designed to support parameter-dependent but time-invariant* a priori *enclosures.*

5.5.1 Interval extension

This subsection discusses a use case of the new framework where $\boldsymbol{f}^L, \boldsymbol{f}^U$ are constructed with interval extensions. The embedded optimization problems in (5.18) then can be solved trivially via choosing lower or upper interval bounds. The classic Harrison's method [59] that constructs interval extension using NIE is included in this category. Beside that, a novel technique that constructs interval extensions using MIE is also discussed.

Assumption 5.4. *Assume that interval function $\bar{F}^B = [\bar{\boldsymbol{f}}^L, \bar{\boldsymbol{f}}^U] : I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ satisfies the following conditions:*

1. $\bar{\boldsymbol{f}}^L$ *and* $\bar{\boldsymbol{f}}^U$ *are continuous,*

2. $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \bar{\boldsymbol{f}}^L(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ *and* $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \bar{\boldsymbol{f}}^U(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ *are Lipschitz continuous on* $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$*, uniformly in* t*,*

3. $\Xi \mapsto \bar{F}^B(t, U, \Xi)$ *is an inclusion function of* $\boldsymbol{f}(t, \boldsymbol{p}, \cdot)$ *on* D *for a.e.* $t \in I$ *and all* $\boldsymbol{p} \in U$*.*

In this use case, consider $\boldsymbol{f}^L, \boldsymbol{f}^U$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} f_i^L(t, \boldsymbol{p}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) &:= \bar{f}_i^L(t, U, \Pi_i^L(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])), \\ f_i^U(t, \boldsymbol{p}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) &:= \bar{f}_i^U(t, U, \Pi_i^U(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])), \end{aligned} \tag{5.21}$$

where Π_i^L, Π_i^U are operators satisfying Assumption 5.3. Since $\bar{f}_i^L, \bar{f}_i^U$ in (5.21) are interval extensions, the optimization problems embedded in the RHS of (5.18) can be solved trivially by choosing the lower or upper bound, respectively. Therefore, (5.18) becomes: for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} \dot{x}_i^L(t) &= \bar{f}_i^L(t, U, \Pi_i^L(t, U, X^B(t))), \quad x_i^L(t_0) = x_{0,i}^L, \\ \dot{x}_i^U(t) &= \bar{f}_i^U(t, U, \Pi_i^U(t, U, X^B(t))), \quad x_i^U(t_0) = x_{0,i}^U. \end{aligned} \tag{5.22}$$

Lemma 5.2. *Under Assumptions 5.3 and 5.4,* $\boldsymbol{f}^L, \boldsymbol{f}^U$ *in* (5.21) *satisfy Assumption 5.1.*

Proof. Condition 2 of Assumption 5.3 and Condition 1 of Assumption 5.4 guarantees the continuity requirement in Condition 1 of Assumption 5.1. Since the composite function of locally Lipschitz continuous functions is locally Lipschitz continuous [117, Theorem 2.5.6], Condition 2 of Assumption 5.3 and Condition 2 of Assumption 5.4 ensures in Condition 2 of Assumption 5.1.

Next, we demonstrate the enclosing dynamics required in Condition 3 of Assumption 5.1. Consider a.e. $t \in I$, any $\boldsymbol{u} \in \mathcal{U}$, $\boldsymbol{x}_0 \in X_0$, $i \in \{1, \ldots, n_x\}$, and $\Xi \equiv [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U] \in \mathbb{IR}^{n_x}$ such that $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in \Xi$. According to Definition 5.5, it is desired to show that,

1. If $\xi_i^L = x_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$, then $\bar{f}_i^L(t, U, \Pi_i^L(t, U, \Xi)) \leq \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$, and
2. If $\xi_i^U = x_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$, then $\bar{f}_i^U(t, U, \Pi_i^U(t, U, \Xi)) \geq \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$.

It will be shown that the first requirement holds; it is analogous to prove the second. If $\xi_i^L = x_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$, Assumption 5.3 ensures that $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in \Pi_i^L(t, U, \Xi)$. Condition 3 of Assumption 5.4 shows that

$$\bar{f}_i^L(t, U, \Pi_i^L(t, U, \Xi)) \leq f_i(t, \boldsymbol{u}(t), \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0)) = \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0),$$

which ensures the first requirement.

Therefore, all three conditions in Assumption 5.1 are satisfied. □

Natural interval extensions

Constructing $\bar{F}^B$ using NIE was proposed by Harrison [59]. $\bar{F}^B$ satisfies Assumption 5.4 according to [117, Section 2.5 and Section 2.3]. If operators Π_i^L, Π_i^U are defined as Ω_i^L, Ω_i^U in Remark 5.2 such that, for each $i \in \{1, \ldots, n_x\}$

$$\begin{aligned} \Pi_i^L(t, U, \Xi) &:= \Omega_i^L(\Xi), \\ \Pi_i^U(t, U, \Xi) &:= \Omega_i^U(\Xi), \end{aligned} \tag{5.23}$$

we recover a formulation proposed in [118]: for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} \dot{x}_i^L(t) &= \bar{f}_i^L(t, U, \Omega_i^L(X^B(t))), \quad x_i^L(t_0) = x_{0,i}^L, \\ \dot{x}_i^U(t) &= \bar{f}_i^U(t, U, \Omega_i^U(X^B(t))), \quad x_i^U(t_0) = x_{0,i}^U. \end{aligned} \tag{5.24}$$

Furthermore, if *a priori* enclosure is not considered, Π_i^L and Π_i^U may be set to the flattening operators in Definition 5.4:

$$\begin{aligned} \Pi_i^L(t, U, \Xi) &:= B_i^L(\Xi), \\ \Pi_i^U(t, U, \Xi) &:= B_i^U(\Xi). \end{aligned}$$

It is readily verified that operators B_i^L and B_i^U satisfy Assumption 5.3. Then, the classic Harrison's method [59] is recovered:

$$\begin{aligned} \dot{x}_i^L(t) &:= \bar{f}_i^L(t, U, B_i^L(X^B(t))), \quad x_i^L(t_0) = x_{0,i}^L, \\ \dot{x}_i^U(t) &:= \bar{f}_i^U(t, U, B_i^U(X^B(t))), \quad x_i^U(t_0) = x_{0,i}^U. \end{aligned} \tag{5.25}$$

If *a priori* enclosure G is available for reducing conservatism, it can be employed by including an interval refining operator I_G^B; see Section 5.6.

McCormick interval extensions

Besides NIE, another applicable interval extension technique for constructing $\bar{F}^B$ is MIE, which uses generalized McCormick-type relaxations. Unlike the traditional usage of GMC and DMC that takes convex and concave relaxations as inputs, MIE takes interval bounds as inputs: $\boldsymbol{x}^L$ and $\boldsymbol{x}^U$ in this case. GMC satisfies Assumption 5.4 according to [117, Section 2.7], and similar arguments hold for DMC.

5.5.2 Optimizing states

This subsection considers a second use case of our new framework. Functions $\boldsymbol{f}^L, \boldsymbol{f}^U$ are constructed with a strategy that involves optimization with respect to state variables [131].

Assumption 5.5. *Assume that interval function* $\tilde{F}^B \equiv [\tilde{\boldsymbol{f}}^L, \tilde{\boldsymbol{f}}^U] : I \times \mathbb{IR}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ *satisfies the following conditions:*

1. $\tilde{\boldsymbol{f}}^L$ *and* $\tilde{\boldsymbol{f}}^U$ *are continuous,*
2. $(\boldsymbol{z}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \tilde{\boldsymbol{f}}^L(t, U, \boldsymbol{z}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ *and* $(\boldsymbol{z}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \tilde{\boldsymbol{f}}^U(t, U, \boldsymbol{z}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ *are Lipschitz continuous, and uniformly in* t*,*
3. $\boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{\xi}) \in \tilde{F}^B(t, U, \boldsymbol{\xi}, \Xi)$ *for a.e.* $t \in I$*, any* $\boldsymbol{p} \in U$*, and any* $\Xi \in \mathbb{IR}^{n_x}$ *and* $\boldsymbol{\xi} \in \Xi$*.*

In this use case, consider $\boldsymbol{f}^L, \boldsymbol{f}^U$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} f_i^L(t, \boldsymbol{p}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U]) &:= \min_{\boldsymbol{z} \in \Pi_i^L(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])} \tilde{f}_i^L(t, U, \boldsymbol{z}, \Pi_i^L(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])), \\ f_i^U(t, \boldsymbol{p}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U]) &:= \max_{\boldsymbol{z} \in \Pi_i^U(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])} \tilde{f}_i^U(t, U, \boldsymbol{z}, \Pi_i^U(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])), \end{aligned} \tag{5.26}$$

where Π_i^L, Π_i^U are operators satisfying Assumption 5.3. Then, (5.18) becomes: for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} \dot{x}_i^L(t) &= \min_{\boldsymbol{z} \in \Pi_i^L(t, U, X^B(t))} \tilde{f}_i^L(t, U, \boldsymbol{z}, \Pi_i^L(t, U, X^B(t))), \qquad x_i^L(t_0) = x_{i,0}^L, \\ \dot{x}_i^U(t) &= \max_{\boldsymbol{z} \in \Pi_i^U(t, U, X^B(t))} \tilde{f}_i^U(t, U, \boldsymbol{z}, \Pi_i^U(t, U, X^B(t))), \qquad x_i^U(t_0) = x_{i,0}^U. \end{aligned} \tag{5.27}$$

Lemma 5.3. *Under Assumptions 5.3 and 5.5,* $\boldsymbol{f}^L, \boldsymbol{f}^U$ *in* (5.26) *satisfy Assumption 5.1.*

Proof. Condition 2 of Assumption 5.3 and Condition 1 of Assumption 5.5 guarantees the continuity requirement in Condition 1 of Assumption 5.1.

Define $\boldsymbol{f}^{L\dagger}, \boldsymbol{f}^{U\dagger} : I \times \mathbb{R}^{n_u} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$f_i^{L\dagger}(t, \boldsymbol{p}, \Xi) := \min_{\boldsymbol{z} \in B_i^L(\Xi)} \tilde{f}_i^L(t, U, \boldsymbol{z}, B_i^L(\Xi)),$$
$$f_i^{U\dagger}(t, \boldsymbol{p}, \Xi) := \max_{\boldsymbol{z} \in B_i^U(\Xi)} \tilde{f}_i^U(t, U, \boldsymbol{z}, B_i^U(\Xi)).$$

Under Assumption 5.5, $\boldsymbol{f}^{L,\dagger}, \boldsymbol{f}^{U\dagger}$ are Lipschitz continuous in $\boldsymbol{\xi}^L, \boldsymbol{\xi}^U$ according to [131, Proposition 2]. Since B_i^L, B_i^U are linear operations, there exists corresponding reverse operators $\check{B}_i^L, \check{B}_i^U$ that are locally Lipschitz continuous and satisfy

$$\check{B}_i^L(B_i^L(\Xi)) = \Xi \quad \text{and} \quad \check{B}_i^U(B_i^U(\Xi)) = \Xi.$$

Then,

$$f_i^{L\dagger}(t, \boldsymbol{p}, \check{B}_i^L(\Pi_i^L(t, U, \Xi))) = \min_{\boldsymbol{z} \in \Pi_i^L(t,U,\Xi)} \tilde{f}_i^L(t, U, \boldsymbol{z}, \Pi_i^L(t, U, \Xi)) \ = f_i^L(t, \boldsymbol{p}, \Xi),$$
$$f_i^{U\dagger}(t, \boldsymbol{p}, \check{B}_i^U(\Pi_i^U(t, U, \Xi))) = \max_{\boldsymbol{z} \in \Pi_i^U(t,U,\Xi)} \tilde{f}_i^U(t, U, \boldsymbol{z}, \Pi_i^U(t, U, \Xi)) = f_i^U(t, \boldsymbol{p}, \Xi).$$

Since the composite function of locally Lipschitz continuous functions is locally Lipschitz continuous [117, Theorem 2.5.6], $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto f_i^{L\dagger}(t, \boldsymbol{p}, \check{B}_i^L(\Pi_i^L(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])))$ and $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto f_i^{U\dagger}(t, \boldsymbol{p}, \check{B}_i^U(\Pi_i^U(t, U, \Xi)))$ are Lipschitz continuous. Thus, $\boldsymbol{f}^L(t, \boldsymbol{p}, \cdot, \cdot)$ and $\boldsymbol{f}^U(t, \boldsymbol{p}, \cdot, \cdot)$ satisfy the Lipschitz continuity in Condition 2 of Assumption 5.1.

Next, we verify Condition 3 of Assumption 5.1. Consider a.e. $t \in I$, any $i \in \{1, \ldots, n_x\}$, $\boldsymbol{u} \in \mathcal{U}$, $\boldsymbol{x}_0 \in X_0$, and $\Xi \equiv [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U] \in \mathbb{IR}^{n_x}$ such that $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in \Xi$.

According to Definition 5.5, it is desired to show that,

1. If $x_i(t) = \xi_i^L$, then $\min_{\boldsymbol{z}\in\Pi_i^L(t,U,\Xi)} \tilde{f}_i^L(t,U,\boldsymbol{z},\Pi_i^L(t,U,\Xi)) \leq \dot{x}_i(t,\boldsymbol{u},\boldsymbol{x}_0)$, and
2. If $x_i(t) = \xi_i^U$, then $\max_{\boldsymbol{z}\in\Pi_i^U(t,U,\Xi)} \tilde{f}_i^U(t,U,\boldsymbol{z},\Pi_i^U(t,U,\Xi)) \geq \dot{x}_i(t,\boldsymbol{u},\boldsymbol{x}_0)$.

It will be shown that the first requirement holds; it is analogous to prove the second. If $x_i(t,\boldsymbol{u},\boldsymbol{x}_0) = \xi_i^L$, Assumption 5.3 ensures that $\boldsymbol{x}(t,\boldsymbol{u},\boldsymbol{x}_0) \in \Pi_i^L(t,U,\Xi)$. Condition 3 of Assumption 5.5 guarantees that

$$\begin{aligned}\min_{\boldsymbol{z}\in\Pi_i^L(t,U,\Xi)} \tilde{f}_i^L(t,U,\boldsymbol{z},\Pi_i^L(t,U,\Xi)) &\leq \tilde{f}_i^L(t,U,\boldsymbol{x}(t,\boldsymbol{u},\boldsymbol{x}_0),\Pi_i^L(t,U,\Xi)) \\ &\leq f_i(t,\boldsymbol{u}(t),\boldsymbol{x}(t,\boldsymbol{u},\boldsymbol{x}_0)) \\ &= \dot{x}_i(t,\boldsymbol{u},\boldsymbol{x}_0),\end{aligned}$$

which ensures the first requirement.

Therefore, all three conditions in Assumption 5.1 are satisfied. □

Nonlinear relaxations

Inclusion function $\tilde{F}^B$ can be constructed using nonlinear convex relaxations GMC and DMC. They satisfy Assumption 5.5 according to the similar reason as described in Section 5.5.1. In this case, the RHS of (5.18) contains convex nonlinear programs (NLPs). Note that local NLP solvers, e.g., IPOPT and CONOPT, typically requires smoothness. While DMC is guaranteed to be continuously differentiable, nonsmoothness may exist in GMC. Although empirically these solvers might still be able to solve these nonsmooth NLPs, their performance will be effected. One

alternative approach for solving nonsmooth NLPs is to use the level method described in [97]. It reformulates the problem into combinations of linear programs (LPs) and quadratic programs (QPs) that can be solved efficiently with advanced LP solvers, e.g., CPLEX and Gurobi.

Piecewise-affine relaxations

Another type of relaxations for constructing $\tilde{F}^B$ is piecewise-affine relaxations. They can be generated using the subtangents of nonlinear convex relaxations at multiple points. Piecewise-affine relaxations are essentially approximations of the original nonlinear relaxations and are certainly looser than the original relaxations. However, they allow the embedded optimization problems in auxiliary RHS to be formulated as LPs, which typically solve much faster than NLPs. The following part discusses about constructing piecewise-affine convex relaxations by linearly approximating nonlinear convex relaxations.

Assumption 5.6. *Assume that interval function* $\tilde{F}^{B,nl} = [\tilde{\boldsymbol{f}}^{L,nl}, \tilde{\boldsymbol{f}}^{L,nl}] : I \times \mathbb{IR}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ *satisfies the following conditions:*

1. $\tilde{\boldsymbol{f}}^{L,nl}$ *and* $\tilde{\boldsymbol{f}}^{L,nl}$ *are continuous*
2. $\tilde{\boldsymbol{f}}^{L,nl}(t, U, \boldsymbol{z}, \Xi)$ *and* $\tilde{\boldsymbol{f}}^{L,nl}(t, U, \boldsymbol{z}, \Xi)$ *are differentiable with respect to* $\boldsymbol{z}$ *for all* $t \in I$ *and* $\Xi \in \mathbb{IR}^{n_x}$,
3. $(\boldsymbol{z}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \frac{\partial \tilde{\boldsymbol{f}}^{L,nl}}{\partial \boldsymbol{z}}(t, U, \boldsymbol{z}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ *and* $(\boldsymbol{z}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \frac{\partial \tilde{\boldsymbol{f}}^{L,nl}}{\partial \boldsymbol{z}}(t, U, \boldsymbol{z}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ *are locally Lipschitz continuous, uniformly over* t, *and*
4. $\tilde{F}^{B,nl}(t, U, \cdot, \Xi)$ *is a convex inclusion function of* $\boldsymbol{f}(t, \boldsymbol{p}, \cdot)$ *on* U *for any* $(t, \Xi) \in I \times \mathbb{IR}^{n_x}$ *and* $\boldsymbol{p} \in U$.

A nonlinear interval function $\tilde{F}^{B,nl}$ satisfying Assumption 5.6 can be constructed with $\mathscr{C}^2$-DMC.

Define function $\chi : \mathbb{IR}^{n_x} \times (0,1) \to \mathbb{R}^{n_x}$ such that

$$\chi(\Gamma, \lambda) := \gamma^L + \lambda(\gamma^U - \gamma^L).$$

Consider an arbitrary positive integer m and choose $\lambda^1, \lambda^2, \ldots, \lambda^m \in (0,1)$. For any $t \in I$, $i \in \{1, \ldots, n_x\}$, and $\Gamma \in \mathbb{IR}^{n_x}$, the subtangents of the nonlinear convex relaxation and the supertangents of the nonlinear concave relaxation at $\lambda^j, j \in \{1, \ldots, m\}$ are

$$(\boldsymbol{a}^{L,i,j}(t,\Gamma))^\top \boldsymbol{z} + b^{L,i,j}(t,\Gamma) = 0 \quad \text{and} \quad (\boldsymbol{a}^{U,i,j}(t,\Gamma))^\top \boldsymbol{z} + b^{U,i,j}(t,\Gamma) = 0,$$

respectively, where

$$\begin{aligned} \boldsymbol{a}^{L,i,j}(t,\Gamma) &= \frac{\partial \tilde{f}_i^{L,nl}}{\partial \boldsymbol{z}}(t, U, \chi(\Gamma, \lambda^j), \Gamma), \\ \boldsymbol{a}^{U,i,j}(t,\Gamma) &= \frac{\partial \tilde{f}_i^{U,nl}}{\partial \boldsymbol{z}}(t, U, \chi(\Gamma, \lambda^j), \Gamma), \end{aligned} \tag{5.28}$$

and

$$\begin{aligned} b^{L,i,j}(t,\Gamma) &= -(\boldsymbol{a}^{L,i,j}(t,\Gamma))^\top \chi(\Gamma, \lambda^j) + \tilde{f}_i^{L,nl}(t, U, \chi(\Gamma, \lambda^j), \Gamma), \\ b^{U,i,j}(t,\Gamma) &= -(\boldsymbol{a}^{U,i,j}(t,\Gamma))^\top \chi(\Gamma, \lambda^j) + \tilde{f}_i^{U,nl}(t, U, \chi(\Gamma, \lambda^j), \Gamma). \end{aligned} \tag{5.29}$$

The piecewise-affine convex (concave) relaxation is the maximum (minimum) of these subtangents (supertangents). So, we define $\tilde{\boldsymbol{f}}^L, \tilde{\boldsymbol{f}}^U$ such that, for each $i \in$

$\{1, \ldots, n_x\}$,

$$\begin{aligned} \tilde{f}_i^L(t, U, \boldsymbol{z}, \Gamma) &:= \max\{(\boldsymbol{a}^{L,i,j}(t,\Gamma))^\top \boldsymbol{z} + b^{L,i,j}(t,\Gamma) : j \in \{1,\ldots,m\}\}, \\ \tilde{f}_i^U(t, U, \boldsymbol{z}, \Gamma) &:= \min\{(\boldsymbol{a}^{U,i,j}(t,\Gamma))^\top \boldsymbol{z} + b^{U,i,j}(t,\Gamma) : j \in \{1,\ldots,m\}\}. \end{aligned} \tag{5.30}$$

Lemma 5.4. *Under Assumptions 5.3 and 5.6, $\tilde{F}^B$ in* (5.30) *satisfies Assumption 5.5.*

Proof. Condition 1 of Assumption 5.6 ensures Condition 1 of Assumption 5.5.

Next, we verify the Lipschtiz continuity required in Condition 2 of Assumption 5.5. According to Section 5.1 in [62], we need to show that $\boldsymbol{a}^{L,j}(t,\cdot,\cdot)$, $\boldsymbol{a}^{U,j}(t,\cdot,\cdot)$ and $b^{L,j}(t,\cdot,\cdot)$, $b^{U,j}(t,\cdot,\cdot)$ are locally Lipschitz continuous. Since $b^{L,j}, b^{U,j}$ are determined through (5.29) and $\boldsymbol{\chi}^j$ only depends on Γ, it suffices to show that $(\boldsymbol{\gamma}^L, \boldsymbol{\gamma}^U) \mapsto \boldsymbol{a}^{L,j}(t, [\boldsymbol{\gamma}^L, \boldsymbol{\gamma}^U])$ and $(\boldsymbol{\gamma}^L, \boldsymbol{\gamma}^U) \mapsto \boldsymbol{a}^{U,j}(t, [\boldsymbol{\gamma}^L, \boldsymbol{\gamma}^U])$ are locally Lipschitz continuous for each $t \in I$. According to (5.28), this always holds under Assumption 5.6.

Lastly, since Condition 4 of Assumption 5.6 holds and since $\tilde{F}^B$ in (5.30) is a piecewise affine approximation of $\tilde{F}^{B,nl}$, Condition 3 of Assumption 5.5 is satisfied.

Hence, all three conditions in Assumption 5.5 are satisfied. □

Affine relaxations

Affine relaxation is a special case of piecewise-affine relaxation, where the original nonlinear convex and concave relaxations are linearized at a single point, typically the midpoint. Therefore, the formulation defined in the previous piecewise-affine relaxation section is applicable here with $m = 1$ and $\lambda = 0.5$. The major advantage of affine relaxations is that we no longer need any local optimization solvers to evaluate the RHS of (5.26), neither an NLP solver or an LP solver. The optimization problem can be solved trivially with a simple algorithm described in Algorithm 1.

Algorithm 1: Min and max affine functions $h^\dagger, h^\ddagger : \mathbb{R}^{n_z} \to \mathbb{R}$, respectively and simultaneously, subject to $\boldsymbol{z} \in Z$

Result: The minimum of $h^\dagger$ on Z, $h^{\dagger *}$; the maximum of $h^\ddagger$ on Z, $h^{\ddagger *}$.

Choose $\bar{\boldsymbol{z}} \in Z$, usually the midpoint;
Evaluate $h^\dagger(\bar{\boldsymbol{z}}), h^\ddagger(\bar{\boldsymbol{z}})$, and the respective slopes $s^\dagger, s^\ddagger$;
$h^{\dagger *} = h^\dagger(\bar{\boldsymbol{z}})$;
$h^{\ddagger *} = h^\ddagger(\bar{\boldsymbol{z}})$;
for $i = 1, \ldots, n_z$ **do**
  if $s^\dagger(\bar{\boldsymbol{z}}) \geq 0$ **then**
    $h^{\dagger *} = h^{\dagger *} + s^\dagger(z_i^L - \bar{z}_i)$;
  else
    $h^{\dagger *} = h^{\dagger *} + s^\dagger(z_i^U - \bar{z}_i)$;
  end
  if $s^\ddagger(\bar{\boldsymbol{z}}) \geq 0$ **then**
    $h^{\ddagger *} = h^{\ddagger *} + s^\ddagger(z_i^U - \bar{z}_i)$;
  else
    $h^{\ddagger *} = h^{\ddagger *} + s^\ddagger(z_i^L - \bar{z}_i)$;
  end
end

5.5.3 Optimizing parameters

This subsection introduces a straightforward use case of the new formulation (5.18). It optimizes the auxiliary RHS functions with respect to the parameter.

Assumption 5.7. *Assume that interval function* $\check{F}^B \equiv [\check{\boldsymbol{f}}^L, \check{\boldsymbol{f}}^U] : I \times \mathbb{R}^{n_u} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ *satisfies the following conditions:*

1. $\check{\boldsymbol{f}}^L$ *and* $\check{\boldsymbol{f}}^U$ *are continuous,*
2. $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \check{\boldsymbol{f}}^L(t, \boldsymbol{p}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ *and* $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \check{\boldsymbol{f}}^U(t, \boldsymbol{p}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ *are Lipschitz continuous on* $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$*, uniformly in* $(t, \boldsymbol{p})$*, and*
3. $\Xi \mapsto \check{F}^B(t, \boldsymbol{p}, \Xi)$ *is an inclusion function of* $\boldsymbol{f}(t, \boldsymbol{p}, \cdot)$ *on* D *for a.e.* $t \in I$ *and all* $\boldsymbol{p} \in U$*.*

In this use case, consider $\boldsymbol{f}^L, \boldsymbol{f}^U$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} f_i^L(t, \boldsymbol{p}, \Xi) &:= \check{f}_i^L(t, \boldsymbol{p}, \Pi_i^L(t, U, \Xi)), \\ f_i^U(t, \boldsymbol{p}, \Xi) &:= \check{f}_i^U(t, \boldsymbol{p}, \Pi_i^U(t, U, \Xi)), \end{aligned} \tag{5.31}$$

where Π_i^L, Π_i^U are operators satisfying Assumption 5.3. Then, the auxiliary system (5.18) becomes: for every $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} \dot{x}_i^L(t) &= \min_{\boldsymbol{p} \in U} \check{f}_i^L(t, \boldsymbol{p}, \Pi_i^L(t, U, X^B(t))), \quad x_i^L(t_0) = x_{0,i}^L, \\ \dot{x}_i^U(t) &= \max_{\boldsymbol{p} \in U} \check{f}_i^U(t, \boldsymbol{p}, \Pi_i^U(t, U, X^B(t))), \quad x_i^U(t_0) = x_{0,i}^U. \end{aligned} \tag{5.32}$$

Lemma 5.5. *Under Assumptions 5.3 and 5.7,* $\boldsymbol{f}^L, \boldsymbol{f}^U$ *in* (5.31) *satisfy Assumption 5.1.*

Proof. Condition 2 of Assumption 5.3 and Condition 1 of Assumption 5.7 guarantees the continuity requirement in Condition 1 of Assumption 5.1.

Since the composite function of locally Lipschitz continuous functions is locally Lipschitz continuous, Condition 2 of Assumption 5.3 and Condition 2 of Assumption 5.7 ensures Condition 2 of Assumption 5.1.

Next, we verify the enclosing dynamics required in Condition 3 of Assumption 5.1. Consider a.e. $t \in I$, any $i \in \{1, \ldots, n_x\}$, $\boldsymbol{u} \in \mathcal{U}$, $\boldsymbol{x}_0 \in X_0$, and $\Xi \equiv [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U] \in \mathbb{IR}^{n_x}$ such that $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in \Xi$. According to Definition 5.5, it is desired to show that

1. If $x_i(t, \boldsymbol{u}, \boldsymbol{x}_0) = \xi_i^L$, then $\check{f}_i^L(t, \boldsymbol{u}(t), \Pi_i^L(t, U, \Xi)) \leq \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$, and
2. If $x_i(t, \boldsymbol{u}, \boldsymbol{x}_0) = \xi_i^U$, then $\check{f}_i^U(t, \boldsymbol{u}(t), \Pi_i^U(t, U, \Xi)) \geq \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$.

It will be shown that the first requirement holds; it is analogous to prove the second. If $x_i(t, \boldsymbol{u}, \boldsymbol{x}_0) = \xi_i^L$, Assumption 5.3 ensures that $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in \Pi_i^L(t, U, \Xi)$. The third condition in Assumption 5.7 guarantees that

$$\check{f}_i^L(t, \boldsymbol{u}(t), \Pi_i^L(t, U, \Xi)) \leq f_i(t, \boldsymbol{u}(t), \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0)) = \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0),$$

which verifies the first requirement.

Therefore, all three conditions in Assumption 5.1 are satisfied. □

Similar to Section 5.5.2, we present three categories of methods to construct $\check{\boldsymbol{f}}^L$ and $\check{\boldsymbol{f}}^U$ and they are distinguished by their linearity.

Nonlinear relaxation

$\check{F}^B$ can be a nonlinear inclusion function of $\boldsymbol{f}(t,\cdot,\cdot,\cdot)$ constructed using GMC and DMC. Arguments analogous to those in previous sections ensure that these two methods satisfy Assumption 5.7.

Piecewise-affine relaxations

$\check{F}^B$ can also be constructed with piecewise-affine approximations of those nonlinear convex relaxations mentioned above. This method is similar to the piecewise-affine relaxations presented in Section 5.5.2.

Assumption 5.8. *Assume that nonlinear interval function* $\check{F}^{B,nl} = [\check{\boldsymbol{f}}^{L,nl}, \check{\boldsymbol{f}}^{L,nl}] : I \times \mathbb{R}^{n_u} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ *satisfies the following conditions:*

1. $\check{\boldsymbol{f}}^{L,nl}$ *and* $\check{\boldsymbol{f}}^{L,nl}$ *are continuous,*
2. $\check{\boldsymbol{f}}^{L,nl}(t,\cdot,\Xi)$ *and* $\check{\boldsymbol{f}}^{L,nl}(t,\cdot,\Xi)$ *are differentiable for all* $t \in I$ *and* $\Xi \in \mathbb{IR}^{n_x}$,
3. $(\boldsymbol{\xi}^L,\boldsymbol{\xi}^U) \mapsto \frac{\partial \check{\boldsymbol{f}}^{L,nl}}{\partial \boldsymbol{p}}(t,\boldsymbol{p},[\boldsymbol{\xi}^L,\boldsymbol{\xi}^U])$ *and* $(\boldsymbol{\xi}^L,\boldsymbol{\xi}^U) \mapsto \frac{\partial \check{\boldsymbol{f}}^{U,nl}}{\partial \boldsymbol{p}}(t,\boldsymbol{p},[\boldsymbol{\xi}^L,\boldsymbol{\xi}^U])$ *are Lipschitz continuous on* $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, *uniformly in* $(t,\boldsymbol{p})$,
4. $\Xi \mapsto \check{F}^B(t,\boldsymbol{p},\Xi)$ *is a convex inclusion function of* $\boldsymbol{f}(t,\boldsymbol{p},\cdot)$ *on* D *for a.e.* $t \in I$ *and all* $\boldsymbol{p} \in U$.

This nonlinear convex inclusion function satisfying Assumption 5.8 may be generated from $\boldsymbol{f}$ with $\mathscr{C}^2$-DMC.

Choose arbitrary fixed points $\boldsymbol{\rho}^j \in U, j \in \{1,\ldots,m\}$. For each $i \in \{1,\ldots,n_x\}$ and $j \in \{1,\ldots,m\}$, defined $\boldsymbol{a}^{L,i,j}, \boldsymbol{a}^{U,i,j} : I \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ and $b^{L,i,j}, b^{U,i,j} : I \times$

$\mathbb{R}^{n_x} \to \mathbb{R}$ such that

$$\boldsymbol{a}^{L,i,j}(t,\Xi) = \frac{\partial \check{f}_i^{L,nl}}{\partial \boldsymbol{p}}(t,\boldsymbol{\rho}^j,\Xi), \qquad b^{L,i,j}(t,\Xi) = -(\boldsymbol{a}^{L,i,j}(t,\Xi))^\top \boldsymbol{\rho}^j + \check{f}_i^{L,nl}(t,\boldsymbol{\rho}^j,\Xi),$$
$$\boldsymbol{a}^{U,i,j}(t,\Xi) = \frac{\partial \check{f}_i^{U,nl}}{\partial \boldsymbol{p}}(t,\boldsymbol{\rho}^j,\Xi), \qquad b^{U,i,j}(t,\Xi) = -(\boldsymbol{a}^{U,i,j}(t,\Xi))^\top \boldsymbol{\rho}^j + \check{f}_i^{U,nl}(t,\boldsymbol{\rho}^j,\Xi).$$

Then, the piecewise-affine relaxation is the maximum (minimum) of these subtangents (super-tangents): For each $i \in \{1,\ldots,n_x\}$ and $t \in I$,

$$\begin{aligned}
\check{f}_i^L(t,\boldsymbol{p},\Xi) &:= \max\{(\boldsymbol{a}^{L,i,j}(t,\Xi))^\top \boldsymbol{p} + b^{L,i,j}(t,\Xi) : j \in \{1,\ldots,m\}\},\\
\check{f}_i^U(t,\boldsymbol{p},\Xi) &:= \min\{(\boldsymbol{a}^{U,i,j}(t,\Xi))^\top \boldsymbol{p} + b^{U,i,j}(t,\Xi) : j \in \{1,\ldots,m\}\}.
\end{aligned} \tag{5.33}$$

It was readily verified that $\check{F}^B$ defined in (5.33) satisfies Assumption 5.7 according to similar arguments as in Section 5.5.2.

Affine relaxations

Similar to the discussion in Section 5.5.2, affine relaxation, a special case of piecewise-affine relaxation, is also applicable here. The optimization problems on the RHS of (5.18) can be solved trivially using Algorithm 1.

5.5.4 Optimizing states and parameters

Similar to the second use case introduced in Section 5.5.2, the last use case of our new framework also construct $\boldsymbol{f}^L$, $\boldsymbol{f}^U$ using optimization problems with respect to states [131]. But here, we combine them with the embedded optimization problems in the RHS of (5.18) and obtain optimization problems with respect to both states

and parameters.

Assumption 5.9. *Assume that interval function* $\hat{F}^B \equiv [\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U] : I \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ *satisfies the following conditions:*

1. $\hat{\boldsymbol{f}}^L$ *and* $\hat{\boldsymbol{f}}^U$ *are continuous,*
2. $(\boldsymbol{\xi}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \hat{\boldsymbol{f}}^L(t, \boldsymbol{p}, \boldsymbol{\xi}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ *and* $(\boldsymbol{\xi}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \hat{\boldsymbol{f}}^U(t, \boldsymbol{p}, \boldsymbol{\xi}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ *are Lipschitz continuous, uniformly in* $(t, \boldsymbol{p})$*,*
3. $\boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{\xi}) \in \hat{F}^B(t, \boldsymbol{p}, \boldsymbol{\xi}, \Xi)$ *for a.e.* $t \in I$*, any* $\boldsymbol{p} \in U$*,* $\Xi \in \mathbb{IR}^{n_x}$ *and* $\boldsymbol{\xi} \in \Xi$*.*

In this use case, consider $\boldsymbol{f}^L, \boldsymbol{f}^U$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} f_i^L(t, \boldsymbol{p}, \Xi) &:= \min_{\boldsymbol{z} \in \Pi_i^L(t, U, \Xi)} \hat{f}_i^L(t, \boldsymbol{p}, \boldsymbol{z}, \Pi_i^L(t, U, \Xi)), \\ f_i^U(t, \boldsymbol{p}, \Xi) &:= \max_{\boldsymbol{z} \in \Pi_i^U(t, U, \Xi)} \hat{f}_i^U(t, \boldsymbol{p}, \boldsymbol{z}, \Pi_i^U(t, U, \Xi)), \end{aligned} \tag{5.34}$$

where Π_i^L, Π_i^U are operators satisfying Assumption 5.3. Substitute (5.34) into (5.18), (5.18) becomes: for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} \dot{x}_i^L(t) &= \min_{\boldsymbol{p} \in U,\ \boldsymbol{z} \in \Pi_i^L(t, U, X^B(t))} \hat{f}_i^L(t, \boldsymbol{p}, \boldsymbol{z}, \Pi_i^L(t, U, X^B(t))), \quad x_i^L(t_0) = x_{i,0}^L, \\ \dot{x}_i^U(t) &= \max_{\boldsymbol{p} \in U,\ \boldsymbol{z} \in \Pi_i^U(t, U, X^B(t))} \hat{f}_i^U(t, \boldsymbol{p}, \boldsymbol{z}, \Pi_i^U(t, U, X^B(t))), \quad x_i^U(t_0) = x_{i,0}^U. \end{aligned} \tag{5.35}$$

Lemma 5.6. *Under Assumptions 5.3 and 5.9,* $\boldsymbol{f}^L, \boldsymbol{f}^U$ *in* (5.34) *satisfy Assumption 5.1.*

Proof. Condition 2 of Assumption 5.3 and Condition 1 of Assumption 5.9 guarantees the continuity requirement in Condition 1 of Assumption 5.1.

Define $\boldsymbol{f}^{L\dagger}, \boldsymbol{f}^{U\dagger} : I \times \mathbb{R}^{n_u} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$f_i^{L\dagger}(t, \boldsymbol{p}, \Xi) := \min_{\boldsymbol{z} \in B_i^L(\Xi)} \hat{f}_i^L(t, \boldsymbol{p}, \boldsymbol{z}, B_i^L(\Xi)),$$
$$f_i^{U\dagger}(t, \boldsymbol{p}, \Xi) := \max_{\boldsymbol{z} \in B_i^U(\Xi)} \hat{f}_i^U(t, \boldsymbol{p}, \boldsymbol{z}, B_i^U(\Xi)).$$

Under Assumption 5.9, the discussion in [131, Sections 5.1 and 5.2] ensures that $\boldsymbol{f}^{L\dagger}, \boldsymbol{f}^{U\dagger}$ are locally Lipschitz continuous in $\boldsymbol{\xi}^L, \boldsymbol{\xi}^U$ for each $(t, \boldsymbol{p}) \in I \times U$. Since B_i^L, B_i^U are linear operations, there exists corresponding reverse operations $\check{B}_i^L, \check{B}_i^U$ of B_i^L, B_i^U that are Lipschitz continuous and satisfy

$$\check{B}_i^L(B_i^L(\Xi)) = \Xi \quad \text{and} \quad \check{B}_i^U(B_i^U(\Xi)) = \Xi.$$

Then,

$$f_i^{L\dagger}(t, \boldsymbol{p}, \check{B}_i^L(\Pi_i^L(t, U, \Xi))) = \min_{\boldsymbol{z} \in \Pi_i^L(t,U,\Xi)} \hat{f}_i^L(t, \boldsymbol{p}, \boldsymbol{z}, \Pi_i^L(t, U, \Xi)) \; = f_i^L(t, \boldsymbol{p}, \Xi),$$
$$f_i^{U\dagger}(t, \boldsymbol{p}, \check{B}_i^U(\Pi_i^U(t, U, \Xi))) = \max_{\boldsymbol{z} \in \Pi_i^U(t,U,\Xi)} \hat{f}_i^U(t, \boldsymbol{p}, \boldsymbol{z}, \Pi_i^U(t, U, \Xi)) = f_i^U(t, \boldsymbol{p}, \Xi).$$

Since the composite function of locally Lipschitz continuous functions is locally Lipschitz continuous [117, Theorem 2.5.6], Condition 2 of Assumption 5.3 and Condition 1 of Assumption 5.9 imply that $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto f_i^{L\dagger}(t, \boldsymbol{p}, \check{B}_i^L(\Pi_i^L(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])))$ and $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto f_i^{U\dagger}(t, \boldsymbol{p}, \check{B}_i^U(\Pi_i^U(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])))$ are locally Lipschitz continuous. Hence, $\boldsymbol{f}^L(t, \boldsymbol{p}, \cdot, \cdot)$ and $\boldsymbol{f}^U(t, \boldsymbol{p}, \cdot, \cdot)$ are Lipschitz continuous for each $(t, \boldsymbol{p}) \in I \times U$ and Condition 2 of Assumption 5.1 is satisfied..

Next, we verify the enclosing dynamics in Condition 3 of Assumption 5.1. Consider a.e. $t \in I$, any $i \in \{1, \ldots, n_x\}$, $\boldsymbol{u} \in \mathcal{U}$, $\boldsymbol{x}_0 \in X_0$, and $\Xi \equiv [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U] \in \mathbb{IR}^{n_x}$ such that $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in \Xi$. According to Definition 5.5, it is desired to show that,

1. If $x_i(t, \boldsymbol{u}, \boldsymbol{x}_0) = \xi_i^L$, then $\min_{\boldsymbol{z} \in \Pi_i^L(t,U,\Xi)} \hat{f}_i^L(t, \boldsymbol{p}, \boldsymbol{z}, \Pi_i^L(t, U, \Xi)) \leq \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$, and

2. If $x_i(t, \boldsymbol{u}, \boldsymbol{x}_0) = \xi_i^U$, then $\max_{\boldsymbol{z} \in \Pi_i^U(t,U,\Xi)} \hat{f}_i^U(t, \boldsymbol{p}, \boldsymbol{z}, \Pi_i^U(t, U, \Xi)) \geq \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0)$.

It will be shown that the first requirement holds; it is analogous to prove the second. If $x_i(t, \boldsymbol{u}, \boldsymbol{x}_0) = \xi_i^L$, Assumption 5.3 ensures that $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in \Pi_i^L(t, U, \Xi)$. Condition 3 in Assumption 5.9 shows that

$$\begin{aligned}\min_{\boldsymbol{z} \in \Pi_i^L(t,U,\Xi)} \hat{f}_i^L(t, \boldsymbol{u}(t), \boldsymbol{z}, \Pi_i^L(t, U, \Xi)) &\leq \hat{f}_i^L(t, \boldsymbol{u}(t), \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0), \Pi_i^L(t, U, \Xi)) \\ &\leq f_i(t, \boldsymbol{u}(t), \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0)) \\ &= \dot{x}_i(t, \boldsymbol{u}, \boldsymbol{x}_0),\end{aligned}$$

which ensures the first requirement.

Therefore, all three conditions in Assumption 5.1 are satisfied. □

In the remainder of this subsection, several approaches are introduced for generating $\hat{F}^B$ that satisfy Assumption 5.9.

Original RHS function

Suppose that $\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U := \boldsymbol{f}$. Then, (5.35) becomes

$$\begin{aligned} \dot{x}_i^L(t) &= \min_{\boldsymbol{p}\in U,\ \boldsymbol{z}\in\Pi_i^L(t,U,X^B(t))} f_i(t,\boldsymbol{p},\boldsymbol{z}), \qquad x_i^L(t_0) = x_{i,0}^L, \\ \dot{x}_i^U(t) &= \max_{\boldsymbol{p}\in U,\ \boldsymbol{z}\in\Pi_i^U(t,U,X^B(t))} f_i(t,\boldsymbol{p},\boldsymbol{z}), \qquad x_i^U(t_0) = x_{i,0}^U. \end{aligned} \tag{5.36}$$

If Π_i^L, Π_i^U is defined as B_i^L, B_i^U without considering an *a priori* enclosure, the generated formulation is similar to the following result proposed in [148, p. 96] and described in [59]:

$$\begin{aligned} \dot{x}_i^L(t) &\leq \min_{\boldsymbol{p}\in U,\ \boldsymbol{z}\in[\boldsymbol{x}^L,\boldsymbol{x}^U],\ z_i=x_i^L} f_i(t,\boldsymbol{p},\boldsymbol{z}), \qquad x_i^L(t_0) \leq x_{i,0}^L, \\ \dot{x}_i^U(t) &\geq \max_{\boldsymbol{p}\in U,\ \boldsymbol{z}\in[\boldsymbol{x}^L,\boldsymbol{x}^U],\ z_i=x_i^U} f_i(t,\boldsymbol{p},\boldsymbol{z}), \qquad x_i^U(t_0) \geq x_{i,0}^U. \end{aligned} \tag{5.37}$$

The solutions of (5.36) are the tightest possible bounds that can be constructed with this new framework. However, since $\boldsymbol{f}$ may be nonlinear and nonconvex, solving the embedded optimization problems in (5.36) during numerical integration is not practical [128].

The other methods generate $\hat{F}^B$ using various relaxations of the original ODE RHS function $\boldsymbol{f}$. To simplify notation in the following part, denote $\boldsymbol{y} = (\boldsymbol{z},\boldsymbol{p}) \in \mathbb{R}^{n_x+n_u}$ and define $Y_i^L, Y_i^U : I \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x} \times U$ such that

$$\begin{aligned} Y_i^L(t,\Xi) &= \Pi_i^L(t,U,\Xi) \times U, \\ Y_i^U(t,\Xi) &= \Pi_i^U(t,U,\Xi) \times U. \end{aligned}$$

Then, (5.35) becomes

$$\begin{aligned} \dot{x}_i^L(t) &= \min_{\boldsymbol{y}\in Y_i^L(t,X^B(t))} \hat{f}_i^L(t,\boldsymbol{y},\Pi_i^L(t,U,X^B(t))), \qquad x_i^L(t_0) = x_{i,0}^L, \\ \dot{x}_i^U(t) &= \max_{\boldsymbol{y}\in Y_i^U(t,X^B(t))} \hat{f}_i^U(t,\boldsymbol{y},\Pi_i^U(t,U,X^B(t))), \qquad x_i^U(t_0) = x_{i,0}^U. \end{aligned} \tag{5.38}$$

Nonlinear relaxation

1. Convex envelopes.

   Suppose that $\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U$ are the convex envelopes of $\boldsymbol{f}$. It was readily verified that Assumption 5.9 is satisfied. Since the minimum of a convex envelope is equivalent to the global minimum of the original function, this method also provides the tightest possible state bounds, equivalent to (5.36). The advantage of using a convex envelope is that, the embedded optimization problem can be solved to its global minimum with a local solver. But obtaining the convex envelope may be cumbersome.

2. McCormick-type relaxations.

   Suppose that $\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U$ are McCormick-type relaxations of $\boldsymbol{f}$. It was readily verified that Assumption 5.9 is satisfied.

3. αBB relaxations.

   Suppose that $\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U$ are αBB relaxations of $\boldsymbol{f}$. This use case has been discussed in [31]. Note that αBB relaxations are convex relaxations that can be optimized with state-of-the-art local NLP solvers, such as IPOPT and CONOPT. Moreover, if the original RHS function is quadratic, then the generated αBB relaxations are also quadratic. In this scenario, these quadratic relaxations

can be optimized with advanced LP solvers, e.g., CPLEX, which typically require less computing time.

4. Edge-concave relaxations.

   Suppose that $\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U$ are edge-concave relaxations developed in [63]. Even though almost every previous method for constructing $\boldsymbol{f}^L, \boldsymbol{f}^U$ involves a convex relaxation technique, our new framework also supports the usage of concave relaxation technique. To the authors' knowledge, this is the first time that a concave relaxation technique is used in enclosing reachable set using differential inequalities. For each $i \in \{1, \ldots, n_x\}$, the corresponding edge-concave under-estimator $\hat{f}_i^L$ and edge-convex over-estimator $\hat{f}_i^U$ of f_i can be constructed as follows.

$$\begin{aligned} \hat{f}_i^L(t, \boldsymbol{y}, \Pi_i^L(t, U, \Xi)) &:= f_i(t, \boldsymbol{y}) - \sum_{i=1}^{n_x+n_u} \theta_i^L \left(y_i - \mathrm{mid}(Y_i^L(t, \Xi)) \right)^2, \\ \hat{f}_i^U(t, \boldsymbol{y}, \Pi_i^L(t, U, \Xi)) &:= f_i(t, \boldsymbol{y}) + \sum_{i=1}^{n_x+n_u} \theta_i^U \left(y_i - \mathrm{mid}(Y_i^U(t, \Xi)) \right)^2, \end{aligned} \tag{5.39}$$

   where $\mathrm{mid} : \mathbb{IR}^{n_x+n_u} \to \mathbb{R}^{n_x+n_u}$ represents the middle point of an interval, and

$$\theta_i^L = \max\left\{0, \frac{1}{2}\left[\frac{\partial^2 f_i}{\partial y_i^2}\right]^U\right\}, \quad \theta_i^U = \max\left\{0, \frac{1}{2}\left[-\frac{\partial^2 f_i}{\partial y_i^2}\right]^U\right\}.$$

   It was readily verified that $\hat{F}^B$ in (5.39) satisfies Assumption 5.9. The optimum of the edge-concave relaxations are obtained by checking all the vertices of the interval domain. The computational cost of this process may be cheap if the numbers of state variables and uncertain parameters are small.

Piecewise-affine inclusion

Suppose that $\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U$ are piecewise-affine relaxations of $\boldsymbol{f}$. Then, the LP-based method from [62] is recovered. A special approach for constructing piecewise-affine relaxations for $\boldsymbol{f}$ with local Lipschitz continuity can be found in [62, Section 5.2].

Another general approach for generating piecewise-affine $\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U$ is to linearly approximating nonlinear convex relaxations, similar to the methods introduced in Sections 5.5.2 and 5.5.3.

Assumption 5.10. *Assume that interval function* $\hat{F}^{B,nl} = [\hat{\boldsymbol{f}}^{L,nl}, \hat{\boldsymbol{f}}^{L,nl}] : I \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ *satisfies the following conditions:*

1. $\hat{\boldsymbol{f}}^{L,nl}$ *and* $\hat{\boldsymbol{f}}^{L,nl}$ *are continuously differentiable,*
2. $\frac{\partial \hat{\boldsymbol{f}}^{L,nl}}{\partial \boldsymbol{y}}(t,\boldsymbol{p},\cdot), \frac{\partial \hat{\boldsymbol{f}}^{L,nl}}{\partial \boldsymbol{y}}(t,\boldsymbol{p},\cdot)$ *are locally Lipschitz continuous for each* $(t,\boldsymbol{p}) \in I \times U$, *and*
3. $(\boldsymbol{p},\boldsymbol{\xi}) \mapsto \hat{F}^{B,nl}(t,\boldsymbol{p},\boldsymbol{\xi})$ *is a convex inclusion function of* $(\boldsymbol{p},\boldsymbol{\xi}) \mapsto \boldsymbol{f}(t,\boldsymbol{p},\boldsymbol{\xi})$ *on* $U \times \Xi$ *for a.e.* $t \in I$ *and any* $\Xi \subset \mathbb{IR}^{n_x}$.

A nonlinear interval function $\hat{F}^{B,nl}$ satisfying Assumption 5.10 may be constructed with twice-continuously differentiable convex relaxations, e.g., $\mathscr{C}^2$-DMC and αBB relaxations.

Define function $\chi^\ddagger : \mathbb{IR}^{n_x+n_u} \times (0,1) \to \mathbb{R}^{n_x+n_u}$ such that

$$\chi^\ddagger(\Gamma,\lambda) = \gamma^L + \lambda(\gamma^U - \gamma^L).$$

Choose $\lambda^1, \lambda^2, \ldots, \lambda^m \in (0,1)$. For any $t \in I$, $i \in \{1, \ldots, n_x\}$, and $\Gamma \in \mathbb{IR}^{n_x}$, the subtangents of nonlinear convex relaxations and the supertangents of nonlinear concave relaxations at $\lambda^j, j \in \{1, \ldots, m\}$, are

$$(\boldsymbol{a}^{L,i,j}(t,\Gamma))^\top \boldsymbol{y} + b^{L,i,j}(t,\Gamma) = 0 \quad \text{and} \quad (\boldsymbol{a}^{U,i,j}(t,\Gamma))^\top \boldsymbol{y} + b^{U,i,j}(t,\Gamma) = 0,$$

respectively, where

$$\boldsymbol{a}^{L,i,j}(t,\Gamma) = \frac{\partial \hat{f}_i^{L,nl}}{\partial \boldsymbol{y}}(t, \chi^\ddagger(\Gamma, \lambda^j)),$$
$$\boldsymbol{a}^{U,i,j}(t,\Gamma) = \frac{\partial \hat{f}_i^{U,nl}}{\partial \boldsymbol{y}}(t, \chi^\ddagger(\Gamma, \lambda^j)),$$

and

$$b^{L,i,j}(t,\Gamma) = -(\boldsymbol{a}^{L,i,j}(t,\Gamma))^\top \chi^\ddagger(\Gamma, \lambda^j) + \hat{f}_i^{L,nl}(t, \chi^\ddagger(\Gamma, \lambda^j)),$$
$$b^{U,i,j}(t,\Gamma) = -(\boldsymbol{a}^{U,i,j}(t,\Gamma))^\top \chi^\ddagger(\Gamma, \lambda^j) + \hat{f}_i^{U,nl}(t, \chi^\ddagger(\Gamma, \lambda^j)).$$

The piecewise-affine convex (concave) relaxation is the maximum (minimum) of these subtangents (supertangents). So, define $\tilde{\boldsymbol{f}}^L, \tilde{\boldsymbol{f}}^{U,pa}$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$\hat{f}_i^L(t, \boldsymbol{y}, \Gamma) := \max\{(\boldsymbol{a}^{L,i,j}(t,\Gamma))^\top \boldsymbol{y} + b^{L,i,j}(t,\Gamma) : j \in \{1, \ldots, m\}\},$$
$$\hat{f}_i^U(t, \boldsymbol{y}, \Gamma) := \min\{(\boldsymbol{a}^{U,i,j}(t,\Gamma))^\top \boldsymbol{y} + b^{U,i,j}(t,\Gamma) : j \in \{1, \ldots, m\}\}.$$

Affine inclusion

Suppose that $\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U$ are affine relaxations of $\boldsymbol{f}$ constructed with the subtangent and supertangent of McCormick-based relaxations. Then, the method developed in [128] is recovered. Since McCormick-based relaxations is typically nonsmooth, $\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U$ are may not be continuous. Thus, the existence and uniqueness of system (5.18) is not guaranteed.

Alternatively, the affine relaxations can be constructed as a special case of the piecewise-affine relaxations introduced in the previous section. The original nonlinear convex and concave relaxations are linearized at a single point, typically the midpoint with parameters $m = 1$ and $\lambda = 0.5$.

5.5.5 Summary of use cases

Four different use cases of the new framework (5.18) have been introduced in Sections 5.5.1-5.5.4. Each of them includes multiple methods for constructing relaxations of the original RHS function. The continuity and bounding properties of $\boldsymbol{f}^L, \boldsymbol{f}^U$ were addressed to ensure Assumption 5.1 holds for every method. A summary of these use cases is presented in Table 5.2.

5.5.6 Comparison of use cases

This section compares the tightness of the state bounds generated with different use cases or different methods. The following lemma shows that tighter relaxations of $\boldsymbol{f}$ generates tighter state bounds in each use case.

Table 5.2: Summary of available methods in Section 5.5

Section	Domain	Relaxation method	Code
5.5.4	$I \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{IR}^{n_x}$	Convex envelope	px-N-E
		Nonlinear GMC	px-N-G
		Nonlinear DMC	px-N-D
		PA with $\mathscr{C}^2$-DMC	px-P-D
		Affine with $\mathscr{C}^2$-DMC	px-A-D
		αBB	px-N-α
		PA with αBB	px-P-α
		Affine with αBB	px-A-α
		Edge-concave	px-N-EC
		PA from [62]	px-P-H
5.5.3	$I \times \mathbb{R}^{n_u} \times \mathbb{IR}^{n_x}$	Nonlinear GMC	p-N-G
		Nonlinear DMC	p-N-D
		PA with $\mathscr{C}^2$-DMC	p-P-D
		Affine with $\mathscr{C}^2$-DMC	p-A-D
5.5.2	$I \times \mathbb{IR}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{IR}^{n_x}$	Nonlinear GMC	x-N-G
		Nonlinear DMC	x-N-D
		PA with $\mathscr{C}^2$-DMC	x-P-D
		Affine with $\mathscr{C}^2$-DMC	x-A-D
5.5.1	$I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x}$	Nonlinear GMC	○-N-G
		Nonlinear DMC	○-N-D
		NIE [59]	○-N-I

Lemma 5.7. *Assume that Assumptions 5.4, 5.5, 5.7, and 5.9 hold. In each use case, consider a pair of corresponding relaxations of* $\boldsymbol{f}$, $\check{F}^\dagger \in \{\bar{F}^\dagger, \tilde{F}^\dagger, \check{F}^\dagger, \hat{F}^\dagger\}$ *and* $\check{F}^\ddagger \in \{\bar{F}^\ddagger, \tilde{F}^\ddagger, \check{F}^\ddagger, \hat{F}^\ddagger\}$ *(with the same accent mark). Suppose that* Π_i^L, Π_i^U *are the flattening operators in Definition 5.4, and* $\check{F}^\ddagger$ *is inclusion monotonic on* $\mathbb{IR}^{n_x}$*. If* $\check{F}^\dagger$ *is tighter than* $\check{F}^\ddagger$ *in the sense of Definition 5.2, then the resulting state bounds* $X^{B,\dagger}$ *is tighter than* $X^{B,\ddagger}$ *with same initial conditions.*

Proof. This result can be proved with Theorem 5.4 by showing $\check{F}^\dagger$ and $\check{F}^\ddagger$ satisfy the conditions in Assumption 5.2.

Consider any $\boldsymbol{p} \in U$, $t \in I$, $\Xi^\dagger \equiv [\boldsymbol{\xi}^{L\dagger}, \boldsymbol{\xi}^{U\dagger}] \subseteq \Xi^\ddagger \equiv [\boldsymbol{\xi}^{L\ddagger}, \boldsymbol{\xi}^{U\ddagger}]$, and $i \in \{1, \ldots, n_x\}$. It will be shown that, if $\xi_i^{L,\dagger} = \xi_i^{L,\ddagger}$, then $\check{f}_i^{L,\ddagger}(t, \boldsymbol{p}, \Xi^\ddagger) \leq \check{f}_i^{L,\dagger}(t, \boldsymbol{p}, \Xi^\dagger)$; it is analogous to show that, if $\xi_i^{U,\dagger} = \xi_i^{U,\ddagger}$, then $\check{f}_i^{U,\ddagger}(t, \boldsymbol{p}, \Xi^\ddagger) \geq \check{f}_i^{U,\dagger}(t, \boldsymbol{p}, \Xi^\dagger)$.

If $\xi_i^{L,\dagger} = \xi_i^{L,\ddagger}$, then $B_i^L(\Xi^\dagger) \subseteq B_i^L(\Xi^\ddagger)$. Since $\check{F}^\dagger$ is tighter than $\check{F}^\ddagger$ and $\check{F}^\ddagger$ is inclusion monotonic,

$$\check{F}^\dagger(t, \boldsymbol{p}, B_i^L(\Xi^\dagger)) \subseteq \check{F}^\ddagger(t, \boldsymbol{p}, B_i^L(\Xi^\dagger)) \subseteq \check{F}^\ddagger(t, \boldsymbol{p}, B_i^L(\Xi^\ddagger)),$$

which implies that $\check{f}_i^{L,\ddagger}(t, \boldsymbol{p}, \Xi^\ddagger) \leq \check{f}_i^{L,\dagger}(t, \boldsymbol{p}, \Xi^\dagger)$. □

Under this new framework, several relaxation techniques can be used in multiple use cases, e.g. GMC and DMC. The following lemma shows that with the same relaxation technique, the use case in Section 5.5.4 generates the tightest state bounds and the use case in Section 5.5.1 produces the worst.

Lemma 5.8. *Consider an inclusion function $\check{F}^B : I \times \mathbb{R}^{n_u} \times \mathbb{IR}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ of $\boldsymbol{f}$ that is inclusion monotonic on $\mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x}$. Suppose that Π_i^L, Π_i^U are the flattening operators in Definition 5.4. Let $X^{B,1}, X^{B,2}, X^{B,3}, X^{B,4}$ be the state bounds generated with* (5.22), (5.27), (5.32), *and* (5.35), *respectively. Then, $X^{B,1}(t) \supseteq X^{B,2}(t) \supseteq X^{B,4}(t)$ and $X^{B,1}(t) \supseteq X^{B,3}(t) \supseteq X^{B,4}(t)$ for each $t \in I$.*

Proof. Define intervals $\Xi^1 \equiv [\boldsymbol{\xi}^{L,1}, \boldsymbol{\xi}^{U,1}], \Xi^2 \equiv [\boldsymbol{\xi}^{L,2}, \boldsymbol{\xi}^{U,2}], \Xi^3 \equiv [\boldsymbol{\xi}^{L,3}, \boldsymbol{\xi}^{U,3}]$ such that $\Xi^1 \subseteq \Xi^2 \subseteq \Xi^3 \subset \mathbb{IR}^{n_x}$, and consider any $t \in I$. If $\xi_i^{L,1} = \xi_i^{L,2}$, $B_i^L(\Xi^1) \subseteq B_i^L(\Xi^2)$. The inclusion monotonicity of $\check{F}^B$ ensures that, for any $\boldsymbol{p} \in U$,

$$\check{f}^L(t, U, B_i^L(\Xi^2)) \leq \check{f}^L(t, \boldsymbol{p}, B_i^L(\Xi^1)).$$

Analogously, it can be shown that when $\xi_i^{U,1} = \xi_i^{U,2}$,

$$\check{f}^U(t, U, B_i^U(\Xi^2)) \geq \check{f}^U(t, \boldsymbol{p}, B_i^U(\Xi^1)).$$

Thus, Theorem 5.4 guarantees that for any $t \in I$, $X^{B,1}(t) \supseteq X^{B,2}(t)$.

Similarly, for any $t \in I$, $\boldsymbol{p} \in U$, and $\boldsymbol{z} \in B_i^L(\Xi^2)$, if $\xi_i^{L,2} = \xi_i^{L,3}$,

$$\begin{aligned}\check{f}^L(t, \boldsymbol{p}, B_i^L(\Xi^3)) &\leq \check{f}^L(t, \boldsymbol{p}, B_i^L(\Xi^2)) \\ &\leq \check{f}^L(t, \boldsymbol{p}, \boldsymbol{z}).\end{aligned}$$

Hence,

$$\check{f}^L(t, U, B_i^L(\Xi^3)) \leq \check{f}^L(t, \boldsymbol{p}, B_i^L(\Xi^2)) \leq \min_{\boldsymbol{z} \in B_i^L(\Xi^2)} \check{f}^L(t, \boldsymbol{p}, \boldsymbol{z}).$$

Analogously,

$$\check{f}^U(t, U, B_i^L(\Xi^3)) \geq \check{f}^U(t, \boldsymbol{p}, B_i^L(\Xi^2)) \geq \max_{\boldsymbol{z} \in B_i^L(\Xi^2)} \check{f}^U(t, \boldsymbol{p}, \boldsymbol{z}).$$

According to Theorem 5.4, for any $t \in I$, $X^{B,2}(t) \supseteq X^{B,4}(t)$.

Similar arguments holds for $X^{B,1}(t) \supseteq X^{B,3}(t) \supseteq X^{B,4}(t)$ for any $t \in I$. □

5.6 Constructing Operators Π_i^L and Π_i^U with an *a priori* Enclosure

This section discusses reducing conservatism in the state bounds using an *a priori* enclosure G in order to construct tighter state bounds.

Assumption 5.11. *Assume that G in Definition 5.7 is described by continuously differentiable functions $\boldsymbol{g} : I \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_g}$ and $\boldsymbol{h} : I \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_h}$ such that*

$$G \equiv \{(t, \boldsymbol{x}, \boldsymbol{u}) \in I \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} : \quad \boldsymbol{g}(t, \boldsymbol{x}, \boldsymbol{u}) \leq \boldsymbol{0}, \quad \boldsymbol{h}(t, \boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{0}\}.$$

Such *a priori* knowledge can be utilized with a novel interval refining operator $I_G^B : I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$. The intuition behind this new operator is that the domain knowledge about states and inputs should be completely translated into refinement over state bounds. This operator is different from the operators $I_G : \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ in [118] and $I_A : \mathbb{IR}^{n_x+n_u} \to \mathbb{IR}^{n_x+n_u}$ in [127], though they are all designed to trim off regions that lie outside of *a priori* G. I_G typically reflects the constraints over the states, such as physical bounds and non-negativity. I_A refines both state bounds and input bounds, but this refinement is only effective when it is combined with the flattening operators B_i^L, B_i^U [127]. Our new framework may

support operator I_A by modifying (5.18) as follows:

$$\dot{x}_i^L(t) = \min_{\boldsymbol{p}\in U_i^L} \bar{f}_i^L(t, \boldsymbol{p}, X_i^{B,L}), \quad x_i^L(t_0) = x_{0,i}^L,$$
$$\dot{x}_i^U(t) = \max_{\boldsymbol{p}\in U_i^U} \bar{f}_i^U(t, \boldsymbol{p}, X_i^{B,U}), \quad x_i^U(t_0) = x_{0,i}^U,$$

where

$$(X_i^{B,L}, U_i^L) = I_A(B_i^L(X^B(t)), U),$$
$$(X_i^{B,U}, U_i^U) = I_A(B_i^U(X^B(t)), U).$$

However, we will proceed with the new operator I_G^B rather than I_A. The reason is that the above auxiliary system involves different input bounds U_i^L, U_i^U for each component of the auxiliary RHS function and these bounds are time-variant. Compare with the formulation in (5.18), this is more complex and may be computationally more expensive in numerical implementation.

Assumption 5.12. *Assume that I_G^B satisfies the following conditions:*

1. *For all $\Xi \in \mathbb{IR}^{n_x}$ with $(\Xi \cap G) \neq \emptyset$,*

$$(\Xi \cap G) \subset I_G^B(t, U, \Xi) \subset \Xi.$$

2. *$(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto I_G^B(t, U, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])$ is locally Lipschitz continuous.*

A valid choice of Π_i^L and Π_i^U satisfying Assumption 5.3 is then obtained in the following definition.

Definition 5.8. *Under Assumption 5.12, define* Π_i^L, Π_i^U *for each* $i \in \{1, \ldots, n_x\}$ *and all* $\Xi \in \mathbb{IR}^{n_x}$ *by*

$$\begin{aligned} \Pi_i^L(t, U, \Xi) &:= I_G^B(t, U, B_i^L(\Xi)), \\ \Pi_i^U(t, U, \Xi) &:= I_G^B(t, U, B_i^U(\Xi)). \end{aligned} \tag{5.40}$$

Based on the interval-Krawczyk method developed in [127, 124], a parametric-interval-Krawczyk method is proposed as follows. Define function $\mu : \{0, 1\} \times I \times \mathbb{IR}^{n_u} \times \mathbb{IR}^{n_x} \to \mathbb{R}$ such that

$$\mu(m, t, U, \Xi) := (-1)^m / \max\left(\epsilon, \left|\left[\frac{\partial g_i}{\partial y_k}\right](t, U, \Xi)\right|\right), \tag{5.41}$$

where ϵ is a fixed user-specified tolerance. Let $V \equiv (-\infty, 0]$. The refinement operator I_G^B is described in Algorithm 2.

Algorithm 2: An implementation of I_G^B

Function $I_G^B(t, \Xi, U)$**:**
  for $i \leftarrow 1, \ldots, n_g$ **do**
    for $j \leftarrow 1, \ldots, n_x$ **do**
      for $m \leftarrow 0, 1$ **do**
        $\boldsymbol{\xi} \leftarrow \mathrm{mid}(\Xi)$
        $\alpha \leftarrow [g_i](t, \boldsymbol{\xi}, U) + \sum_{k \neq j} \left[\frac{\partial g_i}{\partial y_k}\right](t, U, \Xi)(\Xi_k - \xi_k) + V$
        $\widehat{\Xi}_i \leftarrow \xi_i + \mu(m, t, U, \Xi)\alpha$
          $+(1 + \mu(m, t, U, \Xi)\left[\frac{\partial g_i}{\partial y_j}\right](t, U, \Xi)(\Xi_j - \xi_j)$
        $\xi_j^L \leftarrow \min(\max(\widehat{\xi}_j^L, \xi_j^L), \xi_j^U)$
        $\xi_j^U \leftarrow \max(\min(\widehat{\xi}_j^U, \xi_j^U), \xi_j^L)$
      end
    end
  end
  for $i \leftarrow 1, \ldots, n_h$ **do**
    for $j \leftarrow 1, \ldots, n_x$ **do**
      for $m \leftarrow 0, 1$ **do**
        $\boldsymbol{\xi} \leftarrow \mathrm{mid}(\Xi)$
        $\alpha \leftarrow [h_i](t, \boldsymbol{\xi}, U) + \sum_{k \neq j} \left[\frac{\partial h_i}{\partial y_k}\right](t, U, \Xi)(\Xi_k - \xi_k)$
        $\widehat{\Xi}_i \leftarrow \xi_i + \mu(m, t, U, \Xi)\alpha$
          $+(1 + \mu(m, t, U, \Xi)\left[\frac{\partial h_i}{\partial y_j}\right](t, U, \Xi)(\Xi_j - \xi_j)$
        $\xi_j^L \leftarrow \min(\max(\widehat{\xi}_j^L, \xi_j^L), \xi_j^U)$
        $\xi_j^U \leftarrow \max(\min(\widehat{\xi}_j^U, \xi_j^U), \xi_j^L)$
      end
    end
  end
  return Ξ

Theorem 5.5. *Choose any $\epsilon \in \mathbb{R}_{\geq 0}$ and $t \in I$. If $[\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{y}}], [\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{y}}] : I \times \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_u} \mapsto \mathbb{IR}^{n_x}$ are inclusion functions of $\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{y}}$, $\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{y}}$, respectively, and are locally Lipschitz continuous. Then, I_G^B defined in Algorithm 2 satisfies Assumption 5.12.*

Proof. This result of I_G^B satisfying Assumption 5.12 is analogous to [127, Theorem 4] and the proof is similar. □

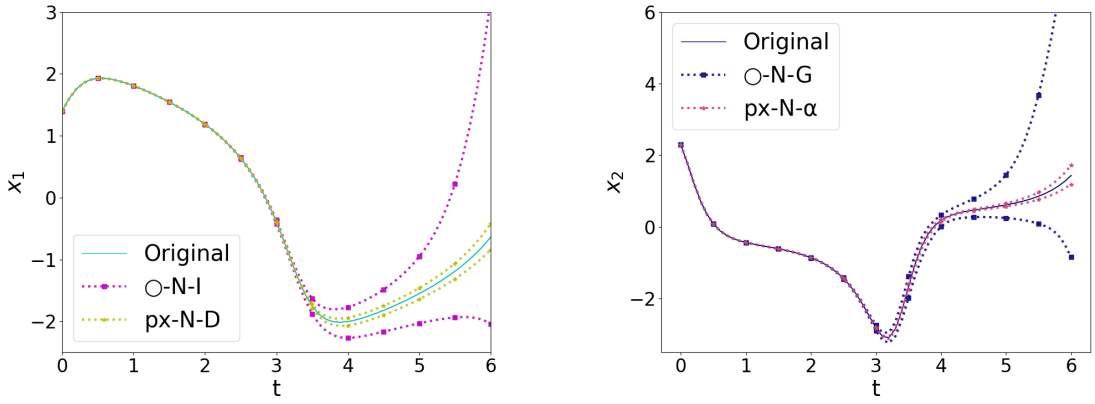
5.7 Numerical Examples

This section presents numerical examples in which state bounds of nonlinear dynamic systems are generated using our new framework (5.18). Functions $\boldsymbol{f}^L$, $\boldsymbol{f}^U$ in the RHS of (5.18) are constructed with various methods introduced in Section 5.5. An implementation of these approaches has been developed in Julia v1.5.3 [20]. The auxiliary system of ODEs (5.18) is solved with DifferentiableEquations.jl [104]. McCormick-type relaxations, including MC, GMC, and $\mathscr{C}^1$-DMC, are generated with McCormick.jl [152]. JuMP v0.21.4 [49] is used as the interface to optimzation solvers; CPLEX v12.10 is used to solve LPs, and IPOPT v3.13.2 is used to solve NLPs. All numerical experiments were performed on a Windows 10 machine with a 3.6 GHz Ryzen 5 2600X CPU and 8 GB memory.

Example 5.1. *This example involves the Van der Pol oscillator, which is a classic dynamic system that has been widely studied in electrical engineering and biological science. Relaxations of this system were obtained by [126]. Here, a two-dimensional form with uncertainty in both initial conditions and RHS functions is considered:*

$$\begin{aligned} \dot{x}_1(t,\boldsymbol{u}) &= x_1, & x_1(t_0,\boldsymbol{u}) &= u_1(t_0), \\ \dot{x}_2(t,\boldsymbol{u}) &= u_1(1-x_1^2)x_2 - x_1, & x_2(t_0,\boldsymbol{u}) &= u_2(t_0), \end{aligned}$$

where $U \equiv [1.399, 1.400] \times [2.299, 2.300]$, $\boldsymbol{u} = (u_1, u_2) \in \mathcal{U}$, *and* $I \equiv [t_0, t_f] = [0, 6]$.



(a) (b)

Figure 5.2: The state bound trajectories generated with different methods for (a) x_1 and (b) x_2 (dotted) in Example 5.1. Solid lines are real trajectories.

State bounds were computed for state variables x_1 and x_2 with methods based on interval extensions and methods that optimize states and parameters. As shown in Figure 5.2, approaches that optimize state and parameters, i.e. px-N-D and px-N-α, construct tighter state bounds than the optimization-free methods, i.e. ○-N-I and ○-N-G. Note that except ○-N-I which was developed by Harrison [59], the other three methods are all newly discovered using our new framework (5.18).

Example 5.2. *This example is an ODE system with a quadratic RHS adapted from [31]:*

$$\dot{x}_1(t,\boldsymbol{u}) = (x_1-u_1)^2 - (x_2-u_1)^2,\ x_1(t_0) = 2.2,$$
$$\dot{x}_2(t,\boldsymbol{u}) = (x_1-u_2)^2 - (x_2-u_2)^2,\ x_2(t_0) = 1.8,$$

where $U \equiv [-2,2]\times[-1,3]$, $\boldsymbol{u} = (u_1,u_2) \in \mathcal{U}$, *and* $I \equiv [t_0,t_f] = [0.0,0.2]$.

Two new state bounding methods, px-N-α and px-N-EC, are used in this example. These two new methods are developed based on αBB relaxations and edge-concave relaxations, respectively, using the new framework. The trajectories of theses methods are plotted in Figure 5.3. For this particular problem, px-N-EC generates tighter state bounds than px-N-α. Furthermore, both of these two new methods generate tighter state bounds than Harrison's method ○-N-I.

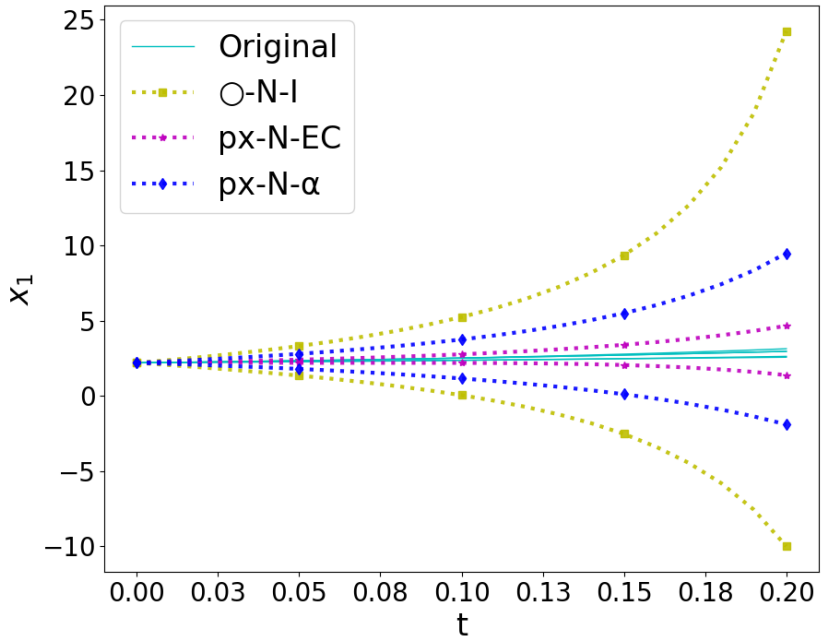


Figure 5.3: The state bound trajectories (dotted) generated with ◯-N-I (square), px-N-EC (star), and px-N-α (diamond) for x_1 in Example 5.2. Solid lines are real trajectories.

Besides those three methods illustrated in Figure 5.3, additional methods from Table 5.2 were tested with this example to evaluate their computational performance. Each method was repeated 10 times and the average computing time is

displayed in Table 5.3. Note that the newly developed method, ○-N-G, has a computing time that is similar to Harrison's method with this implementation. The method that involves nonsmooth nonlinear optimization (i.e. x-N-G) took longer computing time than a similar method than involves smooth nonlinear optimization (i.e. x-N-D). Moreover, since the original ODE RHS function is quadratic, its αBB relaxation is also a quadratic function and can be minimized using CPLEX. The computing time of px-N-α(NLP) and px-N-α(QP) confirms that, solving the embedded optimization problems as a QP using CPLEX is faster than solving it as an NLP using IPOPT. Lastly, minimizing the edge-concave relaxations in px-N-EC were optimized by checking all vertices of the domain box. Since this example has a small number of state variables and controls, this optimization process was not computationally expensive.

Table 5.3: Computing time of Example 5.2 with different methods

Bounding method	○-N-I	○-N-G	x-N-G	x-N-D
Average time (s)	0.0043	0.0042	15.243	4.8706
Bounding method	px-N-D	px-N-EC	px-N-α(NLP)	px-N-α(QP)
Average time (s)	14.143	1.2314	4.7906	1.6311

Example 5.3. *This example describes the dynamic of an anaerobic digestion process originally developed in [18]. Enclosures for the reachable set of this model have been constructed*

in [146, 127].

$$
\begin{aligned}
\dot{X}_1 &= (\mu_1(S_1) - \alpha D)X_1, \\
\dot{X}_2 &= (\mu_2(S_2) - \alpha D)X_2, \\
\dot{S}_1 &= D(S_1^{in} - S_1) - k_1\mu_1(S_1)X_1, \\
\dot{S}_2 &= D(S_2^{in} - S_2) + k_2\mu_1(S_1)X_1 - k_3\mu_2(S_2)X_2, \\
\dot{Z} &= D(Z^{in} - Z), \\
\dot{C} &= D(C^{in} - C) - q_{CO_2} + k_4\mu_1(S_1)X_1 + k_5\mu_2(S_2)X_2,
\end{aligned}
\tag{5.42}
$$

where

$$
\begin{aligned}
q_{CO_2} &= k_L a(C + S_2 - Z - K_H P_{CO_2}, \\
P_{CO_2} &= \frac{\phi_{CO_2} - \sqrt{\phi_{CO_2}^2 - 4K_H P_t(C + S_2 - Z)}}{2K_H}, \\
\phi_{CO_2} &= C + S_2 - Z + K_H P_t + \frac{k_6}{k_L a}\mu_2(S_2)X_2, \\
\mu_1(S_1) &= \bar{\mu}_1 \frac{S_1}{S_1 + K_{S_1}}, \\
\mu_2(S_2) &= \bar{\mu}_2 \frac{S_2}{S_2 + K_{S_2} + S_2^2/K_{I_2}}.
\end{aligned}
$$

The uncertain parameters and initial conditions are summarized in Table 5.4. The other parameters are constants and their values can be found in [146].

A higher-dimensional "lifted" variant of the model (5.42) was developed in

Table 5.4: Parameters and initial conditions for Example 5.3

Symbol	Value	Unit
k_1	$[42.14, 42.98]$	g(COD) g(cell)$^{-1}$
k_2	$[116.5, 118.24]$	mmol g(cell)$^{-1}$
$X_1(t_0)$	$[0.49, 0.51]$	g(COD) L^{-1}
$X_2(t_0)$	$[0.98, 1.02]$	mmol L^{-1}
$C(t_0)$	$[39.2, 40.8]$	mmol L^{-1}
$S_1(t_0)$	1	mmol L^{-1}
$S_2(t_0)$	5	mmol L^{-1}
$Z(t_0)$	50	mmol L^{-1}

[127]. It has two redundant state variables

$$\begin{aligned} \dot{N}_1 &= D(S_1^{\text{in}} + S_1(\alpha - 1) - \alpha N_1), \\ \dot{N}_2 &= D(S_2^{\text{in}} + S_2(\alpha - 1) - \alpha N_2). \end{aligned} \tag{5.43}$$

The new system consisting of (5.42) and (5.43) satisfies an *a priori* enclosure such that

$$\begin{aligned} 0 &= -N_1 + k_1 X_1 + S_1, \\ 0 &= -N_2 - k_2 X_1 + k_3 X_2 + S_2. \end{aligned} \tag{5.44}$$

(5.44) follows the description of *a priori* enclosure G in Assumption 5.12, so that the interval refinement operator I_G^B can be used to reduce the conservatism in state bounds. Figure 5.4 compares three different NIE-based methods. The first one did not consider *a priori* enclosure; the other two incorporated the *a priori* enclosure G with operators I_G^B and I_A, respectively. It was observed that the *a priori* enclosure G prevented the state bounds from diverging within the time horizon. Moreover,

operators I_G^B and I_A provided same refinement to the state bounds of this particular system, and their trajectory overlaps in Figure 5.4.

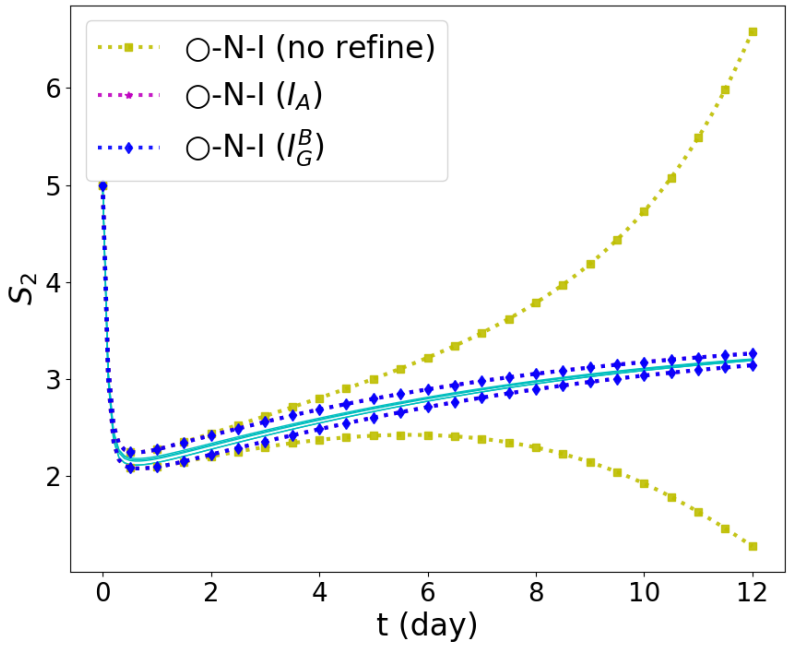


Figure 5.4: The state bound trajectories of S_2 in Example 5.3 (dotted) generated with no refinement (square) and refinement operators I_A (star) and I_G^B (diamond). Solid lines are real trajectories.

5.8 Conclusions

This work concerns the problem of bounding the reachable set of a nonlinear dynamic system with uncertainty. A novel systematic framework (5.18) was developed based on the theory of differential inequalities in Section 5.4. It employs an auxiliary system of ODEs with optimization problems embedded in the RHS. The solutions of these auxiliary ODEs are componentwise lower and upper bounds of

the original system. The intuition behind these embedded optimization problems is to generate lower and upper bounds of the original ODE RHS function by optimizing relaxations of the original ODE RHS function. Four different use cases of this new framework were presented in Section 5.5, and they were distinguished by the decision variables in the embedded optimization problems. Moreover, various methods for generating relaxations of the original ODE RHS function were introduced in each use case. In particular, some of these methods lead to an embedded optimization problem that is trivial to solve and does not require any numerical optimization solver. In Section 5.6, we adapted an approach from [124] to incorporate the *a priori* knowledge of the original system in order to produce tighter bounds. Lastly, a proof-of-concept implementation was developed to support all four use cases. Numerical examples were presented in Section 5.7 to illustrate that several new methods discovered with this framework are capable of constructing tight bounds for the original system efficiently.

The benefits of our new framework for enclosing reachable sets are multi-fold. First, this framework describes a general strategy for bounding nonlinear ODEs using differential inequality. It not only includes several established methods such as [59, 118, 62], but also inspires the discovery of various new methods as long as Assumption 5.1 is satisfied. Second, this framework is versatile. It supports various relaxation techniques that have never been used to construct state bounds, such as GMC, DMC, αBB relaxations, and edge-concave relaxations. Some of the techniques lead to tighter state bounds than established methods, which is fundamental to reachability analysis and global optimization. Third, this framework allows the usage of *a priori* knowledge of the original system for generating tighter

state bounds. Similar approaches have been studied in many researches [126, 146, 121] and were demonstrated to be effective.

For future work, we are interested in using the new state bounds developed in this work to construct convex relaxations for parametric ODEs. These convex relaxations are fundamental to the deterministic global optimization of dynamic systems [131]. To generate convex relaxations for parametric ODEs using differential inequalities, we first need to construct valid state bounds. While established methods [120, 131, 27] typically use Harrison's method [59] to provide state bounds, we expect that the tighter state bounds developed in this work will lead to tighter convex relaxations for parametric ODEs, and therefore help global optimization algorithms converge faster [42].

Chapter 6

Enclosing Reachable Sets for Nonlinear Control Systems using Complementarity-Based Intervals

This chapter is reproduced from a published conference proceeding [31].

6.1 Introduction

The problem of interest in this paper is to compute tight bounds for the reachable set of nonlinear dynamic systems represented as system of parametric ordinary differential equations (ODEs) with uncertain inputs, parameters, and initial conditions. Such enclosures are important in many applications, including state estimation [69], parameter estimation [128], safety verification [66, 142], fault detection [82], and global dynamic optimization [101, 129]. Various strategies have been

proposed to enclose this reachable set, such as solving the Hamilton-Jacobi equations [91], conservatively linearizing nonlinear models [8], constructing zonotopes [78, 155] or ellipsoids [79], and computing validated solutions [96, 84]. This paper focuses on another category of methods that are based on differential inequalities [148].

Differential inequality-based methods generate time-varying interval enclosures for the original nonlinear dynamic system by constructing an auxiliary dynamic system and solving this numerically. The solutions of the auxiliary system are component-wise lower and upper bounds for the original system. Differential inequality-based methods require valid bounding information for the original system's right-hand side (RHS) function. [59] first proposed to use natural interval extensions (NIE) [94] to calculate interval bounds of the RHS function automatically. This strategy was extended using affine relaxation techniques for tighter enclosures [128]. [62] introduced a method to bound the RHS function with the solutions of linear programs (LPs). These LPs optimize piecewise-affine relaxations of the original RHS function that are derived with a special relaxation scheme to ensure the Lipschitz continuity. [37] presented another differential inequality-based approach that applies Taylor series expansion to the original system. Besides the various techniques for constructing an auxiliary bounding system, another direction of research in this area involves generating less conservative enclosures by exposing the "hidden constraints" of the original system, such as physical bounds and implicit conservation laws [118, 126]. This approach may require specialized knowledge of the system of interest to formulate effective constraints for refining the enclosures.

In this work, we propose a novel differential inequality-based method for computing enclosures for nonlinear control systems. Bounding information for the RHS function is obtained by optimizing its convex relaxations. This is distinct from the LP-based method by [62] in which the relaxations are limited to a special type of convex piecewise-affine under-estimators. Our new approach, on the other hand, is applicable to a broad range of convex relaxations. Moreover, complementarity formulations are developed in an effort to solve the optimization problems efficiently. Examples are presented for illustration.

The following notation conventions are used in this paper. Vectors are denoted with boldface lower-case letters (e.g. $\boldsymbol{x}$). Given vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, inequalities such as $\boldsymbol{x} < \boldsymbol{y}$ or $\boldsymbol{x} \leq \boldsymbol{y}$ are to be interpreted component-wise. Convexity of a vector-valued function $\boldsymbol{f}$ refers here to convexity of all components f_i. A matrix is denoted with boldface upper-case letters (e.g. $\boldsymbol{A}$), and its elements are represented by corresponding lower case letters with subscripts indicating the row and column (e.g. a_{ij}). An interval in $\mathbb{R}^n$ is a nonempty subset of $\mathbb{R}^n$ of the form $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}\}$, which is denoted as $[\boldsymbol{a}, \boldsymbol{b}]$. $\mathbb{IR}^n$ denotes the set of all intervals in $\mathbb{R}^n$.

6.2 Problem Statement

Consider $t_0, t_f \in \mathbb{R}$ with $t_0 < t_f$, and define $I := [t_0, t_f]$. Let $U := [\boldsymbol{u}^L, \boldsymbol{u}^U] \subset \mathbb{R}^{n_u}$ be an interval, and $D \subset \mathbb{R}^{n_x}$ be open. Denote the space of all Lebesgue integrable functions $\boldsymbol{h} : I \to \mathbb{R}^n$ as $L^n(I)$. Let $\tilde{U} := \{\boldsymbol{u} \in L^{n_u}(I) : \boldsymbol{u}(t) \in U, t \in I\}$ be a set of admissible controls, and $X_0 := [\boldsymbol{x}_0^L, \boldsymbol{x}_0^U] \in D$ be a set of admissible initial conditions. Given a continuous mapping $\boldsymbol{f} : I \times U \times D \to \mathbb{R}^{n_x}$ for which $\boldsymbol{f}(t, \cdot, \cdot)$ is

twice-continuously differentiable for each $t \in I$, consider an initial-value problem

$$\begin{aligned} \dot{\boldsymbol{x}}(t, \boldsymbol{u}, \boldsymbol{x}_0) &= \boldsymbol{f}(t, \boldsymbol{u}(t), \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0)), \quad \forall t \in (t_0, t_f], \\ \boldsymbol{x}(t_0, \boldsymbol{u}, \boldsymbol{x}_0) &= \boldsymbol{x}_0, \end{aligned} \tag{6.1}$$

where $(\boldsymbol{u}, \boldsymbol{x}_0) \in \tilde{U} \times X_0$, and where dotted quantities indicate time-derivatives (e.g. $\dot{\boldsymbol{x}} \equiv \frac{\partial \boldsymbol{x}}{\partial t}$).

Under these conditions, the ordinary differentiable equation (ODE) (6.1) is guaranteed to have a unique solution by the Picard-Lindelöf Theorem, summarized as [60, Theorem 1.1, Chapter II].

The objective of this work is to compute tight time-varying interval bounds for $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0)$ in (6.1). Here, we use the terminology proposed by [118] to describe such enclosures.

Definition 6.1 (State bounds)**.** *Functions $\boldsymbol{x}^L, \boldsymbol{x}^U : I \to \mathbb{R}^{n_x}$ are* state bounds *for the ODE* (6.1) *if*

$$\boldsymbol{x}^L(t) \leq \boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \leq \boldsymbol{x}^U(t), \quad \forall (t, \boldsymbol{u}, \boldsymbol{x}_0) \in I \times \tilde{U} \times X_0.$$

Let $X^B : I \to \mathbb{IR}^{n_x}$ denote the corresponding interval function: $X^B(t) := [\boldsymbol{x}^L(t), \boldsymbol{x}^U(t)]$ for each $t \in I$.

6.3 Background

The following fundamental differential inequality theorem was presented in [59].

Proposition 6.1. *Let $\boldsymbol{x}^L, \boldsymbol{x}^U : I \to \mathbb{R}^{n_x}$ satisfy the following conditions.*

1. $\boldsymbol{x}_0 \in X^B(t_0)$,

2. *For a.e.* $t \in I$ *and each* $i \in \{1, \ldots, n_x\}$,

$$\dot{x}_i^L(t) \leq \min_{\substack{\boldsymbol{z} \in X^B(t),\, z_i = x_i^L(t), \\ \boldsymbol{u} \in \tilde{U}}} f_i(t, \boldsymbol{u}, \boldsymbol{z}), \tag{6.2a}$$

$$\dot{x}_i^U(t) \geq \max_{\substack{\boldsymbol{z} \in X^B(t),\, z_i = x_i^U(t), \\ \boldsymbol{u} \in \tilde{U}}} f_i(t, \boldsymbol{u}, \boldsymbol{z}), \tag{6.2b}$$

Then $\boldsymbol{x}(t, \boldsymbol{u}, \boldsymbol{x}_0) \in X^B(t)$ *for all* $(t, \boldsymbol{u}) \in I \times \tilde{U}$.

Based on the above result, [59] suggested to compute $\dot{\boldsymbol{x}}^L$ and $\dot{\boldsymbol{x}}^U$ by applying NIE to $\boldsymbol{f}$. This method generates an *inclusion function* of $\boldsymbol{f}$ that is independent of $\boldsymbol{u}$ and $\boldsymbol{z}$, satisfying the following definition.

Definition 6.2 (Inclusion function)**.** *Let* $S \in \mathbb{IR}^n$ *and* $\boldsymbol{h} : S \to \mathbb{R}^m$. *An interval function* $H = [\boldsymbol{h}^L, \boldsymbol{h}^U] : \mathbb{IR}^n \to \mathbb{IR}^m$ *is a* inclusion function *of* $\boldsymbol{h}$ *on* S *if*

$$\{\boldsymbol{h}(\boldsymbol{z}) : \boldsymbol{z} \in Z\} \subseteq H(Z), \quad \forall Z \subseteq S.$$

Harrison also noted that choosing $\dot{\boldsymbol{x}}^L$ and $\dot{\boldsymbol{x}}^U$ close to the bounds given by (6.2) will empirically benefit the generated state bounds. A recent comparison result for ODE solutions [130] also confirms this. So, we explore the possibility of providing bounding information of f_i with convex relaxation, which is typically closer to the original function compared with NIE.

Definition 6.3 (Convex relaxation)**.** *Let* $Z \in \mathbb{IR}^n$ *and* $\boldsymbol{h} : Z \to \mathbb{R}^m$. *Then:*

1. $\boldsymbol{h}^{cv} : Z \to \mathbb{R}^m$ *is a* convex relaxation *of* $\boldsymbol{h}$ *on* Z *if* $\boldsymbol{h}^{cv}(\boldsymbol{z}) \leq \boldsymbol{h}(\boldsymbol{z})$ *for all* $\boldsymbol{z} \in Z$ *and* $\boldsymbol{h}^{cv}$ *is convex on* Z.

2. $\boldsymbol{h}^{cc} : Z \to \mathbb{R}^m$ *is a* concave relaxation *of* $\boldsymbol{h}$ *on* Z *if* $\boldsymbol{h}^{cc}(\boldsymbol{z}) \geq \boldsymbol{h}(\boldsymbol{z})$ *for all* $\boldsymbol{z} \in Z$ *and* $\boldsymbol{h}^{cc}$ *is concave on* Z.

3. *The interval function* $H = [\boldsymbol{h}^{cv}, \boldsymbol{h}^{cc}]$ *is called a* convex inclusion function *of* $\boldsymbol{h}$ *on* Z.

6.4 New Formulation

Define an interval function $F^R = [\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}] : I \times U \times \mathbb{R}^{n_x} \to \mathbb{IR}^{n_x}$, and recall the considered ODE system (6.1).

Assumption 6.1. *Suppose that the interval function* $F^R = [\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}]$ *has the following properties:*

1. $\boldsymbol{f}^{cv}$ *and* $\boldsymbol{f}^{cc}$ *are continuous,*

2. $\boldsymbol{f}^{cv}$ *and* $\boldsymbol{f}^{cc}$ *are locally Lipschitz continuous in* $\boldsymbol{x}$*, uniformly in* $(t, \boldsymbol{p})$*,*

3. $F^R(t, \cdot, \cdot)$ *is an convex inclusion function of* $\boldsymbol{f}(t, \cdot, \cdot)$ *on* $U \times D$ *for a.e.* $t \in I$.

Definition 6.4. *Under Assumption 6.1, define an interval function* $F^B = [\boldsymbol{f}^L, \boldsymbol{f}^U] : I \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ *such that, for each* $i \in \{1, \ldots, n_x\}, t \in I$*, and* $\Xi = [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U] \in \mathbb{IR}^{n_x}$*,*

$$f_i^L(t, \Xi) = \min_{\substack{\boldsymbol{z} \in \Xi,\, z_i = \xi_i^L, \\ \boldsymbol{p} \in U}} f_i^{cv}(t, \boldsymbol{p}, \boldsymbol{z}), \tag{6.3a}$$

$$\text{and} \quad f_i^U(t, \Xi) = \max_{\substack{\boldsymbol{z} \in \Xi,\, z_i = \xi_i^U, \\ \boldsymbol{p} \in U}} f_i^{cc}(t, \boldsymbol{p}, \boldsymbol{z}). \tag{6.3b}$$

Define the following auxiliary ODE system over $t \in I$:

$$\dot{\boldsymbol{x}}^L(t) = \boldsymbol{f}^L(t, X^B(t)),\ \boldsymbol{x}^L(t_0) = \boldsymbol{x}_0^L, \tag{6.4a}$$

$$\dot{\boldsymbol{x}}^U(t) = \boldsymbol{f}^U(t, X^B(t)),\ \boldsymbol{x}^U(t_0) = \boldsymbol{x}_0^U. \tag{6.4b}$$

6.4.1 Existence and uniqueness

This section shows that the auxiliary ODE system (6.4) has exactly one solution under mild assumptions.

Theorem 6.1. *Under Assumption 6.1, the ODE* (6.4) *has unique solutions.*

Proof. Define $\boldsymbol{g}^{cv}, \boldsymbol{g}^{cc} : I \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ such that

$$g_i^{cv}(t, \boldsymbol{z}) = \min_{\boldsymbol{p} \in U} f_i^{cv}(t, \boldsymbol{p}, \boldsymbol{z}),$$

$$g_i^{cc}(t, \boldsymbol{z}) = \max_{\boldsymbol{p} \in U} f_i^{cc}(t, \boldsymbol{p}, \boldsymbol{z}),$$

for each $i \in \{1, \ldots, n_x\}$. Then, (6.3) becomes

$$f_i^L(t, \Xi) = \min_{\boldsymbol{z} \in \Xi,\, z_i = \xi_i^L} g_i^{cv}(t, \boldsymbol{z}),$$

$$f_i^U(t, \Xi) = \max_{\boldsymbol{z} \in \Xi,\, z_i = \xi_i^U} g_i^{cc}(t, \boldsymbol{z}). \tag{6.5}$$

According to Assumption 6.1 and [40, Theorem 2.1], $(\boldsymbol{g}^{cv}, \boldsymbol{g}^{cc})$ are Lipschitz continuous in $\boldsymbol{z}$, uniformly in t. Moreover, because $\boldsymbol{g}^{cv}(t, \cdot)$ and $\boldsymbol{g}^{cc}(t, \cdot)$ are readily verified to be convex and concave, respectively, Proposition 2 from [131] ensures that $(\boldsymbol{f}^L, \boldsymbol{f}^U)$ in (6.5) are Lipschitz continuous with respect to $\boldsymbol{\xi}^L$ and $\boldsymbol{\xi}^U$, uniformly

for $t \in I$. Then, the existence and uniqueness of (6.4) is guaranteed by the Picard-Lindelöf Theorem [60, Theorem 1.1, Chapter II]. □

6.4.2 Bounding the original system

This section shows that the auxiliary ODE (6.4) provides valid state bounds for (6.1).

Theorem 6.2. *Under Assumption 6.1, let* $(\boldsymbol{x}^L, \boldsymbol{x}^U)$ *be solutions of the ODE* (6.4)*. Then,* $(\boldsymbol{x}^L, \boldsymbol{x}^U)$ *are state bounds of ODE* (6.1)*.*

Proof. It suffices to show that the two requirements in Proposition 6.1 are satisfied by $(\boldsymbol{x}^L, \boldsymbol{x}^U)$. First, $\boldsymbol{x}_0 \in X^B(t_0)$ is ensured by the construction of auxiliary ODE system (6.4). Second, Condition 3 in Assumption 6.1 guarantees that, for a.e. $t \in I$ and any $(\boldsymbol{p}, \boldsymbol{z}) \in U \times D$,

$$\boldsymbol{f}^{cv}(t, \boldsymbol{p}, \boldsymbol{z}) \leq \boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{z}).$$

So for a.e. $t \in I$, each $\Xi \in \mathbb{IR}^{n_x}$, and each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} \dot{x}_i^L(t) = f_i^L(t, \Xi) &= \min_{\substack{\boldsymbol{z} \in \Xi,\ z_i = \zeta_i^L, \\ \boldsymbol{p} \in U}} f_i^{cv}(t, \boldsymbol{p}, \boldsymbol{z}) \\ &\leq \min_{\substack{\boldsymbol{z} \in \Xi,\ z_i = \zeta_i^L, \\ \boldsymbol{u} \in \tilde{U}}} f_i(t, \boldsymbol{u}, \boldsymbol{z}). \end{aligned}$$

Similarly,

$$\dot{x}_i^U(t) = f_i^U(t,\Xi) \geq \max_{\substack{\boldsymbol{z}\in\Xi,\, z_i=\zeta_i^U,\\ \boldsymbol{u}\in\tilde{U}}} f_i(t,\boldsymbol{u},\boldsymbol{z}).$$

The second condition in Proposition 6.1 is thus satisfied. □

6.5 Complementarity Reformulation

This section derives a complementarity reformulation of (6.3) based on Karush-Kuhn-Tucker (KKT) conditions. To simplify notation, denote $\boldsymbol{y} = (\boldsymbol{x}, \boldsymbol{p}) \in \mathbb{R}^{n_x+n_u}$ in the remainder of this paper. We also define the following operators as did [118].

Definition 6.5. *For each $i \in \{1,\ldots,n_x\}$, define flattening operators $\underline{B}_i, \overline{B}_i : \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ such that,*

1. *$\underline{B}_i([\boldsymbol{\phi},\boldsymbol{\psi}]) = [\boldsymbol{\phi},\boldsymbol{\psi}']$, where $\psi_i' = \phi_i$, and $\psi_k' = \psi_k$ for all $k \in \{1,\ldots,n_x\}\backslash\{i\}$,*
2. *$\overline{B}_i([\boldsymbol{\phi},\boldsymbol{\psi}]) = [\boldsymbol{\phi}',\boldsymbol{\psi}]$, where $\phi_i' = \psi_i$, and $\phi_k' = \phi_k$ for all $k \in \{1,\ldots,n_x\}\backslash\{i\}$.*

The optimization problem in (6.3a) can be then reformulated as follows; with $\Xi = [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U]$, $[\boldsymbol{\phi}_{(i)}^L, \boldsymbol{\phi}_{(i)}^U] = B_i^L([(\boldsymbol{\xi}^L, \boldsymbol{u}^L),(\boldsymbol{\xi}^U, \boldsymbol{u}^U)])$,

$$\begin{aligned} \min_{\boldsymbol{y}} \quad & f_i^{cv}(t,\boldsymbol{y}), \\ \text{s.t.} \quad & \boldsymbol{\phi}_{(i)}^L \leq \boldsymbol{y} \leq \boldsymbol{\phi}_{(i)}^U. \end{aligned} \tag{6.6}$$

The corresponding KKT conditions are:

$$\begin{aligned} \nabla_{\underline{\boldsymbol{y}}^*} f_i^{cv}(t, \underline{\boldsymbol{y}}^*) + \overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}} &= \mathbf{0}, \\ \boldsymbol{\phi}_{(i)}^L \leq \underline{\boldsymbol{y}}^* &\leq \boldsymbol{\phi}_{(i)}^U, \\ \overline{\boldsymbol{\mu}} \geq \mathbf{0}, \quad \underline{\boldsymbol{\mu}} &\geq \mathbf{0}, \\ (\overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}})^\top \underline{\boldsymbol{y}}^* + \underline{\boldsymbol{\mu}}^\top \boldsymbol{\phi}_{(i)}^L - \overline{\boldsymbol{\mu}}^\top \boldsymbol{\phi}_{(i)}^U &= 0. \end{aligned} \tag{6.7}$$

Under Assumption 6.1, (6.6) is a box-constrained convex optimization problem which satisfies the linearity constraint qualification. So, satisfying the condition (6.7) is equivalent to $\boldsymbol{y}^*$ solving (6.6) directly. A similar formulation can be derived for the optimization problem in (6.3b): with $[\boldsymbol{\psi}_{(i)}^L, \boldsymbol{\psi}_{(i)}^U] = B_i^U([(\boldsymbol{\xi}^L, \boldsymbol{u}^L), (\boldsymbol{\xi}^U, \boldsymbol{u}^U)])$,

$$\begin{aligned} \nabla_{\overline{\boldsymbol{y}}^*} f_i^{cc}(t, \overline{\boldsymbol{y}}^*) - \overline{\boldsymbol{\nu}} + \underline{\boldsymbol{\nu}} &= \mathbf{0}, \\ \boldsymbol{\psi}_{(i)}^L \leq \overline{\boldsymbol{y}}^* &\leq \boldsymbol{\psi}_{(i)}^U, \\ \overline{\boldsymbol{\nu}} \geq \mathbf{0}, \quad \underline{\boldsymbol{\nu}} &\geq \mathbf{0}, \\ (\overline{\boldsymbol{\nu}} - \underline{\boldsymbol{\nu}})^\top \overline{\boldsymbol{y}}^* + \underline{\boldsymbol{\nu}}^\top \boldsymbol{\psi}_{(i)}^L - \overline{\boldsymbol{\nu}}^\top \boldsymbol{\psi}_{(i)}^U &= 0. \end{aligned} \tag{6.8}$$

So (6.3) can be reformulated as

$$\begin{aligned} f_i^L(t, \Xi) &= f_i^{cv}(t, \underline{\boldsymbol{y}}^*), \\ \text{and} \quad f_i^U(t, \Xi) &= f_i^{cc}(t, \overline{\boldsymbol{y}}^*), \end{aligned} \tag{6.9}$$

where $\underline{\boldsymbol{y}}^*$ and $\overline{\boldsymbol{y}}^*$ are the KKT points in (6.7) and (6.8), respectively.

The dynamic system (6.4) with its RHS defined in (6.9) can thus be considered as a mixed nonlinear complementarity system (NCS), for which many numerical

algorithms have been developed [116]. In particular, a software platform Siconos [1] has been developed to solve NCSs efficiently.

6.6 Constructing convex inclusion functions

According to Theorem 6.2, state bounds for (6.1) can be computed by constructing a convex inclusion function of $\boldsymbol{f}$, $F^R = [\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}]$, that satisfies Assumption 6.1. One way to construct F^R is to use convex (concave) envelopes, which are defined as the supremum (infimum) of all convex under-estimators (concave over-estimators) of $\boldsymbol{f}$. In this case, we obtain the tightest bounds that are consistent with Proposition 6.1. However, the convex envelope is generally cumbersome or impossible to evaluate for multivariate functions. A practical and computationally simpler method for generating such a convex inclusion function is to derive αBB relaxations [9] for $\boldsymbol{f}$. Other relaxation approaches, such as McCormick relaxation [123, 72, 73], are also applicable.

6.6.1 αBB relaxation

αBB relaxation is an established technique [3] for constructing convex under-estimators for general nonconvex twice differentiable functions. To construct a relaxation, a negative convex quadratic term is added to the original function, $h : \mathbb{R}^n \to \mathbb{R}^m$:

$$h^{cv}(\boldsymbol{z}) := h(\boldsymbol{z}) + \sum_{i=1}^{n} \alpha_i (z_i^L - z_i)(z_i^U - z_i),$$

where $\boldsymbol{z}^L$ and $\boldsymbol{z}^U$ are the lower and upper bounds of $\boldsymbol{z}$, and $\boldsymbol{\alpha} \in \mathbb{R}^n$ is a constant vector that is determined by h, $\boldsymbol{z}^L$, and $\boldsymbol{z}^U$. [3] propose an approach to construct a

valid $\boldsymbol{\alpha}$ that ensures the convexity of the under-estimator h^{cv}. The first step of this approach is to determine a symmetric interval matrix $[\boldsymbol{H}]$ such that

$$\nabla^2 h(\boldsymbol{z}) \in [\boldsymbol{H}], \qquad \forall \boldsymbol{z} \in [\boldsymbol{z}^L, \boldsymbol{z}^U].$$

This can be accomplished by applying NIE to the Hessian matrix of h, denoted as $\boldsymbol{H}$. Then, each component α_i, $i \in \{1, \ldots, n\}$, can be calculated as

$$\alpha_i = \max\left\{0, -\frac{1}{2}\left(\underline{h}_{ii} - \sum_{j \neq i} |h|_{ij}\right)\right\}, \tag{6.10}$$

where $|h|_{ij} = \max\{|\underline{h}_{ij}|, |\overline{h}_{ij}|\}$, and $\underline{h}_{ij}, \overline{h}_{ij}$ are the lower and upper bounds of h_{ij} in $\boldsymbol{H}$, respectively. Correspondingly, a concave over-estimator can be constructed by taking the negative of the αBB convex under-estimator of $-h(\boldsymbol{z})$.

Using αBB relaxation, a convex inclusion function F^R that satisfies Assumption 6.1 can be constructed as follows.

Definition 6.6. *Define an αBB relaxation $F^\alpha = [\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}] : I \times \mathbb{IR}^{n_y} \to \mathbb{IR}^{n_y}$ such that, for each $i \in \{1, \ldots, n_x\}$,*

$$f_i^{cv}(t, \boldsymbol{y}) = f_i(t, \boldsymbol{y}) + \sum_{j=1}^{n_y} a_{ij}^{cv}(t)(y_j^L - y_j)(y_j^U - y_j), \tag{6.11a}$$

$$f_i^{cc}(t, \boldsymbol{y}) = f_i(t, \boldsymbol{y}) - \sum_{j=1}^{n_y} a_{ij}^{cc}(t)(y_j^L - y_j)(y_j^U - y_j), \tag{6.11b}$$

where the ith rows of matrices $\boldsymbol{A}^{cv}(t)$ and $\boldsymbol{A}^{cc}(t)$ are $\boldsymbol{\alpha}$ factors for $f_i(t, \cdot)$ and $-f_i(t, \cdot)$, respectively, obtained as in [3].

The αBB parameters in $\boldsymbol{A}^{cv}(t)$ and $\boldsymbol{A}^{cc}(t)$ can be calculated via (6.10) at each

$t \in I$ with $\boldsymbol{y}^L = (\boldsymbol{\xi}^L(t), \boldsymbol{u}^L)$ and $\boldsymbol{y}^U = (\boldsymbol{\xi}^U(t), \boldsymbol{u}^U)$. Alternatively, if constant bounds of $\boldsymbol{x}$ are available on $I \times U$, then these can be used to determine another valid combination of $\boldsymbol{y}^L$ and $\boldsymbol{y}^U$. Such bounds may be a rough enclosure of the reachable set, or may be computed by an established state bounding method, such as by [59].

Since the original RHS function $\boldsymbol{f}$ is twice differentiable, it is readily verified that F^α in Definition 6.6 is a valid choice of F^R that satisfies Assumption 6.1, and may be employed in the state bounding system (6.4).

6.6.2 Specialization to quadratic functions

If the original RHS function $\boldsymbol{f}$ in (6.1) is quadratic, then its αBB relaxations $\boldsymbol{f}^{cv}$ and $\boldsymbol{f}^{cc}$ are also quadratic. For an arbitrary $i \in \{1, \ldots, n_x\}$, suppose that

$$f_i(t, \boldsymbol{y}) = \boldsymbol{y}^\top \boldsymbol{Q} \boldsymbol{y} + \boldsymbol{q}^\top \boldsymbol{y} + c,$$

where $\boldsymbol{Q}$ is symmetric.

Definition 6.7. *For matrices (or vectors) $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$, their* Hadamard product $\boldsymbol{A} \odot \boldsymbol{B} \in \mathbb{R}^{m \times n}$ *is a matrix with elements*

$$(\boldsymbol{A} \odot \boldsymbol{B})_{ij} = a_{ij} b_{ij}.$$

Let $\boldsymbol{a}_{(i)}^{cv}$ be the transposed ith row of $\boldsymbol{A}^{cv}$. Then, (6.11a) provides

$$\begin{aligned} f_i^{cv}(t,\boldsymbol{y}) &= \boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y} + \boldsymbol{q}^\top \boldsymbol{y} + c + \sum_{j=1}^{n_y} a_{ij}^{cv}(y_j^L - y_j)(y_j^U - y_j) \\ &= \boldsymbol{y}^\top \tilde{\boldsymbol{Q}}\boldsymbol{y} + \tilde{\boldsymbol{q}}^\top \boldsymbol{y} + \tilde{c}, \end{aligned}$$

where, with $\mathrm{diag}(\boldsymbol{a}_{(i)}^{cv})$ denoting the diagonal matrix with components of $\boldsymbol{a}_{(i)}^{cv}$ along its main diagonal,

$$\begin{aligned} \tilde{\boldsymbol{Q}} &:= \boldsymbol{Q} + \mathrm{diag}(\boldsymbol{a}_{(i)}^{cv}), \\ \tilde{\boldsymbol{q}} &:= \boldsymbol{q} - \boldsymbol{a}_{(i)}^{cv} \odot (\boldsymbol{y}^L + \boldsymbol{y}^U), \\ \tilde{c} &:= c + \sum_{j=1}^{n_y} a_{ij}^{cv} y_j^L y_j^U. \end{aligned}$$

Then, the optimization problem in (6.3a) can be expressed as a convex quadratic program (QP); with $\Xi = [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U]$, $[\boldsymbol{\phi}_{(i)}^L, \boldsymbol{\phi}_{(i)}^U] = B_i^L([(\boldsymbol{\xi}^L, \boldsymbol{u}^L), (\boldsymbol{\xi}^U, \boldsymbol{u}^U)])$,

$$\begin{aligned} \min_{\boldsymbol{y}} \quad & \boldsymbol{y}^\top \tilde{\boldsymbol{Q}}\boldsymbol{y} + \tilde{\boldsymbol{q}}^\top \boldsymbol{y} + \tilde{c}, \\ \text{s.t.} \quad & \boldsymbol{\phi}_{(i)}^L \leq \boldsymbol{y} \leq \boldsymbol{\phi}_{(i)}^U. \end{aligned} \tag{6.12}$$

The KKT conditions of (6.12) can be derived accordingly:

$$\begin{aligned} 2\,\tilde{\boldsymbol{Q}}\boldsymbol{y}^* + \tilde{\boldsymbol{q}} + \overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}} &= \boldsymbol{0}, \\ \boldsymbol{\phi}_{(i)}^L \leq \boldsymbol{y}^* &\leq \boldsymbol{\phi}_{(i)}^U, \\ \overline{\boldsymbol{\mu}} \geq \boldsymbol{0}, \quad \underline{\boldsymbol{\mu}} &\geq \boldsymbol{0}, \\ (\overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}})^\top \boldsymbol{y}^* + \underline{\boldsymbol{\mu}}^\top \boldsymbol{\phi}_{(i)}^L - \overline{\boldsymbol{\mu}}^\top \boldsymbol{\phi}_{(i)}^U &= 0. \end{aligned} \tag{6.13}$$

A vector $\boldsymbol{y}^*$ solves (6.12) if and only if there are multipliers $(\overline{\boldsymbol{\mu}}, \underline{\boldsymbol{\mu}})$ for which $(\boldsymbol{y}^*, \overline{\boldsymbol{\mu}}, \underline{\boldsymbol{\mu}})$ solves (6.13).

Note that the QPs described in (6.12) and (6.13) are solvable by efficient commercial solvers such as CPLEX and Gurobi. They may also be treated as multi-parametric quadratic programs [103], in which the optimum of the optimization problem is considered as a function of varying parameters. The advantage of this strategy is that an analytical expression of the optimum function can in principle be obtained in advance, for quick online evaluation.

Moreover, the KKT conditions in (6.13) also comprise a mixed linear complementarity problem (MLCP). Comprehensive theoretical results and various numerical algorithms for LCPs and MLCPs can be found in literature; see e.g. [45].

6.7 Numerical Examples

This section presents numerical examples in which state bounds are constructed for nonlinear dynamic system with our new method described in Sections 6.4 and 6.6. This method was implemented in Julia v1.4.2 with the auxiliary system of

ODEs solved with DifferentialEquations.jl. All numerical experiments were performed on a Windows 10 machine with an AMD Ryzen 2600X CPU and 16GB memory.

The first example involves a simple ODE system with a quadratic RHS.

Example 6.1. *Consider the quadratic ODEs:*

$$\dot{x}_1(t,\boldsymbol{u}) = (x_1-u_1)^2 - (x_2-u_1)^2,\ x_1(t_0) = 2.2,$$
$$\dot{x}_2(t,\boldsymbol{u}) = (x_1-u_2)^2 - (x_2-u_2)^2,\ x_2(t_0) = 1.8,$$

where $U \equiv [-2,2]\times[-1,3]$, $\boldsymbol{u} = (u_1,u_2) \in \tilde{U}$, *and* $I \equiv [t_0,t_f] = [0.0,0.2]$.

Using the approach from Section 6.6.2, we derived quadratic αBB relaxations of $\boldsymbol{f}$, and the QPs (6.12) in (6.4) were solved with CPLEX v12.10. The resulting bounds are illustrated in Figure 6.1, along with Harrison's NIE-based method and trajectories of the original system. This figure shows that the time-varying bounds generated by our new method are tighter than those by Harrison's method.

Next, we consider the Van der Pol oscillator, which is a classic dynamic system that has been widely studied in electrical engineering and biological science. Relaxations of this system were obtained by [126]. Here, we consider its two-dimensional form with uncertainty in both initial conditions and RHS functions.

Example 6.2. *Consider the Van der Pol oscillator:*

$$\dot{x}_1(t,\boldsymbol{u}) = x_1, \qquad x_1(t_0,\boldsymbol{u}) = u_1(t_0),$$
$$\dot{x}_2(t,\boldsymbol{u}) = u_1(1-x_1^2)x_2 - x_1, \quad x_2(t_0,\boldsymbol{u}) = u_2(t_0),$$

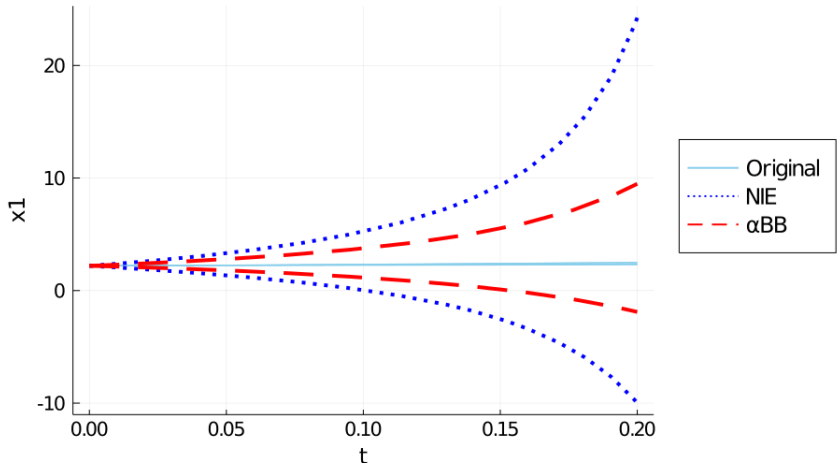


Figure 6.1: State bounds of x_1 in Example 6.1 computed by relaxing RHS functions with NIE (dotted) and αBB relaxation in (6.4) (dashed). Solid (overlapping) lines are real trajectories.

where $U \equiv [1.399, 1.400] \times [2.299, 2.300]$, $\boldsymbol{u} = (u_1, u_2) \in \tilde{U}$, *and* $I \equiv [t_0, t_f] = [0, 6]$.

The αBB relaxations of this ODE's RHS functions were obtained via (6.11) and optimized by IPOPT [147]. State bounds were computed for the state variable x_1 using Harrison's method (NIE) and our new αBB-based method, and are plotted in Figure 6.2. In this case, the new method generates a better enclosure while Harrison's method explodes faster.

The last example involves a bioreactor process [16]. An enclosure of this system was obtained by [83].

Example 6.3. *Consider a microbial growth process described by the following ODE system:*

$$\begin{aligned} \dot{X} &= (\mu - \alpha D)X, & X(t_0) &= 0.82, \\ \dot{S} &= D(S^i - S) - k\mu X, & S(t_0) &= 0.8, \end{aligned}$$

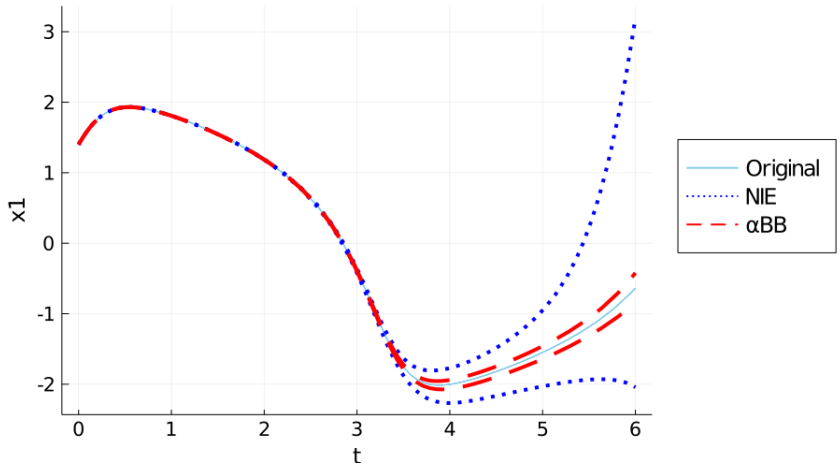


Figure 6.2: State bounds of x_1 in Example 6.2 computed by relaxing RHS functions with NIE (dotted) and αBB relaxation in (6.4) (dashed). Solid (overlapping) lines are real trajectories.

where state variables X and S respectively represent the concentrations of biomass and substrate, $I \equiv [t_0, t_f] = [0, 3]$, and μ is the growth rate

$$\mu = \frac{\mu_m S}{K_S + S + K_I S^2}.$$

The remaining quantities are parameters, whose values and uncertainties are provided in Table 6.1.

Table 6.1: Microbial growth process parameters

Parameter	Symbol	Value	Unit
Process heterogeneity	α	0.5	-
Dilution rate	D	0.36	day^{-1}
Influent concentration	S^i	5.7	g S/l
Yield coefficient	k	10.53	g S/g X
Max growth rate	μ_m	1.2	day^{-1}
Kinetic parameter	K_S	[7.0, 7.2]	g S/l
Kinetic parameter	K_I	[0.4, 0.6]	$(\text{g S/l})^{-1}$

In this numerical experiment, we consider the two kinetic parameters K_S and K_I, to have bounded uncertainties. Corresponding state bounds were constructed with αBB relaxations in (6.4), and are shown in Figure 6.3. This figure shows that the proposed new approach produces a tighter enclosure for the biomass concentration than Harrison's method.

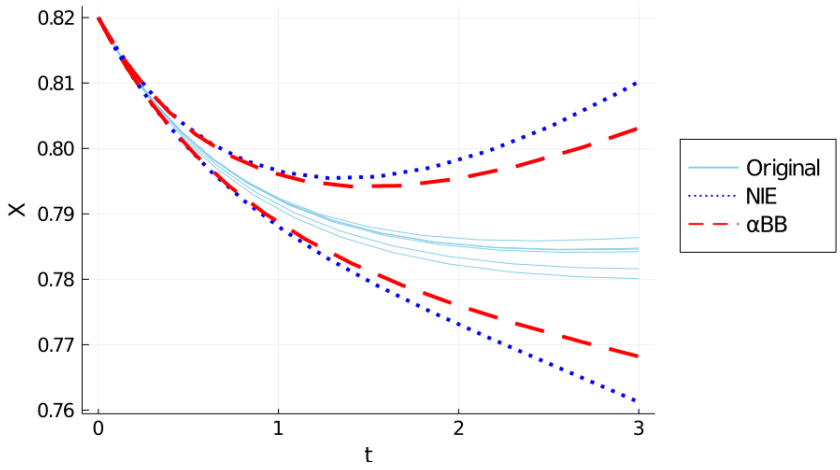


Figure 6.3: State bounds of X in Example 6.3 computed by relaxing RHS functions with NIE (dotted) and αBB relaxation in (6.4) (dashed). Solid lines are real trajectories.

6.8 Conclusion

We have developed an approach for computing tight enclosures for nonlinear control systems based on differential inequalities. Bounding information for the original RHS function $\boldsymbol{f}$ is obtained by optimizing its convex relaxations. We investigated the usage of αBB relaxation in this context, and developed the corresponding complementarity reformulation as an NCS. Our numerical results illustrate the

tightness of the time-varying interval bounds generated by our new method. Future work may involve exploring the usage of other established convex relaxation techniques [123, 72, 73]. Our proof-of-concept implementation involves repeatedly solving optimization problems during integration, which requires a considerable amount of computing effort, especially when the system of interest is nonlinear. As suggested in Section 6.5, a specialized complementarity system solver would help in a more sophisticated implementation.

Chapter 7

A Differential Inequality-Based Framework for Computing Convex Enclosures of Reachable Sets

This chapter represents a manuscript in preparation for submission to a journal.

7.1 Introduction

A reachable set is the set of final states reachable by a dynamic system with uncertain inputs or initial conditions. Enclosing the reachable set provides a quantification for the influence of uncertainty on the dynamic model. This is desired for solving various engineering problems, e.g., state estimation [69], parameter estimation [106], process control [4], fault diagnosis [82, 105], and safety verification [66], in many applications, including chemical reactors [145], biochemical processes [109], and automated vehicles [6]. Another type of problems that requires enclosures of

reachable sets is deterministic global optimization for dynamic systems, including parameter estimation problems [129] and optimal control problems [81, 64]. In this particular usage, reachable set enclosures provide global bounding information for the dynamic model, and they are expected to be convex. Nonconvex enclosures may cause the global optimization algorithm to terminate at a solution that is suboptimal.

Established methods for computing enclosures for the reachable set of general nonlinear dynamic systems can be categorized depending on whether they linearly approximate the nonlinear system or not. Compared with enclosing nonlinear systems, analyzing the reachability of linear systems is significantly easier [108]. Conservative linear and piecewise-linear approximating methods for nonlinear systems are introduced in [8, 7] and [11], respectively. In both techniques, the linearization error is accounted by adding a bounded input to the approximating system. Then, various set representations suitable for linear systems can be applied, such as hyper-rectangles [46], polytopes [39], zonotopes [78, 155], and ellipsoids [79]. Moreover, off-the-shelf packages are available for computing such enclosures, including CORA [5], PHAVer [56], and HSolver [110]. It is worth noting that tight enclosures of reachable sets may require complex set representations and high computational cost [126].

Without linearization, we may generate enclosures for the reachable set of nonlinear systems directly with the following two types of methods. Taylor series methods involve constructing Taylor expansions for the nonlinear dynamic system and then bound them with different techniques, including interval arithmetic

[96], McCormick relaxations [111], and Taylor models [84, 112]. Flow* [38], a reachability analysis tool for polynomial continuous systems, was developed based on this type of methods. Nevertheless, constructing high-order Taylor expansions for high dimensional problems is computationally expensive [120]. The second type of methods relies on the theory of differential inequalities [148]. It constructs an auxiliary system of ODEs to describe time-varying interval bounds for each state of the original system. The right-hand side (RHS) of the auxiliary system are generated from the original RHS functions with various relaxation techniques, e.g., interval arithmetic [59, 118], affine relaxations [128, 61], embedded linear programs [62], and embedded general nonlinear programs [31]. Additional bounding information about the dynamic system, such as physical bounds and conservation laws, can be used to further refine these bounds [121, 62, 127]. Furthermore, Scott and Barton [120] proposed a differential inequality-based framework to generate componentwise convex and concave relaxations for parametric ODEs, given known interval bounds of the original system. They proposed to construct the auxiliary RHS with generalized McCormick relaxations (GMC) [123] of the original RHS function. Recently, Song and Khan [131] developed a new use case of this framework by replacing GMC with embedded optimization problems, whose objective functions can be any convex and concave relaxations of the original RHS functions. Although repeatedly solving optimization problems during numerical integration may not be efficient, their approach generates tighter convex enclosures for nonlinear ODEs than Scott and Barton's GMC-based method. However, Scott and Barton's framework depends on auxiliary ODEs with discontinuous RHS functions, which may lead to difficulties in solving it numerically and evaluating sensitivities

for the generated convex relaxations.

In this work, we propose a new differential inequality-based framework for computing convex enclosures for the reachable set of nonlinear parametric ODEs. It constructs an auxiliary system of ODEs to compute componentwise interval bounds and convex relaxations for the original system simultaneously. Compared with Scott and Barton's framework, our new framework eliminates the discrete jumps by taking into account the dynamics of both interval bounds and convex relaxations. It also supports Song and Khan's optimization-based method for constructing auxiliary RHS function. Moreover, with a proper choice of auxiliary RHS functions, we can generate convex enclosures that are tighter than Scott and Barton's method and Song and Khan's method. Last but not least, smooth convex relaxations that are differentiable with respect to parameters, along with their gradients, can be computed under mild assumptions. This is critical for the application of convex enclosures in global optimization algorithms [131].

This article is organized as follows. Section 7.2 introduces some notations and definitions that are used throughout the article. Problem formulation, along with the relative background, is provided in Section 7.3. Our new framework is presented in Section 7.4. Sections 7.5 and 7.6 introduce various novel methods discovered with this framework for constructing auxiliary RHS functions. Lastly, numerical examples are presented in Section 7.8 to demonstrate the reachable set enclosures generated with these novel methods.

7.2 Preliminaries

This section introduces the mathematical background underlying the methods and results in this article. The following notation conventions are used. The set of positive real numbers is represented as $\mathbb{R}_{>0}$, and $\mathbb{R}_{\geq 0}$ stands for the set of non-negative real numbers. The standard Euclidean norm $\|\cdot\|$ is adopted for any vector space $\mathbb{R}^n$, and $\|\cdot\|_\infty$ represents infinity norm. Vectors are denoted with boldface lower-case letters (e.g. $\boldsymbol{z}$). Given vectors $\boldsymbol{z}^\dagger, \boldsymbol{z}^\ddagger \in \mathbb{R}^n$, inequalities such as $\boldsymbol{z}^\dagger < \boldsymbol{z}^\ddagger$ or $\boldsymbol{z}^\dagger \leq \boldsymbol{z}^\ddagger$ are to be interpreted component-wise. $\boldsymbol{z}_{(-i)} \in \mathbb{R}^{n-1}$ stands for the vector $\boldsymbol{z}$ with the ith component excluded. Throughout this article, the convexity of a vector-valued function $\boldsymbol{h}$ refers to convexity of all components h_i. Dotted quantities indicate time-derivatives (e.g. $\dot{\boldsymbol{z}} \equiv \frac{\partial \boldsymbol{z}}{\partial t}$). The abbreviation "a.e." stands for "almost every" in the Lebesgue sense.

Definition 7.1. *For any $\boldsymbol{z}^L, \boldsymbol{z}^U \in \mathbb{R}^n$ such that $\boldsymbol{z}^L \leq \boldsymbol{z}^U$, define the interval $Z = [\boldsymbol{z}^L, \boldsymbol{z}^U]$ as the nonempty compact connected set of $\{\boldsymbol{z} \in \mathbb{R}^n : \boldsymbol{z}^L \leq \boldsymbol{z} \leq \boldsymbol{z}^U\}$. The set of all interval subsets of $D \subset \mathbb{R}^n$ is denoted as $\mathbb{I}D$, and $\mathbb{IR}^n$ denotes the set of all interval subsets of $\mathbb{R}^n$.*

Definition 7.2. *Let $S \in \mathbb{IR}^n$ and $\boldsymbol{h} : S \to \mathbb{R}^m$.*

1. *An interval function $H \equiv [\boldsymbol{h}^L, \boldsymbol{h}^U] : \mathbb{IR}^n \to \mathbb{IR}^m$ is an* inclusion function *of $\boldsymbol{h}$ on S if for all $Z \subseteq S$,*

$$\{\boldsymbol{h}(\boldsymbol{z}) : \boldsymbol{z} \in Z\} \subseteq H(Z) \equiv [\boldsymbol{h}^L(Z), \boldsymbol{h}^U(Z)].$$

2. *Let* $H^\dagger, H^\ddagger : \mathbb{IR}^n \to \mathbb{IR}^m$ *be interval functions.* $H^\dagger$ *is* tighter *than* $H^\ddagger$ *on* S *if*

$$H^\dagger(Z) \subseteq H^\ddagger(Z), \quad \forall Z \subseteq S.$$

3. H *is* inclusion monotonic *on* S *if for all* $Z^\dagger, Z^\ddagger \in S$ *such that* $Z^\dagger \subseteq Z^\ddagger$,

$$H(Z^\dagger) \subseteq H(Z^\ddagger).$$

Definition 7.3. *Let* $Z \in \mathbb{IR}^n$ *and* $\boldsymbol{h} : Z \to \mathbb{R}^m$.

1. $\boldsymbol{h}^{cv} : Z \to \mathbb{R}^m$ *is a* convex relaxation *of* $\boldsymbol{h}$ *on* Z *if* $\boldsymbol{h}^{cv}(\boldsymbol{z}) \leq \boldsymbol{h}(\boldsymbol{z})$ *for all* $\boldsymbol{z} \in Z$ *and* $\boldsymbol{h}^{cv}$ *is convex on* Z.

2. $\boldsymbol{h}^{cc} : Z \to \mathbb{R}^m$ *is a* concave relaxation *of* $\boldsymbol{h}$ *on* Z *if* $\boldsymbol{h}^{cc}(\boldsymbol{z}) \geq \boldsymbol{h}(\boldsymbol{z})$ *for all* $\boldsymbol{z} \in Z$ *and* $\boldsymbol{h}^{cc}$ *is concave on* Z.

3. *The interval function* $H \equiv [\boldsymbol{h}^{cv}, \boldsymbol{h}^{cc}]$ *is a* convex inclusion function *of* $\boldsymbol{h}$ *on* Z.

Definition 7.4 (Adapted from [120])**.** *For each* $i \in \{1, \ldots, n\}$, *define flattening operators* $B_i^L, B_i^U : \mathbb{IR}^n \to \mathbb{IR}^n$ *such that*

1. $B_i^L([\boldsymbol{\phi}, \boldsymbol{\psi}]) = [\boldsymbol{\phi}, \boldsymbol{\psi}']$, *where* $\psi_i' = \phi_i$ *and* $\boldsymbol{\psi}'_{(-i)} = \boldsymbol{\psi}_{(-i)}$.

2. $B_i^U([\boldsymbol{\phi}, \boldsymbol{\psi}]) = [\boldsymbol{\phi}', \boldsymbol{\psi}]$, *where* $\phi_i' = \psi_i$ *and* $\boldsymbol{\phi}'_{(-i)} = \boldsymbol{\phi}_{(-i)}$.

7.3 Problem Formulation

Consider $t_0, t_f \in \mathbb{R}$ with $t_0 < t_f$ and define $I := [t_0, t_f]$. Let $P \subset \mathbb{R}^{n_p}$ be an interval, and $D \subset \mathbb{R}^{n_x}$ be open. Given a continuous mapping $\boldsymbol{x}_0 : P \to D$ and a Lipschitz

continuous function $\boldsymbol{f} : I \times P \times D \to \mathbb{R}^{n_x}$, consider an initial-value problem

$$\dot{\boldsymbol{x}}(t, \boldsymbol{p}) = \boldsymbol{f}(t, \boldsymbol{p}, \boldsymbol{x}(t, \boldsymbol{p})), \qquad \boldsymbol{x}(t_0, \boldsymbol{p}) = \boldsymbol{x}_0(\boldsymbol{p}). \tag{7.1}$$

Suppose that $k^{\boldsymbol{x}} \in \mathbb{R}_{>0}$ is a Lipschitz constant of $\boldsymbol{f}(t, \boldsymbol{p}, \cdot)$ over D for all $(t, \boldsymbol{p}) \in I \times P$. Then, (7.1) is guaranteed to have a unique solution by the Picard-Lindelöf Theorem as summarized in [60, Theorem 1.1, Chapter II].

State bounds and *State relaxations* [120], defined below, provide valid enclosures for the reachable set of (7.1). They are both time-varying lower and upper bounds for the state variables. While state bounds are independent of parameters, state relaxations are convex with respect to parameters componentwise. Moreover, state relaxations are typically tighter state bounds, so they provide tighter convex enclosures for the reachable set of (7.1) than state bounds. But the computation of state relaxations usually requires state bounds [120].

Definition 7.5 (State bound, adapted from [120])**.** *Functions $\boldsymbol{x}^L, \boldsymbol{x}^U : I \to \mathbb{R}^{n_x}$ are* state bounds *of* (7.1) *if, for each $t \in I$ and $\boldsymbol{p} \in P$,*

$$\boldsymbol{x}^L(t) \leq \boldsymbol{x}(t, \boldsymbol{p}) \leq \boldsymbol{x}^U(t).$$

Let $X^B \equiv [\boldsymbol{x}^L, \boldsymbol{x}^U] : I \to \mathbb{IR}^{n_x}$ denote the corresponding inclusion function of $\boldsymbol{x}$ on $I \times P$.

Definition 7.6 (State relaxation, adapted from [120])**.** *Functions $\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc} : I \times P \to \mathbb{R}^{n_x}$ are* state relaxations *of* (7.1) *on $I \times P$, if, for every $t \in I$,*

1. *$\boldsymbol{x}^{cv}(t, \cdot)$ is convex on P,*

2. $\boldsymbol{x}^{cc}(t,\cdot)$ *is concave on* P*, and*

3. $\boldsymbol{x}^{cv}(t,\boldsymbol{p}) \leq \boldsymbol{x}(t,\boldsymbol{p}) \leq \boldsymbol{x}^{cc}(t,\boldsymbol{p})$ *for all* $\boldsymbol{p} \in P$*.*

Let $X^R \equiv [\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc}] : I \times P \to \mathbb{R}^{n_x}$ *denote the corresponding inclusion function of* $\boldsymbol{x}$ *on* $I \times P$*.*

The main objective of this work is to formulate a general framework for constructing tight and smooth state relaxations for the nonlinear parametric ODE (7.1). We achieve this by modifying Scott and Barton's framework [120] and combining it with our previous work on constructing state bounds [28]. A unified auxiliary system of ODEs is proposed to compute state bounds and state relaxations simultaneously.

7.3.1 Background

This section introduces some properties of the RHS function in an auxiliary ODE system that describes enclosures of (7.1).

Definition 7.7 (Bound-preserving dynamics, adapted from [120])**.** *Functions* $\boldsymbol{u}, \boldsymbol{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ *describe* bound-preserving dynamics *for* (7.1) *if, for any* $\boldsymbol{p} \in P$*, each* $i \in \{1, \ldots, n_x\}$*, a.e.* $t \in I$*, and any* $\boldsymbol{z} \in \mathbb{R}^{n_x}$ *and* $Z \equiv [\boldsymbol{z}^L, \boldsymbol{z}^U] \in \mathbb{IR}^{n_x}$ *such that* $\boldsymbol{z}^L \leq \boldsymbol{z} \leq \boldsymbol{z}^U$*,* $\boldsymbol{u}$ *and* $\boldsymbol{o}$ *satisfy:*

1. *If* $z_i = z_i^L$*, then* $u_i(t,\boldsymbol{p},\boldsymbol{z}^L,\boldsymbol{z}^U) \leq f_i(t,\boldsymbol{p},\boldsymbol{z})$*,*

2. *If* $z_i = z_i^U$*, then* $o_i(t,\boldsymbol{p},\boldsymbol{z}^L,\boldsymbol{z}^U) \geq f_i(t,\boldsymbol{p},\boldsymbol{z})$*.*

Definition 7.8 (Enclosing dynamics, adapted from [28])**.** *Consider arbitrary continuously differentiable functions* $\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger : I \times P \to \mathbb{R}^n$ *such that* $\boldsymbol{\xi}^\dagger(t,\boldsymbol{p}) \leq \boldsymbol{\xi}^\ddagger(t,\boldsymbol{p})$ *for all*

$(t, \boldsymbol{p}) \in I \times P$. *Functions* $\boldsymbol{u}, \boldsymbol{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ *describe* enclosing dynamics *about* $[\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger]$ *if the following holds. For a.e.* $t \in I$, *any* $\boldsymbol{p} \in P$, $Z \equiv [\boldsymbol{z}^L, \boldsymbol{z}^U] \in \mathbb{IR}^{n_x}$ *such that* $\boldsymbol{z}^L \leq \boldsymbol{\xi}^\dagger(t, \boldsymbol{p}) \leq \boldsymbol{\xi}^\ddagger(t, \boldsymbol{p}) \leq \boldsymbol{z}^U$,

1. *If* $z_i^L = \xi_i^\dagger(t, \boldsymbol{p})$, *then* $u_i(t, \boldsymbol{p}, \boldsymbol{z}^L, \boldsymbol{z}^U) \leq \dot{\xi}_i^\dagger(t, \boldsymbol{p})$,
2. *If* $z_i^U = \xi_i^\ddagger(t, \boldsymbol{p})$, *then* $o_i(t, \boldsymbol{p}, \boldsymbol{z}^L, \boldsymbol{z}^U) \geq \dot{\xi}_i^\ddagger(t, \boldsymbol{p})$.

$\boldsymbol{u}, \boldsymbol{o}$ *describe enclosing dynamics about a single trajectory* $\boldsymbol{\xi} : I \times P \to \mathbb{R}^n$ *if we let* $\boldsymbol{\xi}^\dagger \equiv \boldsymbol{\xi}^\ddagger \equiv \boldsymbol{\xi}$.

Observe that if $\boldsymbol{u}, \boldsymbol{o}$ describe enclosing dynamics about $\boldsymbol{x}$, then they also describe bound-preserving dynamics about (7.1). However, this claim does not hold vice versa. Thus, the requirements in enclosing dynamics are weaker than those in bound-preserving dynamics.

Proposition 7.1 (Adapted from [28])**.** *Consider arbitrary continuously differentiable functions* $\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger : I \to \mathbb{R}^n$ *such that* $\boldsymbol{\xi}^\dagger(t) \leq \boldsymbol{\xi}^\ddagger(t)$ *for all* $t \in I$. *Define* $\boldsymbol{\xi}_0^L, \boldsymbol{\xi}_0^U \in \mathbb{R}^n$ *and continuous functions* $\boldsymbol{h}^L, \boldsymbol{h}^U : I \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$. *Let* $(\boldsymbol{\xi}^L, \boldsymbol{\xi}^U)$ *solve the following ODE:*

$$\begin{aligned} \dot{\boldsymbol{\xi}}^L(t) &= \boldsymbol{h}^L(t, \boldsymbol{\xi}^L(t), \boldsymbol{\xi}^U(t)), \quad \boldsymbol{\xi}^L(t_0) = \boldsymbol{\xi}_0^L, \\ \dot{\boldsymbol{\xi}}^U(t) &= \boldsymbol{h}^U(t, \boldsymbol{\xi}^L(t), \boldsymbol{\xi}^U(t)), \quad \boldsymbol{\xi}^U(t_0) = \boldsymbol{\xi}_0^U. \end{aligned} \tag{7.2}$$

If the following holds:

1. *There exists* $k \in \mathbb{R}_{>0}$ *such that, for any* $i \in \{1, \ldots, n\}$, *a.e.* $t \in I$, *and any*

$\boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger \in \mathbb{R}^n$ *for which* $\boldsymbol{\phi}^\ddagger \leq \boldsymbol{\phi}^\dagger \leq \boldsymbol{\psi}^\dagger \leq \boldsymbol{\psi}^\ddagger$,

$$\begin{aligned} &h_i^L(t, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) - h_i^L(t, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) \\ &\quad \leq k(\|\boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger\|_\infty + \|\boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger\|_\infty), \\ &h_i^U(t, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger) - h_i^U(t, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) \\ &\quad \leq k(\|\boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^\ddagger\|_\infty + \|\boldsymbol{\psi}^\dagger - \boldsymbol{\psi}^\ddagger\|_\infty). \end{aligned}$$

2. $\boldsymbol{h}^L, \boldsymbol{h}^U$ *describe enclosing dynamics about* $[\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger]$,

3. $\boldsymbol{\xi}_0^L \leq \boldsymbol{\xi}^\dagger(t_0)$ *and* $\boldsymbol{\xi}^\ddagger(t_0) \leq \boldsymbol{\xi}_0^U$,

then

$$\boldsymbol{\xi}^L(t) \leq \boldsymbol{\xi}^\dagger(t) \leq \boldsymbol{\xi}^\ddagger(t) \leq \boldsymbol{\xi}^U(t), \quad \forall t \in I.$$

Note that Proposition 7.1 can be extended to enclose a single trajectory $\boldsymbol{\xi} : I \to \mathbb{R}^n$ if we let $\boldsymbol{\xi}^\dagger \equiv \boldsymbol{\xi}^\ddagger \equiv \boldsymbol{\xi}$.

Definition 7.9 (Convexity-amplifying dynamics, adapted from [120])**.** *Let* $S \in \mathbb{IR}^{n_x}$. *Functions* $\boldsymbol{u}, \boldsymbol{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ *describe* convexity-amplifying dynamics *on* S *if, for any* $(\lambda, \boldsymbol{p}^\dagger, \boldsymbol{p}^\ddagger) \in (0,1) \times P \times P$, *each* $i \in \{1, \ldots, n_x\}$, *a.e.* $t \in I$, *and* $\boldsymbol{\phi}^\dagger, \boldsymbol{\phi}^\ddagger, \bar{\boldsymbol{\phi}}, \boldsymbol{\psi}^\dagger, \boldsymbol{\psi}^\ddagger, \bar{\boldsymbol{\psi}} \in S$ *such that the following three conditions all hold:*

1. $\bar{\boldsymbol{\phi}} \leq \lambda \boldsymbol{\phi}^\dagger + (1-\lambda)\boldsymbol{\phi}^\ddagger$,

2. $\bar{\boldsymbol{\psi}} \geq \lambda \boldsymbol{\psi}^\dagger + (1-\lambda)\boldsymbol{\psi}^\ddagger$, *and*

3. $\boldsymbol{\phi}^\dagger \leq \boldsymbol{\psi}^\dagger$, $\boldsymbol{\phi}^\ddagger \leq \boldsymbol{\psi}^\ddagger$, $\bar{\boldsymbol{\phi}} \leq \bar{\boldsymbol{\psi}}$,

$\boldsymbol{u}$ and $\boldsymbol{o}$ satisfy:

$$u_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \leq \lambda u_i(t, \boldsymbol{p}^\dagger, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) + (1-\lambda) u_i(t, \boldsymbol{p}^\ddagger, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger),$$
$$o_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \geq \lambda o_i(t, \boldsymbol{p}^\dagger, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) + (1-\lambda) o_i(t, \boldsymbol{p}^\ddagger, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger),$$

where $\bar{\boldsymbol{p}} \equiv \lambda \boldsymbol{p}^\dagger + (1-\lambda)\boldsymbol{p}^\ddagger$.

Definition 7.10 (Convexity-preserving dynamics, adapted from [120])**.** *Let $S \in \mathbb{IR}^{n_x}$. Functions $\boldsymbol{u}, \boldsymbol{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ describe* convexity-preserving dynamics *on S if, for any $(\lambda, \boldsymbol{p}^\dagger, \boldsymbol{p}^\ddagger) \in (0,1) \times P \times P$, each $i \in \{1, \ldots, n_x\}$, a.e. $t \in I$, and $\boldsymbol{\phi}^\dagger, \boldsymbol{\phi}^\ddagger, \bar{\boldsymbol{\phi}}, \boldsymbol{\psi}^\dagger, \boldsymbol{\psi}^\ddagger, \bar{\boldsymbol{\psi}} \in S$ such that the following three conditions all hold:*

1. $\bar{\boldsymbol{\phi}} \leq \lambda \boldsymbol{\phi}^\dagger + (1-\lambda)\boldsymbol{\phi}^\ddagger$,
2. $\bar{\boldsymbol{\psi}} \geq \lambda \boldsymbol{\psi}^\dagger + (1-\lambda)\boldsymbol{\psi}^\ddagger$, *and*
3. $\boldsymbol{\phi}^\dagger \leq \boldsymbol{\psi}^\dagger$, $\boldsymbol{\phi}^\ddagger \leq \boldsymbol{\psi}^\ddagger$, $\bar{\boldsymbol{\phi}} \leq \bar{\boldsymbol{\psi}}$,

$\boldsymbol{u}$ and $\boldsymbol{o}$ satisfy:

1. *If $\bar{\phi}_i = \lambda \phi_i^\dagger + (1-\lambda)\phi_i^\ddagger$, then*

$$u_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \leq \lambda u_i(t, \boldsymbol{p}^\dagger, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) + (1-\lambda) u_i(t, \boldsymbol{p}^\ddagger, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger),$$

2. *If $\bar{\psi}_i = \lambda \psi_i^\dagger + (1-\lambda)\psi_i^\ddagger$, then*

$$o_i(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}) \geq \lambda o_i(t, \boldsymbol{p}^\dagger, \boldsymbol{\phi}^\dagger, \boldsymbol{\psi}^\dagger) + (1-\lambda) o_i(t, \boldsymbol{p}^\ddagger, \boldsymbol{\phi}^\ddagger, \boldsymbol{\psi}^\ddagger),$$

where $\bar{\boldsymbol{p}} \equiv \lambda \boldsymbol{p}^\dagger + (1-\lambda)\boldsymbol{p}^\ddagger$.

Definition 7.11 (Inclusion-amplifying dynamic)**.** *Consider functions* $\boldsymbol{u}^\dagger, \boldsymbol{o}^\dagger, \boldsymbol{u}^\ddagger, \boldsymbol{o}^\ddagger : I \times P \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$. $\boldsymbol{u}^\ddagger, \boldsymbol{o}^\ddagger$ *describe* inclusion-amplifying dynamic *about* $\boldsymbol{u}^\dagger, \boldsymbol{o}^\dagger$ *if, for a.e.* $t \in I$*, any* $i \in \{1, \ldots, n_x\}$, $\boldsymbol{p} \in P$*, and* $\Xi^\dagger, \Xi^\ddagger \in \mathbb{IR}^{n_x}$ *such that* $\Xi^\dagger \subseteq \Xi^\ddagger$,

$$[u_i^\dagger(t, \boldsymbol{p}, \Xi^\dagger), o_i^\dagger(t, \boldsymbol{p}, \Xi^\dagger)] \subseteq [u_i^\ddagger(t, \boldsymbol{p}, \Xi^\ddagger), o_i^\ddagger(t, \boldsymbol{p}, \Xi^\ddagger)].$$

Definition 7.12 (Inclusion-preserving dynamic)**.** *Consider functions* $\boldsymbol{u}^\dagger, \boldsymbol{o}^\dagger, \boldsymbol{u}^\ddagger, \boldsymbol{o}^\ddagger : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$. $\boldsymbol{u}^\ddagger, \boldsymbol{o}^\ddagger$ *describe* inclusion-preserving dynamic *about* $\boldsymbol{u}^\dagger, \boldsymbol{o}^\dagger$ *if, for a.e.* $t \in I$*, any* $i \in \{1, \ldots, n_x\}$, $\boldsymbol{p} \in P$*, and* $\Xi^\dagger \equiv [\boldsymbol{\xi}^{L\dagger}, \boldsymbol{\xi}^{U\dagger}], \Xi^\ddagger \equiv [\boldsymbol{\xi}^{L\ddagger}, \boldsymbol{\xi}^{U\ddagger}] \in \mathbb{IR}^{n_x}$ *such that* $\Xi^\dagger \subseteq \Xi^\ddagger$,

1. *If* $\xi_i^{L\dagger} = \xi_i^{L\ddagger}$*, then* $u_i^\ddagger(t, \boldsymbol{p}, \Xi^\ddagger) \leq u_i^\dagger(t, \boldsymbol{p}, \Xi^\dagger)$.
2. *If* $\xi_i^{U\dagger} = \xi_i^{U\ddagger}$*, then* $o_i^\ddagger(t, \boldsymbol{p}, \Xi^\ddagger) \geq o_i^\dagger(t, \boldsymbol{p}, \Xi^\dagger)$.

Lemma 7.1. *Consider functions* $\boldsymbol{u}^\dagger, \boldsymbol{o}^\dagger, \bar{\boldsymbol{u}}^\dagger, \bar{\boldsymbol{o}}^\dagger, \boldsymbol{u}^\ddagger, \boldsymbol{o}^\ddagger, \bar{\boldsymbol{u}}^\ddagger, \bar{\boldsymbol{o}}^\ddagger : I \times P \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ *such that, for each* $i \in \{1, \ldots, n\}$*, any* $(t, \boldsymbol{p}) \in I \times P$ *and* $Z \in \mathbb{IR}^{n_x}$,

$$\bar{u}_i^\dagger(t, \boldsymbol{p}, Z) \equiv u_i^\dagger(t, \boldsymbol{p}, B_i^L(Z)),\ \bar{o}_i^\dagger(t, \boldsymbol{p}, Z) \equiv o_i^\dagger(t, \boldsymbol{p}, B_i^U(Z)),$$
$$\bar{u}_i^\ddagger(t, \boldsymbol{p}, Z) \equiv u_i^\ddagger(t, \boldsymbol{p}, B_i^L(Z)),\ \bar{o}_i^\ddagger(t, \boldsymbol{p}, Z) \equiv o_i^\ddagger(t, \boldsymbol{p}, B_i^U(Z)).$$

If $\boldsymbol{u}^\ddagger, \boldsymbol{o}^\ddagger$ *describe* inclusion-amplifying dynamics *about* $\boldsymbol{u}^\dagger, \boldsymbol{o}^\dagger$*, then* $\bar{\boldsymbol{u}}^\ddagger, \bar{\boldsymbol{o}}^\ddagger$ *describe* inclusion-preserving dynamics *about* $\bar{\boldsymbol{u}}^\dagger, \bar{\boldsymbol{o}}^\dagger$.

Proof. Consider a.e. $t \in I$, each $i \in \{1, \ldots, n_x\}$, any $\boldsymbol{p} \in P$ and $\Xi^\dagger \equiv [\boldsymbol{\xi}^{L\dagger}, \boldsymbol{\xi}^{U\dagger}], \Xi^\ddagger \equiv [\boldsymbol{\xi}^{L\ddagger}, \boldsymbol{\xi}^{U\ddagger}] \in \mathbb{IR}^{n_x}$ such that $\Xi^\dagger \subseteq \Xi^\ddagger$. According to Definition 7.12, it suffices to show that,

1. If $\xi_i^{L\dagger} = \xi_i^{L\ddagger}$, then $\bar{u}_i^\ddagger(t, \boldsymbol{p}, \Xi^\ddagger) \leq \bar{u}_i^\dagger(t, \boldsymbol{p}, \Xi^\dagger)$.

2. If $\xi_i^{U\dagger} = \xi_i^{U\ddagger}$, then $\bar{o}_i^{\ddagger}(t, \boldsymbol{p}, \Xi^{\ddagger}) \geq \bar{o}_i^{\dagger}(t, \boldsymbol{p}, \Xi^{\dagger})$.

It will be shown that the first condition holds; showing the second is analogous.

If $\xi_i^{L\dagger} = \xi_i^{L\ddagger}$, then the flattening operation ensures that $B_i^L(\Xi^{\dagger}) \subseteq B_i^L(\Xi^{\ddagger})$. Since $\boldsymbol{u}^{\ddagger}, \boldsymbol{u}^{\ddagger}$ describe inclusion-amplifying dynamics about $\boldsymbol{u}^{\dagger}, \boldsymbol{u}^{\dagger}$,

$$u_i^{\ddagger}(t, \boldsymbol{p}, B_i^L(\Xi^{\ddagger})) \leq u_i^{\dagger}(t, \boldsymbol{p}, B_i^L(\Xi^{\dagger})),$$

which is equivalent to

$$\bar{u}_i^{\ddagger}(t, \boldsymbol{p}, \Xi^{\ddagger}) \leq \bar{u}_i^{\dagger}(t, \boldsymbol{p}, \Xi^{\dagger}).$$

Therefore, the first condition is verified. □

7.3.2 Established methods

This subsection briefly reviews Scott and Barton's framework (Scott-Barton framework hereafter) for generating state relaxations of (7.1).

Assumption 7.1. *Assume that $\boldsymbol{x}_0^L, \boldsymbol{x}_0^U \in \mathbb{R}^{n_x}$ and $\boldsymbol{x}_0^{cv}, \boldsymbol{x}_0^{cc} : P \to \mathbb{R}^{n_x}$ satisfy the following:*

1) $\boldsymbol{x}_0^L \leq \boldsymbol{x}_0(\boldsymbol{p}) \leq \boldsymbol{x}_0^U$ for all $\boldsymbol{p} \in P$,

2) $\boldsymbol{x}_0^{cv}$ and $\boldsymbol{x}_0^{cc}$ are convex and concave relaxations of $\boldsymbol{x}_0$, respectively, on P,

3) $[\boldsymbol{x}_0^{cv}(\boldsymbol{p}), \boldsymbol{x}_0^{cc}(\boldsymbol{p})] \subseteq [\boldsymbol{x}_0^L, \boldsymbol{x}_0^U]$ for all $\boldsymbol{p} \in P$.

Note that the last condition in Assumption 7.1 may be enforced by setting $\boldsymbol{x}_0^{cv}(\boldsymbol{p}) \leftarrow \max\{\boldsymbol{x}_0^L, \boldsymbol{x}_0^{cv}(\boldsymbol{p})\}$ and $\boldsymbol{x}_0^{cc}(\boldsymbol{p}) \leftarrow \min\{\boldsymbol{x}_0^U, \boldsymbol{x}_0^{cc}(\boldsymbol{p})\}$.

Given continuously differentiable state bounds X^B of (7.1) on $I \times P$ and initial conditions $\boldsymbol{x}_0^{cv}, \boldsymbol{x}_0^{cc}$ satisfying Assumption 7.1, the Scott-Barton framework [120] provides the following result. If $\boldsymbol{u}, \boldsymbol{o} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ describe bounding-preserving dynamics about (7.1) and convexity-preserving dynamics, then $(\boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ that solves the following auxiliary system of ODEs, provides state relaxations of (7.1) on $I \times P$: for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned}
\dot{x}_i^{cv}(t,\boldsymbol{p}) &= \begin{cases} u_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) \\ \qquad\qquad \text{if } x_i^{cv}(t,\boldsymbol{p}) > x_i^L(t), \\ \max\left\{\dot{x}_i^L(t), u_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p}))\right\} \\ \qquad\qquad \text{if } x_i^{cv}(t,\boldsymbol{p}) \leq x_i^L(t), \end{cases} \\
x_i^{cv}(t_0,\boldsymbol{p}) &= x_{0,i}^{cv}(\boldsymbol{p}), \\
\dot{x}_i^{cc}(t,\boldsymbol{p}) &= \begin{cases} o_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) \\ \qquad\qquad \text{if } x_i^{cc}(t,\boldsymbol{p}) < x_i^U(t), \\ \min\left\{\dot{x}_i^U(t), o_i(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p}))\right\} \\ \qquad\qquad \text{if } x_i^{cc}(t,\boldsymbol{p}) \geq x_i^U(t), \end{cases} \\
x_i^{cc}(t_0,\boldsymbol{p}) &= x_{0,i}^{cc}(\boldsymbol{p}).
\end{aligned} \tag{7.3}$$

Scott and Barton also proposed that state bounds X^B of (7.1) can be computed with Harrison's method [59]. Functions $\boldsymbol{x}_0^{cv}, \boldsymbol{x}_0^{cc}$ satisfying Assumption 7.1 can be generated from $\boldsymbol{x}_0$ with GMC. Functions $\boldsymbol{u}, \boldsymbol{o}$ that describe bounding-preserving dynamics about (7.1) and convexity-preserving dynamics can be constructed by applying the flattening operation in Definition 7.4 to GMC of $\boldsymbol{f}$.

An optimization-based method was developed by Song and Khan in [131] to

compute tighter state relaxations using the Scott-Barton framework. Consider functions $\hat{\boldsymbol{u}}, \hat{\boldsymbol{o}} : I \times P \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ such that $\hat{\boldsymbol{u}}(t,\cdot,\cdot)$ and $\hat{\boldsymbol{o}}(t,\cdot,\cdot)$ are convex and concave relaxations of $\boldsymbol{f}(t,\cdot,\cdot)$, respectively, on $P \times X^B(t)$ for a.e. $t \in I$. For each $i \in \{1,\ldots,n_x\}$, let

$$\begin{aligned} u_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}) &:= \min_{\substack{\boldsymbol{z}\in[\boldsymbol{\phi},\boldsymbol{\psi}],\\ z_i=\phi_i}} \hat{u}_i(t,\boldsymbol{p},\boldsymbol{z}), \\ o_i(t,\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\psi}) &:= \max_{\substack{\boldsymbol{z}\in[\boldsymbol{\phi},\boldsymbol{\psi}],\\ z_i=\psi_i}} \hat{o}_i(t,\boldsymbol{p},\boldsymbol{z}). \end{aligned} \tag{7.4}$$

Song an Khan validated that $\boldsymbol{u}, \boldsymbol{o}$ in (7.4) also describe bounding-preserving dynamics about (7.1) and convexity-preserving dynamics. Thus, $\boldsymbol{u}, \boldsymbol{o}$ in (7.4) can be substituted into (7.3) to compute state relaxations for (7.1).

The Scott-Barton framework described in (7.3) contains discrete jumps in its RHS because of those if-statements. To solve (7.3) numerically, an advanced ODE solver with event detection feature, e.g., CVODES, is required to handle those discrete jumps. This prohibits the usage of many ODE solvers and increases the difficulty of implementation. Moreover, the discontinuities in ODE RHS create obstacles for evaluating gradients or subgradients of state relaxations, which is another limitation of the Scott-Barton framework. To address these problems, a new differential inequality-based framework is proposed in the next section.

7.4 New Framework for State Relaxation

Unlike the Scott-Barton framework that depends on known state bounds of the original system, our new framework computes state bounds and state relaxations

simultaneously using a coupled auxiliary system of ODEs.

Assumption 7.2. *Assume that functions* $\boldsymbol{f}^L, \boldsymbol{f}^U, \boldsymbol{f}^{cv}, \boldsymbol{f}^{cc} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ *satisfy the following:*

1) $\boldsymbol{f}^L, \boldsymbol{f}^U, \boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ *are continuous,*

2) $\boldsymbol{f}^L(t,\boldsymbol{p},\cdot,\cdot), \boldsymbol{f}^U(t,\boldsymbol{p},\cdot,\cdot)$ *and* $\boldsymbol{f}^{cv}(t,\boldsymbol{p},\cdot,\cdot), \boldsymbol{f}^{cc}(t,\boldsymbol{p},\cdot,\cdot)$ *are Lipschitz continuous on* $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$*, uniformly in* $(t,\boldsymbol{p}) \in I \times P$*,*

3) *Both* $\boldsymbol{f}^L, \boldsymbol{f}^U$ *and* $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ *describe enclosing dynamics about* $\boldsymbol{x}$*, and*

4) $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ *describe convexity-preserving dynamics on* $\mathbb{R}^{n_x}$*.*

Assumption 7.3. *Assume that* $\boldsymbol{f}^L, \boldsymbol{f}^U$ *describe inclusion-preserving dynamics about* $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$*.*

Under Assumptions 7.2 and 7.3, consider the following auxiliary initial-value problem in parametric ODEs: for $i \in \{1,\ldots,n_x\}$,

$$\begin{aligned} \dot{x}_i^L(t) &= \min_{\boldsymbol{p}\in P}\ f_i^L(t,\boldsymbol{p},\boldsymbol{x}^L(t),\boldsymbol{x}^U(t)),\ x_i^L(t_0) = x_{0,i}^L, \\ \dot{x}_i^U(t) &= \max_{\boldsymbol{p}\in P} f_i^U(t,\boldsymbol{p},\boldsymbol{x}^L(t),\boldsymbol{x}^U(t)),\ x_i^U(t_0) = x_{0,i}^U, \end{aligned} \tag{7.5a}$$

$$\begin{aligned} \dot{x}_i^{cv}(t,\boldsymbol{p}) &= f_i^{cv}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})),\ x_i^{cv}(t_0,\boldsymbol{p}) = x_{0,i}^{cv}(\boldsymbol{p}), \\ \dot{x}_i^{cc}(t,\boldsymbol{p}) &= f_i^{cc}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})),\ x_i^{cc}(t_0,\boldsymbol{p}) = x_{0,i}^{cc}(\boldsymbol{p}). \end{aligned} \tag{7.5b}$$

Theorem 7.1. *Under Assumptions 7.2 and 7.3,* (7.5) *has one unique solution.*

Proof. The existence and uniqueness of a solution of (7.5a) have been verified in [28, Theorem 2].

Conditions 1) and 2) in Assumption 7.2 ensure the existence and uniqueness of a solution of (7.5b) according to the Picard-Lindelöf Theorem as summarized in [60, Theorem 1.1]. □

Theorem 7.2. *Under Assumptions 7.1, 7.2, and 7.3, let* $(\boldsymbol{x}^L, \boldsymbol{x}^U, \boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ *be a solution of* (7.5) *on* $I \times P$*. Then, the following holds:*

1) $\boldsymbol{x}^L, \boldsymbol{x}^U$ *are state bounds of* $\boldsymbol{x}$ *on* $I \times P$,

2) $\boldsymbol{x}^{cv}(t,\boldsymbol{p}) \leq \boldsymbol{x}(t,\boldsymbol{p}) \leq \boldsymbol{x}^{cc}(t,\boldsymbol{p})$ *for all* $(t,\boldsymbol{p}) \in I \times P$,

3) $[\boldsymbol{x}^{cv}(t,\boldsymbol{p}), \boldsymbol{x}^{cc}(t,\boldsymbol{p})] \subseteq [\boldsymbol{x}^L(t), \boldsymbol{x}^U(t)]$ *for all* $(t,\boldsymbol{p}) \in I \times P$,

Proof. Theorem 7.1 ensures that such a unique solution exists. Result 1) was readily verified in [28, Theorem 3].

Consider any $\bar{\boldsymbol{p}} \in P$. Result 2) is verified by showing that all three requirements in Proposition 7.1 are satisfied with $\boldsymbol{x}(\cdot,\bar{\boldsymbol{p}})$ in place of $\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger$ and $\boldsymbol{x}^{cv}(\cdot,\bar{\boldsymbol{p}}), \boldsymbol{x}^{cc}(\cdot,\bar{\boldsymbol{p}})$ in place of $\boldsymbol{\xi}^L, \boldsymbol{\xi}^U$. Conditions 2) and 3) in Assumption 7.2 ensures the first and second requirements, respectively. The third requirement is guaranteed by Condition 2) in Assumption 7.1.

Similarly, Result 3) is verified by showing that all three requirements in Proposition 7.1 are satisfied with $\boldsymbol{x}^{cv}(\cdot,\bar{\boldsymbol{p}}), \boldsymbol{x}^{cc}(\cdot,\bar{\boldsymbol{p}})$ in place of $\boldsymbol{\xi}^\dagger, \boldsymbol{\xi}^\ddagger$ and $\boldsymbol{x}^L(\cdot,\bar{\boldsymbol{p}}), \boldsymbol{x}^U(\cdot,\bar{\boldsymbol{p}})$ in place of $\boldsymbol{\xi}^L, \boldsymbol{\xi}^U$. Condition 2) in Assumption 7.2 ensures the first requirement, and Condition 3) in Assumption 7.1 ensures the third requirement. Next, we verify the enclosing dynamics in the second requirement. It suffices to show that, for a.e. $t \in I$, any $[\boldsymbol{\xi}^L, \boldsymbol{\xi}^U] \in \mathbb{IR}^{n_x}$ such that $[\boldsymbol{x}^{cv}(t,\bar{\boldsymbol{p}}), \boldsymbol{x}^{cc}(t,\bar{\boldsymbol{p}})] \subseteq [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U]$,

1. If $\zeta_i^L = x_i^{cv}(t,\bar{\boldsymbol{p}})$, then $\min_{\boldsymbol{p}\in P} f_i^L(t,\boldsymbol{p},\boldsymbol{\xi}^L,\boldsymbol{\xi}^U) \leq \dot{x}_i^{cv}(t,\bar{\boldsymbol{p}})$.

2. If $\xi_i^U = x_i^{cc}(t, \bar{\boldsymbol{p}})$, then $\max_{\boldsymbol{p}\in P} f_i^U(t, \boldsymbol{p}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \geq \dot{x}_i^{cc}(t, \bar{\boldsymbol{p}})$.

It will be shown that the first condition holds; verifying the second is analogous. According to Assumption 7.3, if $\xi_i^L = x_i^{cv}(t, \bar{\boldsymbol{p}})$, then

$$f_i^L(t, \bar{\boldsymbol{p}}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \leq f_i^{cv}(t, \bar{\boldsymbol{p}}, \boldsymbol{x}^{cv}(t, \bar{\boldsymbol{p}}), \boldsymbol{x}^{cc}(t, \bar{\boldsymbol{p}})).$$

It follows that

$$\begin{aligned}\min_{\boldsymbol{p}\in P} f_i^L(t, \boldsymbol{p}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) &\leq f_i^L(t, \bar{\boldsymbol{p}}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \\ &\leq f_i^{cv}(t, \bar{\boldsymbol{p}}, \boldsymbol{x}^{cv}(t, \bar{\boldsymbol{p}}), \boldsymbol{x}^{cc}(t, \bar{\boldsymbol{p}})) \\ &= \dot{x}_i^{cv}(t, \bar{\boldsymbol{p}}),\end{aligned}$$

which ensures the first condition. Thus, all three requirements in Proposition 7.1 are satisfied. □

Theorem 7.2 shows that $\boldsymbol{x}^L(t) \leq \boldsymbol{x}^U(t)$ and $\boldsymbol{x}^{cv}(t, \boldsymbol{p}) \leq \boldsymbol{x}^{cc}(t, \boldsymbol{p})$ for all $(t, p) \in I \times P$, and they form intervals functions X^B and X^R according to Definitions 7.5 and 7.6. This result will be used implicitly in the remainder of this article.

Theorem 7.3. *Under Assumptions 7.1, 7.2, and 7.3, let* $(\boldsymbol{x}^L, \boldsymbol{x}^U, \boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ *be a solution of* (7.5) *on* $I \times P$. *Then,* $\boldsymbol{x}^{cv}(t, \cdot)$ *and* $\boldsymbol{x}^{cc}(t, \cdot)$ *are, respectively, convex and concave on* P *for every* $t \in I$.

Proof. We proceed very similarly to the proof of [120, Theorem 3]. Theorem 7.2 shows $\boldsymbol{x}(t, \boldsymbol{p}) \in X^R(t, \boldsymbol{p}) \subseteq X^B(t)$ for each $(t, \boldsymbol{p}) \in I \times P$. Choose any fixed $\boldsymbol{p}^\dagger, \boldsymbol{p}^\ddagger \in$

P and $\lambda \in (0,1)$. For all $t \in I$, define

$$\begin{aligned}\bar{\boldsymbol{p}} &:= \lambda \boldsymbol{p}^{\mathsf{T}} + (1-\lambda)\boldsymbol{p}^{\mathsf{T}},\\ \bar{\boldsymbol{x}}^{cv}(t) &:= \lambda \boldsymbol{x}^{cv}(t,\boldsymbol{p}^{\mathsf{T}}) + (1-\lambda)\boldsymbol{x}^{cv}(t,\boldsymbol{p}^{\mathsf{T}}),\\ \bar{\boldsymbol{x}}^{cc}(t) &:= \lambda \boldsymbol{x}^{cc}(t,\boldsymbol{p}^{\mathsf{T}}) + (1-\lambda)\boldsymbol{x}^{cc}(t,\boldsymbol{p}^{\mathsf{T}}).\end{aligned}$$

To achieve a contradiction, assume that there exists $\hat{t} \in I$ such that either $x_j^{cv}(\hat{t},\bar{\boldsymbol{p}}) > \bar{x}_j^{cv}(\hat{t})$ or $x_j^{cc}(\hat{t},\bar{\boldsymbol{p}}) < \bar{x}_j^{cc}(\hat{t})$ for at least one index $j \in \{1,\ldots,n_x\}$. Define $\boldsymbol{\delta} : I \to \mathbb{R}^{2n_x}$ by

$$\boldsymbol{\delta}(t) := (\boldsymbol{x}^{cv}(t,\bar{\boldsymbol{p}}) - \bar{\boldsymbol{x}}^{cv}(t),\ \bar{\boldsymbol{x}}^{cc}(t) - \boldsymbol{x}^{cc}(t,\bar{\boldsymbol{p}})), \quad \forall t \in I.$$

Then, there is $\delta_j(\hat{t}) > 0$ or $\delta_{j+n_x}(\hat{t}) < 0$ for at least one j. Let $k^r \in \mathbb{R}_{>0}$ denote the Lipschitz constant in Condition 2) in Assumption 7.2. According to Lemma 3 in [120], there exists $j \in \{1,\ldots,n_x\}$, $t_1, t_2 \in I$ with $t_1 < t_2$, and a continuously differentiable function $\rho : I \to \mathbb{R}$ satisfying

$$0 < \rho(t) \quad \text{and} \quad \dot{\rho}(t) > (2k^r)\rho(t), \quad \forall t \in I,$$

and the inequalities

$$\boldsymbol{x}^{cv}(t,\bar{\boldsymbol{p}}) \leq \bar{\boldsymbol{x}}^{cv}(t) + \mathbf{1}\rho(t), \quad \forall t \in [t_1, t_2), \tag{7.6a}$$

$$\boldsymbol{x}^{cc}(t,\bar{\boldsymbol{p}}) \geq \bar{\boldsymbol{x}}^{cc}(t) - \mathbf{1}\rho(t), \quad \forall t \in [t_1, t_2), \tag{7.6b}$$

$$\bar{x}_j^{cv}(t) < x_j^{cv}(t,\bar{\boldsymbol{p}}) < \bar{x}_j^{cv}(t) + \rho(t), \quad \forall t \in (t_1, t_2), \tag{7.6c}$$

$$x_j^{cv}(t_2,\bar{\boldsymbol{p}}) = \bar{x}_j^{cv}(t_2) + \rho(t_2), \tag{7.6d}$$

$$x_j^{cv}(t_1,\bar{\boldsymbol{p}}) = \bar{x}_j^{cv}(t_1). \tag{7.6e}$$

A violation of concavity of $x_j^{cc}(t,\cdot)$ on $[t_1, t_2]$ for some j can be shown analogously by altering the above inequalities. Here, we assume that (7.6) holds.

Define $\boldsymbol{x}^{cv\dagger}(t) := \min(\boldsymbol{x}^{cv}(t,\bar{\boldsymbol{p}}), \bar{\boldsymbol{x}}^{cv}(t))$ and $\boldsymbol{x}^{cc\dagger}(t) := \max(\boldsymbol{x}^{cc}(t,\bar{\boldsymbol{p}}), \bar{\boldsymbol{x}}^{cc}(t))$ for all $t \in [t_1, t_2]$. Since $\boldsymbol{x}^{cv}(t,\bar{\boldsymbol{p}}) \geq \boldsymbol{x}^{cv\dagger}(t)$ for all $t \in [t_1, t_2]$, Condition 2) in Assumption 7.2 provides

$$\begin{aligned}
&\dot{x}_i^{cv}(t,\bar{\boldsymbol{p}}) \\
&\quad = f_j^{cv}(t,\bar{\boldsymbol{p}},\boldsymbol{x}^{cv}(t,\bar{\boldsymbol{p}}),\boldsymbol{x}^{cc}(t,\bar{\boldsymbol{p}})) \\
&\quad \leq f_j^{cv}(t,\bar{\boldsymbol{p}},\boldsymbol{x}^{cv\dagger}(t),\boldsymbol{x}^{cc\dagger}(t)) \\
&\qquad + k^r(\|\boldsymbol{x}^{cv}(t,\bar{\boldsymbol{p}}) - \boldsymbol{x}^{cv\dagger}(t)\| + \|\boldsymbol{x}^{cc}(t,\bar{\boldsymbol{p}}) - \boldsymbol{x}^{cc\dagger}(t)\|),
\end{aligned}$$

for a.e. $t \in [t_1, t_2]$. By (7.6a) and (7.6b), it follows that, for a.e. $t \in [t_1, t_2]$,

$$\begin{aligned}
\dot{x}_i^{cv}(t,\bar{\boldsymbol{p}}) &\leq f_j^{cv}(t,\bar{\boldsymbol{p}},\boldsymbol{x}^{cv\dagger}(t),\boldsymbol{x}^{cc\dagger}(t)) + 2k^r\rho(t) \\
&< f_j^{cv}(t,\bar{\boldsymbol{p}},\boldsymbol{x}^{cv\dagger}(t),\boldsymbol{x}^{cc\dagger}(t)) + \dot{\rho}(t).
\end{aligned}$$

Next, following Definition 7.10, we use the assumption that $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ describe convexity-preserving dynamics to show that, for a.e. $t \in [t_1, t_2]$,

$$\begin{aligned}
&\dot{x}_j^{cv}(t, \bar{\boldsymbol{p}}) \\
&\quad \leq f_j^{cv}(t, \bar{\boldsymbol{p}}, \boldsymbol{x}^{cv\dagger}(t), \boldsymbol{x}^{cc\dagger}(t)) + \dot{\rho}(t) \\
&\quad \leq \lambda f_j^{cv}(t, \boldsymbol{p}^\dagger, \boldsymbol{x}^{cv}(t, \boldsymbol{p}^\dagger), \boldsymbol{x}^{cc}(t, \boldsymbol{p}^\dagger)) \\
&\qquad + (1-\lambda) f_j^{cv}(t, \boldsymbol{p}^\ddagger, \boldsymbol{x}^{cv}(t, \boldsymbol{p}^\ddagger), \boldsymbol{x}^{cc}(t, \boldsymbol{p}^\ddagger)) + \dot{\rho}(t).
\end{aligned} \tag{7.7}$$

First, it is assured that, for a.e. $t \in [t_1, t_2]$,

$$\begin{aligned}
\boldsymbol{x}^{cv}(t, \boldsymbol{p}^\dagger) &\leq \boldsymbol{x}^{cc}(t, \boldsymbol{p}^\dagger), \\
\boldsymbol{x}^{cv}(t, \boldsymbol{p}^\ddagger) &\leq \boldsymbol{x}^{cc}(t, \boldsymbol{p}^\ddagger), \\
\boldsymbol{x}^{cv\dagger}(t) \leq \boldsymbol{x}^{cv}(t, \bar{\boldsymbol{p}}) &\leq \boldsymbol{x}^{cc}(t, \bar{\boldsymbol{p}}) \leq \boldsymbol{x}^{cc\dagger}(t).
\end{aligned}$$

Moreover, it is trivial to see that $\boldsymbol{x}^{cv}(t, \boldsymbol{q}), \boldsymbol{x}^{cc}(t, \boldsymbol{q}) \in X^B(t)$, $\forall \boldsymbol{q} \in \{\boldsymbol{p}^\dagger, \boldsymbol{p}^\ddagger, \bar{\boldsymbol{p}}\}$, and thus $\bar{\boldsymbol{x}}^{cv}(t), \bar{\boldsymbol{x}}^{cc}(t), \boldsymbol{x}^{cv\dagger}(t), \boldsymbol{x}^{cc\dagger}(t) \in X^B(t)$. Finally,

$$\begin{aligned}
\boldsymbol{x}^{cv\dagger}(t) \leq \bar{\boldsymbol{x}}^{cv}(t) &= \lambda \boldsymbol{x}^{cv}(t, \boldsymbol{p}^\dagger) + (1-\lambda)\boldsymbol{x}^{cv}(t, \boldsymbol{p}^\ddagger), \\
\boldsymbol{x}^{cc\dagger}(t) \geq \bar{\boldsymbol{x}}^{cc}(t) &= \lambda \boldsymbol{x}^{cc}(t, \boldsymbol{p}^\dagger) + (1-\lambda)\boldsymbol{x}^{cc}(t, \boldsymbol{p}^\ddagger),
\end{aligned}$$

and (7.6c) shows that, for a.e. $t \in [t_1, t_2]$,

$$x_j^{cv\dagger}(t) = \bar{x}_j^{cv}(t) = \lambda x_j^{cv}(t, \boldsymbol{p}^\dagger) + (1-\lambda) x_j^{cv}(t, \boldsymbol{p}^\ddagger).$$

Thus, (7.7) follows from applying Definition 7.10 with

$$\begin{aligned}\boldsymbol{\phi}^\dagger &:= \boldsymbol{x}^{cv}(t,\boldsymbol{p}^\dagger), \quad \boldsymbol{\phi}^\ddagger := \boldsymbol{x}^{cv}(t,\boldsymbol{p}^\ddagger), \quad \bar{\boldsymbol{\phi}} := \boldsymbol{x}^{cv\dagger}(t),\\ \boldsymbol{\psi}^\dagger &:= \boldsymbol{x}^{cc}(t,\boldsymbol{p}^\dagger), \quad \boldsymbol{\psi}^\ddagger := \boldsymbol{x}^{cc}(t,\boldsymbol{p}^\ddagger), \quad \bar{\boldsymbol{\psi}} := \boldsymbol{x}^{cc\dagger}(t).\end{aligned}$$

Observe that (7.7) is equivalent to

$$\dot{x}_j^{cv}(t,\bar{\boldsymbol{p}}) \leq \lambda \dot{x}_j^{cv}(t,\boldsymbol{p}^\dagger) + (1-\lambda)\dot{x}_j^{cv}(t,\boldsymbol{p}^\ddagger) + \dot{\rho}(t). \tag{7.8}$$

According to Theorem 1 in [120], (7.8) implies that $\dot{x}_j^{cv}(t,\bar{\boldsymbol{p}}) - \dot{\bar{x}}_j^{cv}(t) - \dot{\rho}(t)$ is non-increasing on $[t_1,t_2]$. So,

$$x_j^{cv}(t_2,\bar{\boldsymbol{p}}) - \bar{x}_j^{cv}(t_2) - \rho(t_2) \leq x_j^{cv}(t_1,\bar{\boldsymbol{p}}) - \bar{x}_j^{cv}(t_1) - \rho(t_1).$$

(7.6d) and (7.6e) suggest that $0 \geq \rho(t_1)$, which is a contradiction. Hence,

$$\begin{aligned}\boldsymbol{x}^{cv}(t,\lambda\boldsymbol{p}^\dagger + (1-\lambda)\boldsymbol{p}^\ddagger) &= \boldsymbol{x}^{cv}(t,\bar{\boldsymbol{p}})\\ &\leq \bar{\boldsymbol{x}}^{cv}(t)\\ &= \lambda\boldsymbol{x}^{cv}(t,\boldsymbol{p}^\dagger) + (1-\lambda)\boldsymbol{x}^{cv}(t,\boldsymbol{p}^\ddagger),\\ \boldsymbol{x}^{cc}(t,\lambda\boldsymbol{p}^\dagger + (1-\lambda)\boldsymbol{p}^\ddagger) &= \boldsymbol{x}^{cc}(t,\bar{\boldsymbol{p}})\\ &\geq \bar{\boldsymbol{x}}^{cc}(t)\\ &= \lambda\boldsymbol{x}^{cc}(t,\boldsymbol{p}^\dagger) + (1-\lambda)\boldsymbol{x}^{cc}(t,\boldsymbol{p}^\ddagger),\end{aligned}$$

for all $t \in I$. Since $\boldsymbol{p}^\dagger, \boldsymbol{p}^\ddagger \in P$ and $\lambda \in (0,1)$ was chosen arbitrarily, the above inequalities hold for all $\boldsymbol{p}^\dagger, \boldsymbol{p}^\ddagger \in P$ and $\lambda \in (0,1)$. □

Combining Theorems 7.2 and 7.3, it can be concluded that the solution of (7.5) provides valid state bounds and state relaxations for (7.1) on $I \times P$. Moreover, the state relaxations are tighter than the state bounds.

Next, we show that if $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ are smooth, then the generated state relaxations from (7.5b) are smooth.

Assumption 7.4. *$\boldsymbol{f}^{cv}(t,\cdot,\cdot,\cdot)$, $\boldsymbol{f}^{cc}(t,\cdot,\cdot,\cdot)$ are differentiable on $P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ for every $t \in I$.*

Theorem 7.4. *Under Assumptions 7.1, 7.2, 7.3, and 7.4, let $(\boldsymbol{x}^L, \boldsymbol{x}^U, \boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ be a solution of* (7.5) *on $I \times P$. Then, $\boldsymbol{x}^{cv}$ and $\boldsymbol{x}^{cc}$ are continuously differentiable on $I \times P$.*

Proof. The claimed result follows from [60, Chapter V, Theorem 3.1]. □

Sections 7.5 and 7.6 introduce methods for constructing the RHS of state bound system (7.5a) and the RHS of the state relaxation system (7.5b), so that Assumptions 7.2 and 7.3 are satisfied.

7.5 Constructing State Bound RHS

Assumption 7.5. *Assume that interval function $\check{\boldsymbol{F}}^B \equiv [\check{\boldsymbol{f}}^L, \check{\boldsymbol{f}}^U] : I \times P \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ satisfies the following:*

1) *$\check{\boldsymbol{f}}^L$, $\check{\boldsymbol{f}}^U$ are continuous,*
2) *$\check{\boldsymbol{f}}^L(t,\boldsymbol{p},\cdot,\cdot)$, $\check{\boldsymbol{f}}^U(t,\boldsymbol{p},\cdot,\cdot)$ are Lipschitz continuous on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$, uniformly in $(t,\boldsymbol{p})$,*
3) *$\Xi^B \mapsto \check{\boldsymbol{F}}^B(t,\boldsymbol{p},\Xi^B)$ is an inclusion function of $\boldsymbol{f}(t,\boldsymbol{p},\cdot)$ on $\mathbb{IR}^{n_x}$ for a.e. $t \in I$ and any $\boldsymbol{p} \in P$.*

Under Assumption 7.5, consider $\boldsymbol{f}^L, \boldsymbol{f}^U$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} f_i^L(t, \boldsymbol{p}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) &:= \check{f}_i^L(t, \boldsymbol{p}, B_i^L([\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])), \\ f_i^U(t, \boldsymbol{p}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) &:= \check{f}_i^U(t, \boldsymbol{p}, B_i^U([\boldsymbol{\xi}^L, \boldsymbol{\xi}^U])). \end{aligned} \tag{7.9}$$

It was verified in [28] that $\boldsymbol{f}^L, \boldsymbol{f}^U$ in (7.9) satisfy Conditions 1)-3) in Assumption 7.2.

Four approaches for constructing $\check{\boldsymbol{f}}^L, \check{\boldsymbol{f}}^U$ have been developed in [28]. A brief summary is provided as follows.

The first approach constructs $\check{\boldsymbol{f}}^L, \check{\boldsymbol{f}}^U$ using GMC and differentiable McCormick relaxations (DMC) [74, 72, 73], as well as piecewise-affine (PA) approximations and affine approximations of twice-continuously differentiable McCormick relaxations $\mathscr{C}^2$-DMC [74]. These methods are listed in Category II of Table 7.1. Note that techniques for generating PA approximations and affine approximations of nonlinear convex relaxations have been introduced in [28].

In the second approach, we consider $\hat{\boldsymbol{f}}^L, \hat{\boldsymbol{f}}^U : I \times P \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ such that $\hat{\boldsymbol{f}}^L(t, \cdot, \cdot), \hat{\boldsymbol{f}}^U(t, \cdot, \cdot)$ are convex and concave relaxations of $\boldsymbol{f}(t, \cdot, \cdot)$, respectively, on $P \times X^B(t)$ for each $t \in I$. They can be generated using GMC, DMC, αBB relaxations, as well as PA approximations and affine approximations of $\mathscr{C}^2$-DMC and αBB relaxations; see Category I of Table 7.1. Then, we let

$$\begin{aligned} \check{f}_i^L(t, \boldsymbol{p}, \Xi^B) &:= \min_{\boldsymbol{p} \in P, \boldsymbol{z} \in \Xi^B} \hat{f}_i^L(t, \boldsymbol{p}, \boldsymbol{z}), \\ \check{f}_i^U(t, \boldsymbol{p}, \Xi^B) &:= \max_{\boldsymbol{p} \in P, \boldsymbol{z} \in \Xi^B} \hat{f}_i^U(t, \boldsymbol{p}, \boldsymbol{z}). \end{aligned} \tag{7.10}$$

In the third approach, we consider $\tilde{\boldsymbol{f}}^L, \tilde{\boldsymbol{f}}^U : I \times \mathbb{IR}^{n_p} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ such that $\tilde{\boldsymbol{f}}^L(t, P, \cdot), \tilde{\boldsymbol{f}}^U(t, P, \cdot)$ are convex and concave relaxations of $\boldsymbol{f}(t, \boldsymbol{p}, \cdot)$, respectively,

on $X^B(t)$ for each $(t,\boldsymbol{p}) \in I \times P$. They can be generated using GMC, DMC, as well as PA approximations and affine approximations of $\mathscr{C}^2$-DMC; see Category III of Table 7.1. Then, we let

$$\begin{aligned} \check{f}_i^L(t,\boldsymbol{p},\Xi^B) &:= \min_{\boldsymbol{z}\in\Xi^B} \tilde{f}_i^L(t,P,\boldsymbol{z}), \\ \check{f}_i^U(t,\boldsymbol{p},\Xi^B) &:= \max_{\boldsymbol{z}\in\Xi^B} \tilde{f}_i^U(t,P,\boldsymbol{z}). \end{aligned} \tag{7.11}$$

In the last approach, we consider $\bar{\boldsymbol{f}}^L, \bar{\boldsymbol{f}}^U : I \times \mathbb{IR}^{n_p} \times \mathbb{IR}^{n_x} \to \mathbb{R}^{n_x}$ such that $(\hat{P},\Xi) \mapsto [\bar{\boldsymbol{f}}^L(t,\hat{P},\Xi), \bar{\boldsymbol{f}}^U(t,\hat{P},\Xi)]$ is an inclusion function of $\boldsymbol{f}(t,\cdot,\cdot)$ on $P \times X(t)$ for each $t \in I$. They can be generated using interval extension methods, including NIE, GMC, and DMC; see Category IV of Table 7.1. Then, we let

$$\begin{aligned} \check{f}_i^L(t,\boldsymbol{p},\Xi^B) &:= \bar{f}_i^L(t,P,\Xi^B), \\ \check{f}_i^U(t,\boldsymbol{p},\Xi^B) &:= \bar{f}_i^U(t,P,\Xi^B). \end{aligned} \tag{7.12}$$

7.6 Constructing State Relaxation RHS

This section introduces two strategies for constructing $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ in (7.5b), given that state bound X^B is available on I. They both involve applying the flattening operators to inclusion functions of $\boldsymbol{f}(t,\cdot,\cdot)$.

Assumption 7.6. *Assume that interval function $\check{F}^R \equiv [\check{\boldsymbol{f}}^{cv}, \check{\boldsymbol{f}}^{cc}] : I \times \mathbb{R}^{n_p} \times \mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ satisfies the following conditions: for any $\Xi^B \in \mathbb{IR}^{n_x}$,*

1) *$\check{\boldsymbol{f}}^{cv}, \check{\boldsymbol{f}}^{cc}$ are continuous,*

Table 7.1: Summary of available methods

Category	Domain	Relaxation method	Code	Category score	Relax. score	Approx. score
I	$I \times \mathbb{R}^{n_p} \times \mathbb{R}^{n_x}$	Nonlinear GMC	px-N-G	1	2	0
		Nonlinear DMC	px-N-D	1	3	0
		PA $\mathscr{C}^2$-DMC	px-P-D	1	3	1
		Affine $\mathscr{C}^2$-DMC	px-A-D	1	3	2
		αBB	px-N-α	1	$1+\alpha$	0
		PA with αBB	px-P-α	1	$1+\alpha$	1
		Affine with αBB	px-A-α	1	$1+\alpha$	2
II	$I \times \mathbb{R}^{n_p} \times \mathbb{IR}^{n_x}$	Nonlinear GMC	p-N-G	2	2	0
		Nonlinear DMC	p-N-D	2	3	0
		PA $\mathscr{C}^2$-DMC	p-P-D	2	3	1
		Affine $\mathscr{C}^2$-DMC	p-A-D	2	3	2
III	$I \times \mathbb{IR}^{n_p} \times \mathbb{R}^{n_x}$	Nonlinear GMC	x-N-G	2	2	0
		Nonlinear DMC	x-N-D	2	3	0
		PA $\mathscr{C}^2$-DMC	x-P-D	2	3	1
		Affine $\mathscr{C}^2$-DMC	x-A-D	2	3	2
IV	$I \times \mathbb{IR}^{n_p} \times \mathbb{IR}^{n_x}$	Nonlinear GMC	○-N-G	3	2	0
		Nonlinear DMC	○-N-D	3	3	0
		NIE	○-N-I	3	4	0

2) $\check{\boldsymbol{f}}^{cv}(t,\boldsymbol{p},\cdot,\cdot,\Xi^B)$, $\check{\boldsymbol{f}}^{cc}(t,\boldsymbol{p},\cdot,\cdot,\Xi^B)$ *are locally Lipschitz continuous on* $\Xi^B \times \Xi^B$ *, uniformly in* $(t,\boldsymbol{p})$*,*

3) $\Xi^R \mapsto \check{F}^R(t,\boldsymbol{p},\Xi^R,\Xi^B)$ *is an inclusion function of* $\boldsymbol{f}(t,\boldsymbol{p},\cdot)$ *on* Ξ^B *for a.e.* $t \in I$ *and any* $\boldsymbol{p} \in P$*,*

4) $(t,\boldsymbol{p},\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc}) \mapsto \check{F}^R(t,\boldsymbol{p},[\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc}],\Xi^B)$ *describe convexity-amplifying dynamics on* Ξ^B*.*

Assumption 7.7. *Assume that* $\check{\boldsymbol{f}}^U, \check{\boldsymbol{f}}^L$ *describe inclusion-amplifying dynamics about* $\check{\boldsymbol{f}}^{cv}, \check{\boldsymbol{f}}^{cc}$*.*

Consider $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} f_i^{cv}(t, \boldsymbol{p}, \boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}) &:= \breve{f}_i^{cv}(t, \boldsymbol{p}, B_i^L([\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}]), X^B(t)), \\ f_i^{cc}(t, \boldsymbol{p}, \boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}) &:= \breve{f}_i^{cc}(t, \boldsymbol{p}, B_i^U([\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}]), X^B(t)). \end{aligned} \tag{7.13}$$

Since B_i^L, B_i^U are linear operations, it has been readily verified that $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ in (7.13) satisfy continuous and Lipschitz conditions in Assumption 7.2 under Assumption 7.6. Moreover, Lemma 7.1 ensures that, under Assumption 7.7, $\boldsymbol{f}^L, \boldsymbol{f}^U$ in (7.13) describe inclusion-preserving dynamics about $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ as in Assumption 7.3. Next, we verify that $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ in (7.13) satisfy the enclosing and convexity properties desired in Assumptions 7.2.

Lemma 7.2. *Under Assumption 7.6, $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$ defined in* (7.13) *describe enclosing dynamics about $\boldsymbol{x}$.*

Proof. According to Definition 7.8, it suffices to show that, for a.e. $t \in I$, any $\bar{\boldsymbol{p}} \in P$ and $\Xi^R \equiv [\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}] \in \mathbb{IR}^{n_x}$ such that $\boldsymbol{x}(t, \bar{\boldsymbol{p}}) \subseteq \Xi^R \subseteq X^B(t)$,

1. If $\xi_i^{cv} = \dot{x}_i(t, \bar{\boldsymbol{p}})$, then $f_i^{cv}(t, \bar{\boldsymbol{p}}, \Xi^R) \leq \dot{x}_i(t, \bar{\boldsymbol{p}})$.
2. If $\xi_i^{cc} = \dot{x}_i(t, \bar{\boldsymbol{p}})$, then $f_i^{cc}(t, \bar{\boldsymbol{p}}, \Xi^R) \geq \dot{x}_i(t, \bar{\boldsymbol{p}})$.

It will be shown that the first condition holds; verifying the second is analogous.

If $\xi_i^{cv} = \dot{x}_i(t, \bar{\boldsymbol{p}})$, the flattening operation ensures that $\boldsymbol{x}(t, \bar{\boldsymbol{p}}) \in B_i^L(\Xi^R)$. The third condition in Assumption 7.6 ensures that

$$\breve{f}_i^{cv}(t, \bar{\boldsymbol{p}}, B_i^L(\Xi^R), X^B(t)) \leq f_i(t, \bar{\boldsymbol{p}}, \boldsymbol{x}(t, \bar{\boldsymbol{p}})),$$

which is equivalent to

$$f_i^{cv}(t, \bar{\boldsymbol{p}}, \Xi^R) \leq \dot{x}_i(t, \bar{\boldsymbol{p}}).$$

Thus, the first condition is verified. □

Lemma 7.3. *Under Assumption 7.6, $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ defined in* (7.13) *describe convexity-preserving dynamics on $X^B(t)$ for every $t \in I$.*

Proof. Condition 4) in Assumption 7.6 shows that, for every $t \in I$, $(t, \boldsymbol{p}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \check{F}^R(t, \boldsymbol{p}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U], X^B(t))$ describes convexity-amplifying dynamics on $X^B(t)$. After flattening operation, $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ defined in (7.13) describe convexity-preserving dynamics according to [120, Lemma 11]. □

Assumption 7.8. *$\check{\boldsymbol{f}}^{cv}(t, \cdot, \cdot, \cdot, \Xi^B)$, $\check{\boldsymbol{f}}^{cc}(t, \cdot, \cdot, \cdot, \Xi^B)$ are differentiable on $P \times \Xi^B \times \Xi^B$ for all $t \in I$, $\Xi^B \in \mathbb{IR}^{n_x}$.*

Lemma 7.4. *Under Assumption 7.8, $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ defined in* (7.13) *satisfy Assumption 7.4.*

Proof. Since B_i^L, B_i^U are linear operators, the claimed result holds. □

In the remainder of this subsection, we introduce two strategies to construct $\check{\boldsymbol{f}}^{cv}$, $\check{\boldsymbol{f}}^{cc}$ that satisfy Assumption 7.6. Assumption 7.7 will be discussed in Section 7.7.

7.6.1 Generalized convex relaxations

The first strategy is adapted from Scott and Barton [120] in which $\check{\boldsymbol{f}}^{cv}$, $\check{\boldsymbol{f}}^{cc}$ are constructed with GMC. It was validated in [120, Section 4.2] that $\check{\boldsymbol{f}}^{cv}$, $\check{\boldsymbol{f}}^{cc}$ satisfy Assumption 7.6. Similar arguments hold for DMC. Moreover, the PA approximations

and affine approximations of $\mathscr{C}^2$-DMC are also valid options for generating $\check{\boldsymbol{f}}^{cv}$, $\check{\boldsymbol{f}}^{cc}$ and they were validated in [28]. Theses methods are summarized in Category II in Table 7.1.

7.6.2 Optimization-based method

The second strategy is adapted from [131] where convex optimization problems are embedded in the RHS of the auxiliary ODEs.

Assumption 7.9. *Assume that interval function* $\hat{F}^R \equiv [\hat{\boldsymbol{f}}^{cv}, \hat{\boldsymbol{f}}^{cc}] : I \times P \times \mathbb{R}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ *satisfies the following conditions: for any* $\Xi^B \in \mathbb{IR}^{n_x}$,

1) $\hat{\boldsymbol{f}}^{cv}$, $\hat{\boldsymbol{f}}^{cc}$ *are continuous,*

2) $\hat{\boldsymbol{f}}^{cv}(t, \boldsymbol{p}, \cdot, \Xi^B)$, $\hat{\boldsymbol{f}}^{cc}(t, \boldsymbol{p}, \cdot, \Xi^B)$ *are Lipschitz continuous on* Ξ^B*, uniformly in* $(t, \boldsymbol{p})$*,*

3) $\hat{\boldsymbol{f}}^{cv}(t, \cdot, \cdot, \Xi^B)$, $\hat{\boldsymbol{f}}^{cc}(t, \cdot, \cdot, \Xi^B)$ *are convex and concave relaxations of* $\boldsymbol{f}(t, \cdot, \cdot)$*, respectively, on* $P \times \Xi^B$ *for a.e.* $t \in I$*.*

Under Assumption 7.9, consider $\check{F}^R$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned} \check{f}_i^{cv}(t, \boldsymbol{p}, \Xi^R, \Xi^B) &:= \min_{\boldsymbol{z} \in \Xi^R} \hat{f}_i^{cv}(t, \boldsymbol{p}, \boldsymbol{z}, \Xi^B), \\ \check{f}_i^{cc}(t, \boldsymbol{p}, \Xi^R, \Xi^B) &:= \max_{\boldsymbol{z} \in \Xi^R} \hat{f}_i^{cc}(t, \boldsymbol{p}, \boldsymbol{z}, \Xi^B). \end{aligned} \tag{7.14}$$

It was verified in [131] that $\check{F}^R$ in (7.14) satisfy Conditions 1) and 2) in Assumption 7.6. We will show that $\check{F}^R$ also satisfy Conditions 3) and 4) in Assumption 7.6.

Lemma 7.5. *Under Assumption 7.9, for a.e.* $t \in I$*, any* $\boldsymbol{p} \in P$*, and any* $\Xi^B \in \mathbb{IR}^{n_x}$*,* $\Xi^R \mapsto \check{F}^R(t, \boldsymbol{p}, \Xi^R, \Xi^B)$ *is an inclusion function of* $\boldsymbol{f}(t, \boldsymbol{p}, \cdot)$ *on* Ξ^B*.*

Proof. According to Definition 7.2, it suffices to show that, for each $i \in \{1, \ldots, n_x\}$, any $\boldsymbol{p} \in P$, $Z \subseteq \Xi^B$, and $\bar{\boldsymbol{z}} \in Z$,

$$\check{f}_i^{cv}(t, \boldsymbol{p}, Z, \Xi^B) \leq f_i(t, \boldsymbol{p}, \bar{\boldsymbol{z}}),$$
$$\check{f}_i^{cc}(t, \boldsymbol{p}, Z, \Xi^B) \geq f_i(t, \boldsymbol{p}, \bar{\boldsymbol{z}}).$$

It will be shown that the first inequality holds; showing the second is analogous.

Condition 3) in Assumption 7.9 shows that

$$\hat{f}_i^{cv}(t, \boldsymbol{p}, \bar{\boldsymbol{z}}, \Xi^B) \leq f_i(t, \boldsymbol{p}, \bar{\boldsymbol{z}}).$$

It follows that

$$\begin{aligned}\check{f}_i^{cv}(t, \boldsymbol{p}, Z, \Xi^B) &= \min_{\boldsymbol{z} \in Z} \hat{f}_i^{cv}(t, \boldsymbol{p}, \boldsymbol{z}, \Xi^B) \\ &\leq \hat{f}_i^{cv}(t, \boldsymbol{p}, \bar{\boldsymbol{z}}, \Xi^B) \\ &\leq f_i(t, \boldsymbol{p}, \bar{\boldsymbol{z}}),\end{aligned}$$

which ensures the first inequality. □

Lemma 7.6. *Under Assumption 7.9, for each* $\Xi^B \in \mathbb{IR}$, $(t, \boldsymbol{p}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \check{\boldsymbol{f}}^{cv}(t, \boldsymbol{p}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U], \Xi^B)$ *and* $(t, \boldsymbol{p}, \boldsymbol{\xi}^L, \boldsymbol{\xi}^U) \mapsto \check{\boldsymbol{f}}^{cc}(t, \boldsymbol{p}, [\boldsymbol{\xi}^L, \boldsymbol{\xi}^U], \Xi^B)$ *defined in* (7.14) *describe convexity-amplifying dynamics on* Ξ^B.

Proof. Consider any $(\lambda, \boldsymbol{p}^\dagger, \boldsymbol{p}^\ddagger) \in (0,1) \times P \times P$, each $i \in \{1, \ldots, n_x\}$, a.e. $t \in I$, and any $\boldsymbol{\phi}^\dagger, \boldsymbol{\phi}^\ddagger, \bar{\boldsymbol{\phi}}, \boldsymbol{\psi}^\dagger, \boldsymbol{\psi}^\ddagger, \bar{\boldsymbol{\psi}} \in \Xi^B$ such that the following three conditions all hold:

1. $\bar{\boldsymbol{\phi}} \leq \lambda \boldsymbol{\phi}^\dagger + (1-\lambda)\boldsymbol{\phi}^\ddagger$,

2. $\bar{\boldsymbol{\psi}} \geq \lambda \boldsymbol{\psi}^{\dagger} + (1-\lambda)\boldsymbol{\psi}^{\ddagger}$, and

3. $\boldsymbol{\phi}^{\dagger} \leq \boldsymbol{\psi}^{\dagger}$, $\boldsymbol{\phi}^{\ddagger} \leq \boldsymbol{\psi}^{\ddagger}$, $\bar{\boldsymbol{\phi}} \leq \bar{\boldsymbol{\psi}}$.

It suffices to show that $\check{\boldsymbol{f}}^{cv}$ and $\check{\boldsymbol{f}}^{cc}$ satisfy:

$$\begin{aligned}
&\check{f}_i^{cv}(t, \bar{\boldsymbol{p}}, [\bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}], \Xi^B) \\
&\quad \leq \lambda \check{f}_i^{cv}(t, \boldsymbol{p}^{\dagger}, [\boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}], \Xi^B) + (1-\lambda)\check{f}_i^{cv}(t, \boldsymbol{p}^{\ddagger}, [\boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}], \Xi^B), \\
&\check{f}_i^{cc}(t, \bar{\boldsymbol{p}}, [\bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}}], \Xi^B) \\
&\quad \geq \lambda \check{f}_i^{cc}(t, \boldsymbol{p}^{\dagger}, [\boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}], \Xi^B) + (1-\lambda)\check{f}_i^{cc}(t, \boldsymbol{p}^{\ddagger}, [\boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}], \Xi^B),
\end{aligned}$$

where $\bar{\boldsymbol{p}} \equiv \lambda \boldsymbol{p}^{\dagger} + (1-\lambda)\boldsymbol{p}^{\ddagger}$. It will be verified that the first inequality holds; showing the second is analogous.

Since $\boldsymbol{\phi}^{\dagger} \leq \boldsymbol{\psi}^{\dagger}$, $\boldsymbol{\phi}^{\ddagger} \leq \boldsymbol{\psi}^{\ddagger}$, and $\bar{\boldsymbol{\phi}} \leq \bar{\boldsymbol{\psi}}$, the first equation in (7.14) shows that, for all $(\boldsymbol{q}, \boldsymbol{\phi}, \boldsymbol{\psi}) \in \{(\boldsymbol{p}^{\dagger}, \boldsymbol{\phi}^{\dagger}, \boldsymbol{\psi}^{\dagger}), (\boldsymbol{p}^{\ddagger}, \boldsymbol{\phi}^{\ddagger}, \boldsymbol{\psi}^{\ddagger}), (\bar{\boldsymbol{p}}, \bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\psi}})\}$,

$$\check{f}_i^{cv}(t, \boldsymbol{q}, [\boldsymbol{\phi}, \boldsymbol{\psi}], \Xi^B) = \min_{\boldsymbol{z} \in [\boldsymbol{\phi}, \boldsymbol{\psi}]} \hat{f}_i^{cv}(t, \boldsymbol{q}, \boldsymbol{z}, \Xi^B). \tag{7.15}$$

Let $\boldsymbol{z}^{\dagger,*}$ and $\boldsymbol{z}^{\ddagger,*}$ solve the above optimization problem at $\boldsymbol{q} \equiv \boldsymbol{p}^{\dagger}$ and $\boldsymbol{q} \equiv \boldsymbol{p}^{\ddagger}$, respectively. Define $\bar{\boldsymbol{z}}^* := \lambda \boldsymbol{z}^{\dagger,*} + (1-\lambda)\boldsymbol{z}^{\ddagger,*}$. Because $\hat{f}_i^{cv}(t, \cdot, \cdot, \Xi^B)$ is convex on $P \times \Xi^B$,

$$\begin{aligned}
&\hat{f}_i^{cv}(t, \bar{\boldsymbol{p}}, \bar{\boldsymbol{z}}^*, \Xi^B) \\
&\quad \leq \lambda \hat{f}_i^{cv}(t, \boldsymbol{p}^{\dagger}, \boldsymbol{z}^{\dagger,*}, \Xi^B) + (1-\lambda)\hat{f}_i^{cv}(t, \boldsymbol{p}^{\ddagger}, \boldsymbol{z}^{\ddagger,*}, \Xi^B).
\end{aligned}$$

Thus,

$$
\begin{aligned}
&\check{f}_i^{cv}(t,\bar{\boldsymbol{p}},[\bar{\boldsymbol{\phi}},\bar{\boldsymbol{\psi}}],\Xi^B)\\
&\quad=\min_{\boldsymbol{z}\in[\bar{\boldsymbol{\phi}},\bar{\boldsymbol{\psi}}]}\hat{f}_i^{cv}(t,\bar{\boldsymbol{p}},\boldsymbol{z},\Xi^B)\\
&\quad\le\hat{f}_i^{cv}(t,\bar{\boldsymbol{p}},\bar{\boldsymbol{z}}^*,\Xi^B)\\
&\quad\le\lambda\hat{f}_i^{cv}(t,\boldsymbol{p}^\dagger,\boldsymbol{z}^{\dagger,*},\Xi^B)+(1-\lambda)\hat{f}_i^{cv}(t,\boldsymbol{p}^\ddagger,\boldsymbol{z}^{\ddagger,*},\Xi^B)\\
&\quad=\lambda\min_{\boldsymbol{z}\in[\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger]}\hat{f}_i^{cv}(t,\boldsymbol{p}^\dagger,\boldsymbol{z},\Xi^B)\\
&\qquad+(1-\lambda)\min_{\boldsymbol{z}\in[\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger]}\hat{f}_i^{cv}(t,\boldsymbol{p}^\ddagger,\boldsymbol{z},\Xi^B)\\
&\quad=\lambda\check{f}_i^{cv}(t,\boldsymbol{p}^\dagger,[\boldsymbol{\phi}^\dagger,\boldsymbol{\psi}^\dagger],\Xi^B)\\
&\qquad+(1-\lambda)\check{f}_i^{cv}(t,\boldsymbol{p}^\ddagger,[\boldsymbol{\phi}^\ddagger,\boldsymbol{\psi}^\ddagger],\Xi^B),
\end{aligned}
$$

which verifies the first inequality. □

7.7 Ensuring Inclusion-amplifying Dynamics

In this section, we address Assumption 7.7 by discussing three different scenarios of $\check{F}^B$ and $\check{F}^R$.

7.7.1 Tighter relaxations

The first scenario assumes that $\check{F}^R$ is tighter than $\check{F}^B$ and is inclusion monotonic.

Assumption 7.10. *Assume the following holds: for a.e. $t\in I$, any $\boldsymbol{p}\in P$ and $\Xi^B\in\mathbb{IR}^{n_x}$.*

1. *$\Xi\mapsto\check{F}^R(t,\boldsymbol{p},\Xi,\Xi^B)$ is inclusion monotonic on Ξ^B.*

2. $\Xi \mapsto \check{F}^R(t, \boldsymbol{p}, \Xi, \Xi^B)$ *is tighter than* $\Xi \mapsto \check{F}^B(t, \boldsymbol{p}, \Xi)$ *on* Ξ^B.

According to Definitions 7.2 and 7.11, it was readily verified that, if interval functions $\check{F}^R$ and $\check{F}^B$ satisfy Assumption 7.10, then they also satisfy the inclusion-amplifying dynamics in Assumption 7.7.

If $\check{F}^R$ is constructed with the strategy in Section 7.6.1, then $\check{F}^R$ being tighter than $\check{F}^B$ is straightforward according to Definition 7.2. We elaborate more on the second strategy in Section 7.6.2 where $\check{F}^R$ is generated with optimizing convex relaxations. We show that, if $\check{F}^B$ is an inclusion function of $\hat{\boldsymbol{f}}^{cv}$ and $\hat{\boldsymbol{f}}^{cc}$, then $\check{F}^R$ is tighter than $\check{F}^B$.

Lemma 7.7. *Assume that for a.e.* $t \in I$, *any* $\boldsymbol{p} \in P$, *and any* $\Xi^B \in \mathbb{IR}^{n_x}$, $\Xi \mapsto \check{F}^B(t, \boldsymbol{p}, \Xi)$ *is an inclusion function of* $\hat{\boldsymbol{f}}^{cv}(t, \boldsymbol{p}, \cdot, \Xi^B)$ *and* $\hat{\boldsymbol{f}}^{cc}(t, \boldsymbol{p}, \cdot, \Xi^B)$ *on* Ξ^B. *Then,* $\Xi \mapsto \check{F}^R(t, \boldsymbol{p}, \Xi, \Xi^B)$ *is tighter than* $\Xi \mapsto \check{F}^B(t, \boldsymbol{p}, \Xi)$ *in* (7.14) *on* Ξ^B.

Proof. Consider for a.e. $t \in I$, each $i \in \{1, \ldots, n_x\}$, any $\boldsymbol{p} \in P$ and any $Z \subseteq \Xi^B$. According to Definition 7.2, it suffices to show that,

$$\check{f}_i^L(t, \boldsymbol{p}, Z) \leq \check{f}_i^{cv}(t, \boldsymbol{p}, Z, \Xi^B),$$
$$\check{f}_i^U(t, \boldsymbol{p}, Z) \geq \check{f}_i^{cc}(t, \boldsymbol{p}, Z, \Xi^B).$$

It will be shown that the first inequality holds; showing the second is analogous.

$\Xi \mapsto \check{F}^B(t, \boldsymbol{p}, \Xi)$ being an inclusion function of $\hat{\boldsymbol{f}}^{cv}(t, \boldsymbol{p}, \cdot, \Xi^B)$ ensures that, for any $\bar{\boldsymbol{z}} \in Z$,

$$\check{f}_i^L(t, \boldsymbol{p}, Z) \leq \hat{f}_i^{cv}(t, \boldsymbol{p}, \bar{\boldsymbol{z}}, \Xi^B).$$

It follows that

$$\check{f}_i^L(t, \boldsymbol{p}, Z) \leq \min_{\boldsymbol{z} \in Z} \hat{f}_i^{cv}(t, \boldsymbol{p}, \boldsymbol{z}, \Xi^B) = \check{f}_i^{cv}(t, \boldsymbol{p}, Z, \Xi^B),$$

which ensures the first inequality. □

In addition to summarizing methods for constructing $\check{F}^B$ and $\check{F}^R$, Table 7.1 is also a useful tool for choosing a pair of $\check{F}^B$ and $\check{F}^R$ that satisfy Assumptions 7.10. The steps of using Table 7.1 are described as follows:

i. Choose $\check{F}^R$ from Categories I or II in Table 7.1.

ii. Choose $\check{F}^B$ from any category in Table 7.1 with higher or equivalent category score, relaxation score, and approximation score. However, if we chose $\check{F}^R$ is from the Category II, then $\check{F}^B$ cannot be from Category III.

iii. If $\check{F}^B$ and $\check{F}^R$ are affine or piecewise-affine methods, then the linearization points for constructing $\check{F}^B$ must contain the those linearization points used for constructing $\check{F}^R$.

7.7.2 Max-landing

In the second scenario, we suppose that Assumption 7.10 does not hold. We present an approach to enforce Assumption 7.7.

Assumption 7.11. *Suppose that $\check{\boldsymbol{f}}^L$, $\check{\boldsymbol{f}}^U$ describe convexity-amplifying dynamics.*

Under Assumption 7.11, consider interval function $\check{F}^{R*} = [\check{\boldsymbol{f}}^{cv*}, \check{\boldsymbol{f}}^{cc*}] : I \times P \times$

$\mathbb{IR}^{n_x} \times \mathbb{IR}^{n_x} \to \mathbb{IR}^{n_x}$ such that, for each $i \in \{1, \ldots, n_x\}$,

$$\begin{aligned}
&\check{f}_i^{cv*}(t, \boldsymbol{p}, \Xi^R, \Xi^B) \\
&\quad := \max\left\{\check{f}_i^{cv}(t, \boldsymbol{p}, \Xi^R, \Xi^B), \check{f}_i^{L}(t, \boldsymbol{p}, \Xi^R)\right\}, \qquad (7.16a)\\
&\check{f}_i^{cc*}(t, \boldsymbol{p}, \Xi^R, \Xi^B) \\
&\quad := \min\left\{\check{f}_i^{cc}(t, \boldsymbol{p}, \Xi^R, \Xi^B), \check{f}_i^{U}(t, \boldsymbol{p}, \Xi^R)\right\}. \qquad (7.16b)
\end{aligned}$$

In the max-landing approach, we use $\check{\boldsymbol{f}}^{cv*}, \check{\boldsymbol{f}}^{cc*}$ to replace $\check{\boldsymbol{f}}^{cv}, \check{\boldsymbol{f}}^{cc}$ in (7.13). So, we need to verify that they satisfy Assumptions 7.6 and 7.7.

Lemma 7.8. *Under Assumptions 7.5, 7.6 and 7.11, $\check{\boldsymbol{f}}^{cv*}, \check{\boldsymbol{f}}^{cc*}$ defined in* (7.16) *satisfy Assumptions 7.6 and 7.7 in place of $\check{\boldsymbol{f}}^{cv}, \check{\boldsymbol{f}}^{cc}$.*

Proof. Conditions 1)-3) in Assumption 7.5 and Conditions 1)-3) in Assumption 7.6 ensure that $\check{\boldsymbol{f}}^{cv*}, \check{\boldsymbol{f}}^{cc*}$ satisfy Conditions 1)-3) in Assumption 7.6 in place of $\check{\boldsymbol{f}}^{cv}, \check{\boldsymbol{f}}^{cc}$. Since max and min functions preserve convexity and concavity [25], respectively, Condition 4) in Assumption 7.6 and Assumption 7.11 ensure that $\check{\boldsymbol{f}}^{cv*}, \check{\boldsymbol{f}}^{cc*}$ satisfy Conditions 4) in Assumption 7.6 in place of $\check{\boldsymbol{f}}^{cv}, \check{\boldsymbol{f}}^{cc}$. Hence, Assumption 7.6 is verified.

Next, we verify Assumption 7.7 by showing that $\check{\boldsymbol{f}}^{L}, \check{\boldsymbol{f}}^{U}$ describe inclusion-amplifying dynamics about $\check{\boldsymbol{f}}^{cv*}, \check{\boldsymbol{f}}^{cc*}$. According to Definition 7.11, it suffices to show that, for each $i \in \{1, \ldots, n_x\}$ and any $Z^*, Z \subseteq \Xi^B$ such that $Z^* \subseteq Z$,

$$\begin{aligned}
\check{f}_i^{L}(t, \boldsymbol{p}, Z) &\leq \check{f}_i^{cv*}(t, \boldsymbol{p}, Z^*, \Xi^B),\\
\check{f}_i^{U}(t, \boldsymbol{p}, Z) &\geq \check{f}_i^{cc*}(t, \boldsymbol{p}, Z^*, \Xi^B).
\end{aligned}$$

(7.16a) shows that

$$\begin{aligned}\check{f}_i^{cv*}(t,\boldsymbol{p},Z^*,\Xi^B) &= \max\left\{\check{f}_i^{cv}(t,\boldsymbol{p},Z^*,\Xi^B),\ \check{f}_i^{L}(t,\boldsymbol{p},Z^*)\right\}\\ &\geq \check{f}_i^{L}(t,\boldsymbol{p},Z^*)\\ &\geq \check{f}_i^{L}(t,\boldsymbol{p},Z),\end{aligned}$$

which verifies the first inequality. Similar arguments hold for the second inequality, □

The formulation in (7.16) can be extended so that extra $\check{\boldsymbol{f}}^{cv}$, $\check{\boldsymbol{f}}^{cc}$ functions can be used in the max and min functions. For example,

$$\begin{aligned}\check{f}_i^{cv*}(t,\boldsymbol{p},\Xi^R,\Xi^B) := \max\{&\check{f}_i^{cv}(t,\boldsymbol{p},\Xi^R,\Xi^B),\check{f}_i^{L}(t,\boldsymbol{p},\Xi^R),\\ &\check{f}_i^{cv\dagger}(t,\boldsymbol{p},\Xi^R,\Xi^B)\},\\ \check{f}_i^{cc*}(t,\boldsymbol{p},\Xi^R,\Xi^B) := \min\{&\check{f}_i^{cc}(t,\boldsymbol{p},\Xi^R,\Xi^B),\check{f}_i^{U}(t,\boldsymbol{p},\Xi^R),\\ &\check{f}_i^{cc\dagger}(t,\boldsymbol{p},\Xi^R,\Xi^B)\},\end{aligned}$$

where $\check{\boldsymbol{f}}^{cv\dagger}$, $\check{\boldsymbol{f}}^{cc\dagger}$ satisfy Assumption 7.6. This result can be verified similarly as the lemma above. The benefit of using multiple $\check{\boldsymbol{f}}^{cv}$, $\check{\boldsymbol{f}}^{cc}$ functions is to construct tighter state relaxation RHS functions, and therefore tighter state relaxations.

Next, we show that if $\check{\boldsymbol{f}}^{cv}$, $\check{\boldsymbol{f}}^{cc}$ and $\check{\boldsymbol{f}}^{L}$, $\check{\boldsymbol{f}}^{U}$ in (7.16) are smooth, then the generated state relaxations are smooth. Observe that in max-landing, we use $\check{\boldsymbol{f}}^{cv*}$, $\check{\boldsymbol{f}}^{cc*}$ to replace $\check{\boldsymbol{f}}^{cv}$, $\check{\boldsymbol{f}}^{cc}$ in (7.13). So, consider $\boldsymbol{f}^{cv*}$, $\boldsymbol{f}^{cc*} : I \times P \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ such

that, for each $i \in \{1,\ldots,n_x\}$,

$$\begin{aligned}
&f_i^{cv*}(t,\boldsymbol{p},\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc})\\
&\quad:= \check{f}_i^{cv*}(t,\boldsymbol{p},B_i^L(\Xi^R),X^B(t))\\
&\quad= \max\left\{\check{f}_i^{cv}(t,\boldsymbol{p},B_i^L(\Xi^R),X^B(t)),\check{f}_i^{L}(t,\boldsymbol{p},B_i^L(\Xi^R))\right\},\\
&\quad= \max\left\{f_i^{cv}(t,\boldsymbol{p},\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc}),f_i^{L}(t,\boldsymbol{p},\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc})\right\},\\
&f_i^{cc*}(t,\boldsymbol{p},\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc})\\
&\quad:= \check{f}_i^{cc*}(t,\boldsymbol{p},B_i^U(\Xi^R),X^B(t))\\
&\quad= \min\left\{\check{f}_i^{cc}(t,\boldsymbol{p},B_i^U(\Xi^R),X^B(t)),\check{f}_i^{U}(t,\boldsymbol{p},B_i^U(\Xi^R))\right\},\\
&\quad= \min\left\{f_i^{cc}(t,\boldsymbol{p},\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc}),f_i^{U}(t,\boldsymbol{p},\boldsymbol{\xi}^{cv},\boldsymbol{\xi}^{cc})\right\}.
\end{aligned}$$

Then, $\boldsymbol{f}^{cv*}$, $\boldsymbol{f}^{cc*}$ will replace $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ in the RHS of (7.5b).

Assumption 7.12. *Assume the following holds.*

1. $\check{\boldsymbol{f}}^L(t,\cdot,\cdot,\cdot)$, $\check{\boldsymbol{f}}^U(t,\cdot,\cdot,\cdot)$ *are differentiable on* $P\times\Xi^B\times\Xi^B$ *for any* $t\in I$, $\Xi^B\in\mathbb{IR}^{n_x}$.

2. *Let* $(\boldsymbol{x}^L,\boldsymbol{x}^U,\boldsymbol{x}^{cv},\boldsymbol{x}^{cc})$ *be a solution of* (7.5) *on* $I\times P$. *For all* $(t,\boldsymbol{p})\in I\times P$ *and* $i\in\{1,\ldots,n_x\}$,

   (a) *If* $f_i^{cv}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc}) = f_i^L(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc})$, *then*

$$\begin{aligned}
&\frac{\partial f_i^{cv}}{\partial \boldsymbol{x}^{cv}}\boldsymbol{f}^{cv*}+\frac{\partial f_i^{cv}}{\partial \boldsymbol{x}^{cc}}\boldsymbol{f}^{cc*}+\frac{\partial f_i^{cv}}{\partial t}\\
&\quad\neq \frac{\partial f_i^{L}}{\partial \boldsymbol{x}^{cv}}\boldsymbol{f}^{cv*}+\frac{\partial f_i^{L}}{\partial \boldsymbol{x}^{cc}}\boldsymbol{f}^{cc*}+\frac{\partial f_i^{L}}{\partial t},
\end{aligned}$$

(b) If $f_i^{cc}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc}) = f_i^{U}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc})$*, then*

$$\frac{\partial f_i^{cc}}{\partial \boldsymbol{x}^{cv}}\boldsymbol{f}^{cv*} + \frac{\partial f_i^{cc}}{\partial \boldsymbol{x}^{cc}}\boldsymbol{f}^{cc*} + \frac{\partial f_i^{cc}}{\partial t} \neq \frac{\partial f_i^{U}}{\partial \boldsymbol{x}^{cv}}\boldsymbol{f}^{cv*} + \frac{\partial f_i^{U}}{\partial \boldsymbol{x}^{cc}}\boldsymbol{f}^{cc*} + \frac{\partial f_i^{U}}{\partial t},$$

where arguments $(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc})$ *are omitted for simplicity.*

Lemma 7.9. *Under Assumptions 7.1, 7.5, 7.6, 7.8, 7.11, and 7.12, let* $(\boldsymbol{x}^{L},\boldsymbol{x}^{U},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc})$ *be a solution of* (7.5) *on* $I \times P$*. Consider any* $\boldsymbol{p} \in P$*,* $m \in \{1,\ldots,n_p\}$*, and* $i \in \{1,\ldots,n_x\}$*. Then, the partial derivative* $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}$ *exists at every* $t \in I$ *such that* $f_i^{cv}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) = f_i^{L}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p}))$*, and partial derivative* $\frac{\partial \boldsymbol{x}^{cc}}{\partial p_m}$ *exist at every* $t \in I$ *such that* $f_i^{cc}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) = f_i^{U}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p}))$*.*

Proof. Consider any $\boldsymbol{p} \in P$, $m \in \{1,\ldots,n_p\}$, and $i \in \{1,\ldots,n_x\}$. We will show that $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}$ exists at every t such that $f_i^{cv}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) = f_i^{L}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p}))$. It is analogous to show the second result.

According to Assumptions 7.8 and 7.12, the partial derivatives

$$\frac{\partial f_i^{cv}}{\partial \boldsymbol{x}^{cv}}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc}), \quad \frac{\partial f_i^{cv}}{\partial p_m}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc}),$$
$$\frac{\partial f_i^{L}}{\partial \boldsymbol{x}^{cv}}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc}), \quad \frac{\partial f_i^{L}}{\partial p_m}(t,\boldsymbol{p},\boldsymbol{x}^{cv},\boldsymbol{x}^{cc})$$

exist and are continuous in a neighborhood of $x_i^{cv}(t,\boldsymbol{p})$ and $x_i^{cc}(t,\boldsymbol{p})$. When f_i^{cv*} changes between f_i^{cv} and f_i^{L}, we consider the hybrid system to be moving from some mode s_j to some mode s_{j+1}. Assumption 7.12 ensures that the Jacobian matrix corresponding to $\boldsymbol{x}_j^{cv}, \boldsymbol{x}_j^{cc}, \boldsymbol{x}_{j+1}^{cv}, \boldsymbol{x}_{j+1}^{cc}, t$ is invertible. According to Theorem 1

and Remark 6 in [57], the sensitivities in mode s_j to mode s_{j+1} are equal such that

$$\frac{\partial x_{i,j}^{cv}}{\partial p_m}(t,\boldsymbol{p}) = \frac{\partial x_{i,j+1}^{cv}}{\partial p_m}(t,\boldsymbol{p}).$$

So, partial derivative $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}(t,\boldsymbol{p})$ exists at every t such that $f_i^{cv}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p})) = f_i^{L}(t,\boldsymbol{p},\boldsymbol{x}^{cv}(t,\boldsymbol{p}),\boldsymbol{x}^{cc}(t,\boldsymbol{p}))$. □

Theorem 7.5. *Under Assumptions 7.1, 7.5, 7.6, 7.8, 7.11, and 7.12, let $\boldsymbol{f}^{cv*}, \boldsymbol{f}^{cc*}$ be the RHS of* (7.5b) *and let $(\boldsymbol{x}^L, \boldsymbol{x}^U, \boldsymbol{x}^{cv}, \boldsymbol{x}^{cc})$ be a solution of* (7.5) *on $I \times P$. For any $m \in \{1,\ldots,n_p\}$, the partial derivatives $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}$ and $\frac{\partial \boldsymbol{x}^{cc}}{\partial p_m}$ exist and are continuous in $I \times P$.*

Proof. Consider any $m \in \{1,\ldots,n_p\}$, and $\boldsymbol{p} \in P$. Let $[s_1, s_2, \ldots, s_j, \ldots, s_{n_j}]$ be the ordered succession of modes a hybrid system described by (7.5). At each event time $t \in I$ between two successive modes, Lemma 7.9 ensures that $\frac{\partial x_i^{cv}}{\partial p_m}(t,\boldsymbol{p})$ or $\frac{\partial x_i^{cc}}{\partial p_m}(t,\boldsymbol{p})$ exists, and respectively

$$\frac{\partial x_{i,j}^{cv}}{\partial p_m}(t,\boldsymbol{p}) = \frac{\partial x_{i,j+1}^{cv}}{\partial p_m}(t,\boldsymbol{p}) \quad \text{or} \quad \frac{\partial x_{i,j}^{cc}}{\partial p_m}(t,\boldsymbol{p}) = \frac{\partial x_{i,j+1}^{cc}}{\partial p_m}(t,\boldsymbol{p}).$$

Next, we consider each mode as an individual ODE system. The RHS function of the ODE (7.5b) is either $\boldsymbol{f}^{cv}$ and $\boldsymbol{f}^{L}$, and the initial condition is the terminal state of the previous mode. Since $\check{\boldsymbol{f}}^{cv}(t,\cdot,\cdot,\cdot)$, $\check{\boldsymbol{f}}^{cc}(t,\cdot,\cdot,\cdot)$ and $\check{\boldsymbol{f}}^{L}(t,\cdot,\cdot,\cdot)$, $\check{\boldsymbol{f}}^{U}(t,\cdot,\cdot,\cdot)$ are differentiable, flattening operations B_i^L, B_i^U ensure that $\boldsymbol{f}^{cv}(t,\cdot,\cdot,\cdot)$, $\boldsymbol{f}^{cc}(t,\cdot,\cdot,\cdot)$ and $\boldsymbol{f}^{L}(t,\cdot,\cdot,\cdot)$, $\boldsymbol{f}^{U}(t,\cdot,\cdot,\cdot)$ are also differentiable. Then, [60, Theorem 3.1] ensures the existence and continuity of $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}$ and $\frac{\partial \boldsymbol{x}^{cc}}{\partial p_m}$ within each mode. Therefore, $\frac{\partial \boldsymbol{x}^{cv}}{\partial p_m}$ and $\frac{\partial \boldsymbol{x}^{cc}}{\partial p_m}$ exist and are continuous on $I \times P$. □

7.7.3 Safe-landing

In the last scenario, we suppose that neither Assumption 7.10 or Assumption 7.11 holds. The safe-landing strategy introduced in [27] can be applied here. Its formulation is as follows:for all $(t, \boldsymbol{p}) \in I \times P$ and $\Xi^R \in X^B(t)$,

$$\check{f}_i^{cv}(t, \boldsymbol{p}, \Xi^R, X^B(t)) := \max\left\{\check{f}_i^{cv}(t, \boldsymbol{p}, \Xi^R, X^B(t)), \dot{x}_i^L(t) - \sqrt{2\underline{k}_i(\check{\zeta}_i^{cv} - x_i^L(t))}\right\}, \tag{7.18a}$$

$$\check{f}_i^{cc}(t, \boldsymbol{p}, \Xi^R, X^B(t)) := \min\left\{\check{f}_i^{cc}(t, \boldsymbol{p}, \Xi^R, X^B(t)), \dot{x}_i^U(t) + \sqrt{2\overline{k}_i(x_i^U(t) - \check{\zeta}_i^{cc})}\right\}, \tag{7.18b}$$

where $\underline{\boldsymbol{k}}, \overline{\boldsymbol{k}} \in \mathbb{R}^{n_x}_{>0}$ are safe-landing constants [27, Definition 7] determined by $\dot{\boldsymbol{x}}^L, \dot{\boldsymbol{x}}^U$, and $\dot{\boldsymbol{x}}$.

Using similar approaches as in [27], it has been readily verified that $\boldsymbol{f}^{cv}, \boldsymbol{f}^{cc}$, constructed with flattened $\check{\boldsymbol{f}}^{cv}, \check{\boldsymbol{f}}^{cc}$ from (7.18), describe enclosing dynamics and convexity-preserving dynamics. Note that $\check{\boldsymbol{f}}^{cv}, \check{\boldsymbol{f}}^{cc}$ in (7.18) are not Lipschitz continuous with respect to state variables, but one-sided Lipschitz continuous, due to those square root functions. But the existence theorem and unique theorem in [148] ensure that (7.5b) always has a unique solution under this circumstance.

7.8 Numerical Examples

In this section, we use the new framework to construct state relaxations for parametric ODEs. A proof-of-concept implementation was developed in Julia [20]. DifferentialEquations.jl [104] was used as the ODE solver, and GMC and DMC were

computed with McCormick.jl [134]. The numerical results reported below were obtained by running this implementation on a Windows 10 machine with a 3.6 GHz AMD Ryzen 5 2600X CPU and 8GB memory.

The first example is adapted from [31]. It involves a simple ODE system with a quadratic RHS. We will use it to illustrate constructing state relaxations using our new framework (7.5) and compare the generated state relaxations with established methods.

Example 7.1. *Consider the quadratic ODEs:*

$$\begin{aligned} \dot{x}_1(t,\boldsymbol{p}) &= (x_1-p_1)^2-(x_2-p_1)^2,\ x_1(t_0)=2.2,\\ \dot{x}_2(t,\boldsymbol{p}) &= (x_1-p_2)^2-(x_2-p_2)^2,\ x_2(t_0)=1.8, \end{aligned} \tag{7.19}$$

where $\boldsymbol{p}=(p_1,p_2)\in P\equiv[-2,2]\times[-1,3]$, *and* $I\equiv[t_0,t_f]=[0.0,0.16]$.

To start with, we choose $\check{\boldsymbol{f}}^L$, $\check{\boldsymbol{f}}^U$ and $\check{\boldsymbol{f}}^{cv}$, $\check{\boldsymbol{f}}^{cc}$ to construct $\boldsymbol{f}^L$, $\boldsymbol{f}^U$ and $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ as in (7.9) and (7.13), respectively. According to Section 7.7.1, we may choose them conveniently using Table 7.1 following steps i-iii. Firstly, we selected p-N-G from Category II to construct $\check{F}^R$. Secondly, we selected ○-N-G from Category IV which has higher or equivalent scores than p-N-G. Since both p-N-G and ○-N-G use nonlinear relaxations, the last step can be ignored. After substituting $\boldsymbol{f}^L$, $\boldsymbol{f}^U$ and $\boldsymbol{f}^{cv}$, $\boldsymbol{f}^{cc}$ into (7.5), we obtained a method, p-N-G w/ ○-N-G, for constructing state relaxations for (7.1) using the new framework. To the best of the authors' knowledge, this method has never been reported in literature, and is therefore a newly discovered method using our new framework. Following similar steps, we also obtained another new method, p-N-D w/ p-A-D, using Table 7.1. Note that in this method, $\check{\boldsymbol{f}}^L$, $\check{\boldsymbol{f}}^U$ are affine approximations of DMC linearized at the middle

point of P, so that the optimization problem in the RHS of (7.5a) can be solved trivially without a numerical optimization solver.

Figure 7.1 illustrates the state relaxations generated with the two new state relaxation methods, as well as the Scott-Barton method [120]. Compared with the Scott-Barton method, the first new method, p-N-G w/ ○-N-G, constructed looser state relaxations in this example. However, according to Table 7.2, it required less computing time. This may due to the fact that our new framework eliminates the discrete jumps, while Scott-Barton method requires continuously checking if-statements in (7.3) using event detection feature during integration which may be computationally expensive. The second method, p-N-D w/ p-A-D, generated state relaxations are tighter than the first method, but the computing time is longer. Compared with the Scott-Barton method, p-N-D w/ p-A-D generated significantly tighter concave relaxation and similar convex relaxation. Note that the DMC implemented in [152] is from [72], which is not twice-continuously differentiable. Therefore, the uniqueness of a solution of (7.5a) is not guaranteed according to [28]. Here, we use this implementation for demonstration and assume that there exists only one solution.

Figure 7.2 illustrates another two new methods for constructing state relaxations of (7.1) using the new framework, and they both depend on αBB relaxations [3]. The first new method, px-N-α w/ px-N-α, was discovered using Table 7.1. Since the αBB relaxations of quadratic functions are also quadratic [31], the embedded optimization in (7.5) are quadratic programs and can be solved with CPLEX. The second method, px-N-α&○-N-G w/ ○-N-G, was obtained following the max-landing method described in Section 7.7.2. $\check{\boldsymbol{f}}^{cv}$, $\check{\boldsymbol{f}}^{cc}$ in (7.16) were constructed with

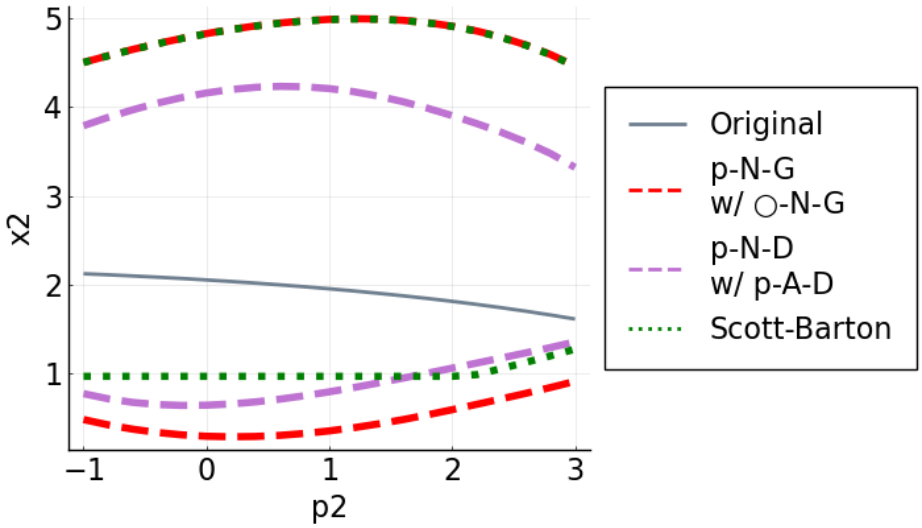


Figure 7.1: The parametric solution of x_2 in (7.19) (black solid), along with the convex and concave relaxations generated with the Scott-Barton method (green dotted) and our new methods (violet dashed and red dashed) with $p_1 = 0$ at $t_f = 0.15s$

.

αBB relaxations and $\check{\boldsymbol{f}}^L, \check{\boldsymbol{f}}^U$ were constructed with GMC. Observe that the state relaxations constructed with both new methods are significantly tighter the Scott-Barton method. Moreover, method px-N-α&○-N-G w/ ○-N-G constructed tighter concave relaxations than the method by Song and Khan [131].

Figure 7.3 shows smooth state relaxations constructed using new method p-N-D w/ ○-N-D. These state relaxations are differentiable with respect to parameters. Their gradients were evaluated using a subgradient computation method from [132] and were plotted as tangents in Figure 7.3.

The second example considers a biochemical process adapted from [16]. Convex enclosures of this system have been obtained in [83, 31].

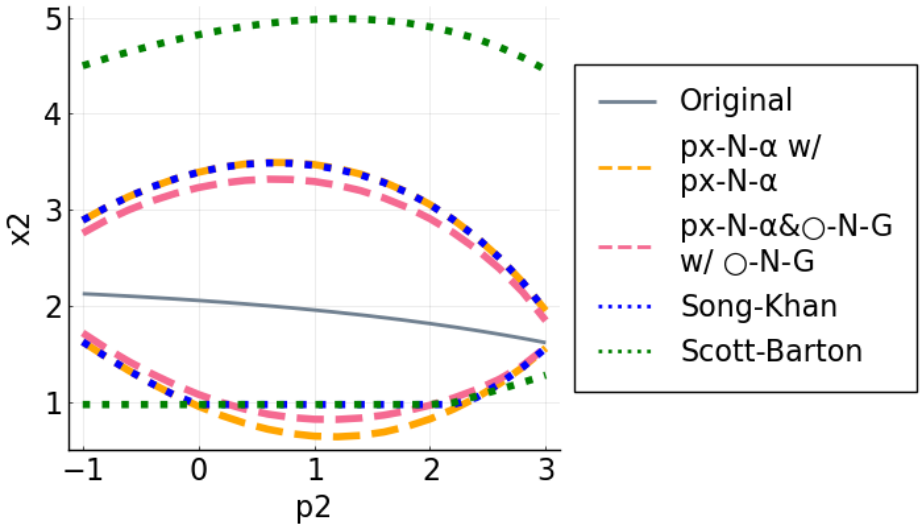


Figure 7.2: The parametric solution of x_2 in (7.19) (black solid), along with the convex and concave relaxations generated with Scott-Barton method [120] (green dotted) and Song-Khan method [131] (blue dotted) and our new methods (pink dashed and orange dashed) with $p_1 = 0$ at $t_f = 0.15s$

.

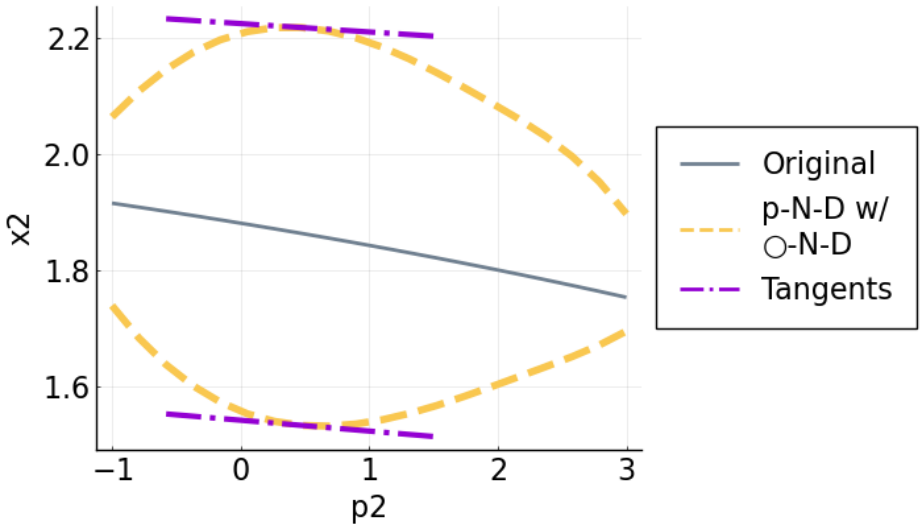


Figure 7.3: The parametric solution of x_2 in (7.19) (black solid), along with the differentiable convex/concave relaxations (yellow dashed) and their tangents (purple dash-dot), with $p_1 = 0$ at $t_f = 0.05s$

Table 7.2: Computing times of methods shown in Figures 7.1 and 7.2

State relaxation	State bound	Label	CPU time [1]
GMC	GMC	p-N-G w/ ○-N-G	0.0035
DMC	Affine DMC	p-N-D w/ p-A-D	0.0538
GMC	NIE	Scott-Barton	0.0104
αBB	αBB	px-N-α w/ px-N-α	4.0869
αBB and GMC	GMC	px-N-α & ○-N-G w/ ○-N-G	3.3831
αBB	NIE	Song-Khan	2.2989

1 Each number is the average of 20 runs

Example 7.2. *Consider a microbial growth process described by the following ODE system:*

$$\begin{aligned} \dot{x}_1(t) &= (\mu - \alpha D)x_1, & x_1(t_0) &= 0.82, \\ \dot{x}_2(t) &= D(S^i - x_2) - k\mu x_1, & x_2(t_0) &= 0.8, \end{aligned} \tag{7.20}$$

where state variables x_1 and x_2 respectively represent the concentrations of biomass and substrate, $\boldsymbol{p} = (K_I, K_S)$ are uncertain kinetic parameters, $I \equiv [t_0, t_f] = [0, 5]$, and μ is the growth rate

$$\mu = \frac{\mu_m x_2}{K_S + x_2 + K_I x_2^2}.$$

The remaining quantities are parameters, whose values and uncertainties are provided in Table 7.3.

Figure 7.4 presents two new state relaxations constructed for (7.2) and compares them with the Scott-Barton method. The first method, p-N-G w/ ○-N-G,

Table 7.3: Microbial growth process parameters

Parameter	Symbol	Value	Unit
Process heterogeneity	α	0.5	-
Dilution rate	D	0.36	day^{-1}
Influent concentration	S^i	5.7	g S/l
Yield coefficient	k	10.53	g S/g X
Max growth rate	μ_m	1.2	day^{-1}
Kinetic parameter	K_I	[0.4, 0.6]	$(\text{g S/l})^{-1}$
Kinetic parameter	K_S	[7.0, 7.2]	g S/l

generates state relaxations that overlap with Scott-Barton relaxations. But its computational cost is slightly less than Scott-Barton method, probably due to elimination of discrete jumps. The other method, p-N-G w/ px-N-G, utilizes the optimization-based state relaxation method introduced in Section 7.6.2 where $\hat{\boldsymbol{f}}^{cv}$, $\hat{\boldsymbol{f}}^{cc}$ are generated with GMC. The embedded optimization problems were solved with IPOPT. Compared with the Scott-Barton method, though the overall computing process of the second new method takes a longer time according to Table 7.3, the constructed state relaxations are significantly tighter.

Table 7.4: Computing times of methods shown in Figures 7.4 of Example 7.2

State relaxation	State bound	Label	CPU time [2]
GMC	GMC	p-N-G w/ ○-N-G	0.0076
GMC	GMC	p-N-G w/ px-N-G	74.7005
GMC	NIE	Scott-Barton	0.0106

2 Each number is the average of 20 runs

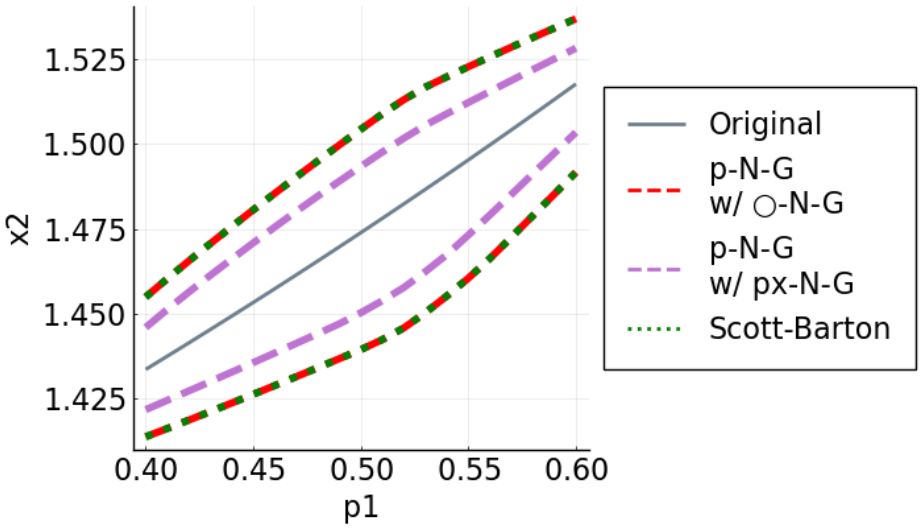


Figure 7.4: The parametric solution of x_2 in (7.20) (black solid), along with the convex and concave relaxations generated with the Scott-Barton method [120] (green dotted) and our new methods (violet dashed and red dashed) with $K_S = 7.1$ at $t_f = 5s$

.

7.9 Conclusion

In this article, we proposed a general framework for computing convex enclosures for nonlinear parametric ODEs (7.1) using differential inequalities. Componentwise convex and concave relaxations of the original state variables were obtained by constructing and solving an auxiliary system of ODEs (7.5). Unlike Scott and Barton's framework [120] which contains discrete jumps in the auxiliary ODE RHS (7.3), our new framework employs continuous ODEs. This modification not only makes the auxiliary system (7.5) easier to solve numerically than (7.3), but also permits subgradient evaluation for the generated state relaxations [132]. Subgradients are useful sensitivity information for local optimization solvers. They also can be used to generate computationally cheap outer approximations of nonlinear

convex relaxations, which are favorable in deterministic global optimization [33, 152]. Moreover, this new framework is versatile. Section 7.6 introduced various methods for constructing the auxiliary RHS functions in (7.5), including both established methods [120, 131] and newly discovered methods. They are all summarized in Table 7.1. Some of the new methods lead to tighter convex relaxations than the established methods. To demonstrate this, we developed a proof-of-concept implementation of the new framework and presented multiple numerical examples in Section 7.8. Tighter convex relaxations supply more accurate quantification on the influence of uncertainty to the original system. They also facilitate branch-and-bound algorithms to converge faster in deterministic global optimization [42].

Future work may involve using these new convex relaxations of parametric ODEs in deterministic global dynamic optimization. Compare with using established relaxations from [120, 131], we expect to see an improvement in computational performance, since our new relaxations are tighter and their subgradients are available.

Chapter 8

Concluding Remarks

8.1 Conclusions

In this thesis, novel formulations and supporting theory have been developed to automatically construct improved bounds for implicit functions and dynamic systems in order to improve the computational performance of deterministic global optimization algorithms.

Chapter 2 presented a new approach for constructing convex relaxations for implicit functions using parametric programming. Unlike state-of-the-art approaches [136, 151], this new approach does not assume the uniqueness of the implicit function and does not require the original residual function to be factorable. The validity of these convex relaxations was verified and a proof-of-concept implementation of this new approach was developed. Multiple numerical examples illustrated that this new approach constructed tighter convex relaxations than established methods. This new approach was also extended to generate convex relaxations for the

numerical solutions of parametric ODEs obtained with implicit integration methods.

Chapters 5 and 6 presented an optimization-based framework for constructing time-varying interval bounds of ODEs using differential inequalities. This framework includes several established bounding approaches, but also includes many new approaches. Some of these new approaches were implemented, and they generated tighter ODE bounds than established methods in many numerical examples. These tighter interval bounds are useful in constructing tighter convex relaxation for ODEs [131]. Complementarity reformulations of these approaches were also provided, so that tighter ODE bounds may be computed without adding significantly to the cost of bound evaluation.

Chapters 3 and 7 introduced two new approaches for generating smooth convex relaxations of the parametric ODEs. They both improved the Scott-Barton method [120] by eliminating the discrete jumps in the bounding ODEs. To achieve this, Chapter 3 developed a novel smoothing technique, and Chapter 7 adapted the new optimization-based framework from Chapter 5. Both new approaches were demonstrated to construct convex relaxations that are at least as tight as the Scott-Barton method. Under additional conditions, both approaches can generate differentiable relaxations. Such tightness, smoothness, and availability of gradients are beneficial to the application of these ODE relaxations in branch-and-bound algorithms for deterministic global dynamic optimization.

Chapter 4 presented a new approach for generating guaranteed lower bounds of a nonconvex optimal control problem (OCP) by constructing a relaxed OCP. The optimal solution value of this relaxed OCP was verified that to be a lower bound of

the optimal solution value of the original OCP, which is useful in the deterministic global optimization of the original OCP [119, 64]. A two-point boundary-value problem was developed and implemented to solve the relaxed OCP to global optimality using the Pontryagin's Minimum Principle conditions. Numerical examples illustrated that new bounds are much tighter than established relaxations, and theoretical results support this.

8.2 Outlook

The ultimate goal of of this line of research is to develop and implement computationally efficient methods for deterministic global dynamic optimization, and use them to solve real-world engineering problems. The bottleneck is the construction of tight and smooth convex relaxations for dynamic systems for use in branch-and-bound algorithms [112, 120]. The overall approach of this thesis uses differential inequalities to construct convex relaxations for ODEs [122, 120, 131]. These convex relaxations are described by bounding systems of ODEs, whose RHS functions are relaxations of the original ODE RHS function. However, this overall approach can be vulnerable to the wrapping effect [59] and similar effects, which might lead to conservative bounds or even make the bounding systems "explode" before reaching the terminal time. Various formulations and techniques have been developed to address this problem [59, 128, 120, 131], including in this thesis. However, this is still one of the main obstacles of generating tight enclosures for ODEs using differential inequalities and applying them in deterministic global dynamic optimization.

The developments of this thesis suggest two avenues for research in the immediate future to further improve these differential inequality-based approaches. First, the new theories presented in Chapter 5 and 7, as well as recent developments in [130], imply that tighter relaxations of the original ODE RHS function will lead to tighter relaxations of the ODE solutions. Therefore, tighter convex relaxation techniques of general nonlinear functions are useful for reducing the wrapping effect. Second, recent research by Scott et al. [118, 126, 127, 124] constructs tighter time-varying interval bounds of ODEs with some *a priori* knowledge of the original system, such as conservative laws, physical bounds, constraints in an optimization problem, and algebraic equations in DAEs. Nevertheless, to the best of the author's knowledge, this strategy has not been applied to refine the convex relaxations of parametric ODEs. This might be a potential approach to generate tight convex relaxations for ODEs and improve the computational efficiency of deterministic global dynamic optimization.

Chapters 3 and 7 in this thesis provided two new approaches that generate smooth convex relaxations for parametric ODEs by eliminating the discrete jumps in Scott and Barton's bounding system of ODEs [120]. This removed the discontinuity obstacles in evaluating gradients for these convex relaxations. However, additional techniques must also be implemented, such as forward and adjoint sensitivity analysis, to compute the gradients of these bounding ODEs, so that these gradients can be used for minimizing convex relaxations in deterministic global dynamic optimization. Compared with the traditional forward sensitivity analysis, adjoint sensitivity analysis [35] is typically much faster at computing gradients for smooth ODEs with many parameters. Developing and implementing adjoint

sensitivity analysis for ODE relaxations may significantly improve the computational performance of branch-and-bound algorithms and increase the size of dynamic optimization problems that can be solved to global optimality in a reasonable time. However, there are currently no established adjoint sensitivity solvers for smooth ODEs in Julia, as it is a relatively new language. Once these are established, then we expect that adjoint sensitivity analysis may be applied to the various dynamic relaxations of this thesis.

The work underlying this thesis also included the development of a Julia package for deterministic global dynamic optimization, requiring minimal user input, and generating the various relaxations completely automatically using operator overloading. This implementation has successfully solved several benchmark problems and chemical engineering cases studies from literature to global optimality. Nevertheless, to make this implementation applicable for large-scale engineering problems, we need to further improve its computational efficiency. One possible approach is to exploit any sparsity in the original system (1.1), and to harness parallel computing techniques effectively. Furthermore, to facilitate the on-going research in this area, a representative library of benchmark problems ought to be collected for assessing and comparing the performance of algorithms for reachability analysis and global dynamic optimization. Such libraries have been constructed for static global optimization. The COCONUT library [114] contains over 1000 benchmark problems and has been used to compare state-of-the-art global optimization solvers, including BARON [139] and ANTIGONE [90]. The development of such a benchmark library can provide a thorough evaluation framework for the performance of global dynamic optimization algorithms and provide fair

comparisons between different algorithms. Such a library would also help to identify the kind of problems that are amenable to solution by one particular algorithm, in the spirit of the "no free lunch" theorem [154].

Bibliography

[1] V. Acary and F. Pérignon. "Siconos: A Software Platform for Modeling, Simulation, Analysis and Control of Nonsmooth Dynamical Systems". In: *Simulation News Europe* 17.3/4 (2007), pp. 19–26.

[2] B. Açıkmeşe and L. Blackmore. "Lossless Convexification of a Class of Optimal Control Problems with Non-Convex Control Constraints". In: *Automatica* 47.2 (Feb. 2011), pp. 341–347.

[3] C. Adjiman, S. Dallwig, C. Floudas, and A. Neumaier. "A Global Optimization Method, αBB, for General Twice-Differentiable Constrained NLPs – I. Theoretical Advances". In: *Computers & Chemical Engineering* 22.9 (1998), pp. 1137–1158.

[4] T. Alamo, D. Limon, E. Camacho, and J. Bravo. "Robust MPC of Constrained Nonlinear Systems Based on Interval Arithmetic". In: *IEE Proceedings - Control Theory and Applications* 152.3 (May 2005), pp. 325–332.

[5] M. Althoff. "An Introduction to CORA 2015". In: *ARCH14-15. 1st and 2nd International Workshop on Applied veRification for Continuous and Hybrid Systems*. 2015, pp. 120–87.

[6] M. Althoff and J. M. Dolan. "Online Verification of Automated Road Vehicles Using Reachability Analysis". In: *IEEE Transactions on Robotics* 30.4 (Aug. 2014), pp. 903–918.

[7] M. Althoff and B. H. Krogh. "Reachability Analysis of Nonlinear Differential-Algebraic Systems". In: *IEEE Transactions on Automatic Control* 59.2 (Feb. 2014), pp. 371–383.

[8] 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*. Cancun, Mexico: IEEE, 2008, pp. 4042–4048.

[9] I. P. Androulakis, C. D. Maranas, and C. A. Floudas. "aBB: A Global Optimization Method for General Constrained Nonconvex Problems". In: *Journal of Global Optimization* 7.4 (1995), pp. 337–363.

[10] R. Angira and A. Santosh. "Optimization of Dynamic Systems: A Trigonometric Differential Evolution Approach". In: *Computers & Chemical Engineering* 31.9 (Sept. 2007), pp. 1055–1063.

[11] E. Asarin, T. Dang, and A. Girard. "Reachability Analysis of Nonlinear Systems Using Conservative Approximation". In: *Hybrid Systems: Computation and Control*. Vol. 2623. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 20–35.

[12] E. Asarin, T. Dang, G. Frehse, A. Girard, C. Le Guernic, and O. Maler. "Recent Progress in Continuous and Hybrid Reachability Analysis". In: *2006*

IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control. Munich, Germany: IEEE, Oct. 2006, pp. 1582–1587.

[13] V. Azhmyakov and J. Raisch. "Convex Control Systems and Convex Optimal Control Problems with Constraints". In: *IEEE Transactions on Automatic Control* 53.4 (May 2008), pp. 993–998.

[14] J. R. Banga, C. G. Moles, and A. A. Alonso. "Global Optimization of Bioprocesses Using Stochastic and Hybrid Methods". In: *Frontiers in Global Optimization*. Vol. 74. Boston, MA: Springer US, 2004, pp. 45–70.

[15] J. R. Banga and W. D. Seider. "Global Optimization of Chemical Processes Using Stochastic Algorithms". In: *In "State of the Art in Global Optimization", CA Floudas and PM Pardalos (Eds*. Kluwer Academic Pub, 1996, pp. 563–583.

[16] G. Bastin and D. Dochain. *On-Line Estimation and Adaptive Control of Bioreactors*. Elsevier, 1990.

[17] R. Bellman. *Dynamic Programming*. Reprint edition. Dover Publications, Apr. 2013.

[18] 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". In: *Biotechnology and Bioengineering* 75.4 (2001), p. 15.

[19] D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Belmont, Mass: Athena Scientific, 1995.

[20] J. Bezanson, A. Edelman, S. Karpinski, and V. Shah. "Julia: A Fresh Approach to Numerical Computing". In: *SIAM Review* 59.1 (Jan. 2017), pp. 65–98.

[21] L. T. Biegler, A. M. Cervantes, and A. Wächter. "Advances in Simultaneous Strategies for Dynamic Process Optimization". In: *Chemical Engineering Science* 57.4 (Feb. 2002), pp. 575–593.

[22] L. Blackmore, B. Açıkmeşe, and J. M. Carson. "Lossless Convexification of Control Constraints for a Class of Nonlinear Optimal Control Problems". In: *Systems & Control Letters* 61.8 (Aug. 2012), pp. 863–870.

[23] H. Bock and K. Plitt. "A Multiple Shooting Algorithm for Direct Solution of Optimal Control Problems". In: *IFAC Proceedings Volumes* 17.2 (July 1984), pp. 1603–1608.

[24] A. Bompadre and A. Mitsos. "Convergence Rate of McCormick Relaxations". In: *Journal of Global Optimization* 52.1 (Jan. 2012), pp. 1–28.

[25] S. Boyd and L. Vandenberghe. *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[26] A. Bressan and B. Piccoli. *Introduction to the Mathematical Theory of Control*. First. Springfield, MO: American Institute of Mathematical Sciences, Aug. 2007.

[27] H. Cao and K. A. Khan. "A Smoothing Method for Generating Tighter Reachable Sets Enclosures of Parametric Ordinary Differential Equations". In: In preparation (2021).

[28] H. Cao and K. A. Khan. "An Optimization-Based Framework for Enclosing Reachable Sets Using Differential Inequalities". In: In preparation (2021).

[29] H. Cao and K. A. Khan. "Bounding Nonconvex Optimal Control Problems Using Pontryagin's Minimum Principle". In: In preparation (2021).

[30] H. Cao and K. A. Khan. "Convex Relaxations of Implicit Functions". In: In preparation (2021).

[31] H. Cao and K. A. Khan. "Enclosing Reachable Sets for Nonlinear Control Systems Using Complementarity-Based Intervals". In: *IFAC-PapersOnLine*. 16th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM 2021 54.3 (Jan. 2021), pp. 590–595.

[32] H. Cao and K. A. Khan. "Improved Convex Relaxations for Global Dynamic Optimization". In: In preparation (2021).

[33] H. Cao, Y. Song, and K. A. Khan. "Convergence of Subtangent-Based Relaxations of Nonlinear Programs". In: *Processes* 7.4 (Apr. 2019), p. 221.

[34] Y. Cao, C. L. E. Swartz, J. Flores-Cerrillo, and J. Ma. "Dynamic Modeling and Collocation-Based Model Reduction of Cryogenic Air Separation Units". In: *AIChE Journal* 62.5 (May 2016), pp. 1602–1615.

[35] Y. Cao, S. Li, L. Petzold, and R. Serban. "Adjoint Sensitivity Analysis for Differential-Algebraic Equations: The Adjoint DAE System and Its Numerical Solution". In: *SIAM Journal on Scientific Computing* 24.3 (Jan. 2003), pp. 1076–1089.

[36] B. Chachuat. *MC++: A Toolkit for Bounding Factorable Functions*. 2014.

[37] B. Chachuat and M. Villanueva. "Bounding the Solutions of Parametric Odes: When Taylor Models Meet Differential Inequalities". In: *Computer Aided Chemical Engineering*. Vol. 30. Elsevier, 2012, pp. 1307–1311.

[38] X. Chen, E. Ábrahám, and S. Sankaranarayanan. "Flow*: An Analyzer for Non-Linear Hybrid Systems". In: *Computer Aided Verification*. Vol. 8044. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 258–263.

[39] A. Chutinan and B. H. Krogh. "Computational Techniques for Hybrid System Verification". In: *IEEE Transactions on Automatic Control* 48.1 (Jan. 2003), pp. 64–75.

[40] F. Clarke. *Optimization and Nonsmooth Analysis*. Classics in Applied Mathematics. SIAM, Jan. 1990.

[41] F. Clarke. *Functional Analysis, Calculus of Variations and Optimal Control*. Graduate Texts in Mathematics. London: Springer-Verlag, 2013.

[42] J. Clausen. *Branch and Bound Algorithms - Principles and Examples*. Tech. rep. Copenhagen, Denmark: University of Copenhagen, 1999, p. 30.

[43] E. A. Coddington and N. Levinson. *Theory of Ordinary Differential Equations*. McGraw-Hill, 1955.

[44] S. D. Cohen and A. C. Hindmarsh. "CVODE, a Stiff/Nonstiff ODE Solver in C". In: *Computers in Physics* 10.2 (Mar. 1996), pp. 138–143.

[45] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The Linear Complementarity Problem*. SIAM, Jan. 1992.

[46] T. Dang and O. Maler. "Reachability Analysis via Face Lifting". In: *Hybrid Systems: Computation and Control*. Vol. 1386. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 96–109.

[47] A. L. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings: A View from Variational Analysis*. Second. Springer Series in Operations Research and Financial Engineering. New York: Springer-Verlag, 2014.

[48] K. Du and R. B. Kearfott. "The Cluster Problem in Multivariate Global Optimization". In: *Journal of Global Optimization* 5.3 (Oct. 1994), pp. 253–265.

[49] I. Dunning, J. Huchette, and M. Lubin. "JuMP: A Modeling Language for Mathematical Optimization". In: *SIAM Review* 59.2 (Jan. 2017), pp. 295–320.

[50] M. A. Duran and I. E. Grossmann. "An Outer-Approximation Algorithm for a Class of Mixed-Integer Nonlinear Programs". In: *Mathematical Programming* 36.3 (Oct. 1986), pp. 307–339.

[51] R. Faber, T. Jockenhövel, and G. Tsatsaronis. "Dynamic Optimization with Simulated Annealing". In: *Computers & Chemical Engineering* 29.2 (Jan. 2005), pp. 273–290.

[52] A. V. Fiacco and J. Kyparisis. "Convexity and Concavity Properties of the Optimal Value Function in Parametric Nonlinear Programming". In: *Journal of Optimization Theory and Applications* 48.1 (Jan. 1986), pp. 95–126.

[53] A. F. Filippov. *Differential Equations with Discontinuous Right-Hand Sides: Control Systems*. Mathematics and Its Applications. Springer Netherlands, 1988.

[54] A. Flores-Tlacuahuac, S. T. Moreno, and L. T. Biegler. "Global Optimization of Highly Nonlinear Dynamic Systems". In: *Industrial & Engineering Chemistry Research* 47.8 (Apr. 2008), pp. 2643–2655.

[55] C. A. Floudas, P. 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*. Nonconvex Optimization and Its Applications. Springer US, 1999.

[56] G. Frehse. "PHAVer: Algorithmic Verification of Hybrid Systems Past HyTech". In: *Hybrid Systems: Computation and Control, 2005*. Zurich, Switzerland, 2005, p. 273.

[57] S. Galán, W. F. Feehery, and P. I. Barton. "Parametric Sensitivity Functions for Hybrid Discrete/Continuous Systems". In: *Applied Numerical Mathematics* 31.1 (Sept. 1999), pp. 17–47.

[58] A. Griewank and A. Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. 2nd ed. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2008.

[59] G. W. Harrison. "Dynamic Models with Uncertain Parameters". In: *Proceedings of the 1st International Conference on Mathematical Modeling* 1 (1977), pp. 295–304.

[60] P. Hartman. *Ordinary Differential Equations*. Philadelphia, PA: SIAM, Jan. 2002.

[61] S. M. Harwood and P. I. Barton. "Affine Relaxations for the Solutions of Constrained Parametric Ordinary Differential Equations". In: *Optimal Control Applications and Methods* 39.2 (Mar. 2018), pp. 427–448.

[62] S. M. Harwood, J. K. Scott, and P. I. Barton. "Bounds on Reachable Sets Using Ordinary Differential Equations with Linear Programs Embedded". In: *IMA Journal of Mathematical Control and Information* 33.2 (June 2016), pp. 519–541.

[63] M. M. F. Hasan. "An Edge-Concave Underestimator for the Global Optimization of Twice-Differentiable Nonconvex Problems". In: *Journal of Global Optimization* 71.4 (Aug. 2018), pp. 735–752.

[64] B. Houska and B. Chachuat. "Branch-and-Lift Algorithm for Deterministic Global Optimization in Nonlinear Optimal Control". In: *Journal of Optimization Theory and Applications* 162.1 (July 2014), pp. 208–248.

[65] S. Hu and N. S. Papageorgiou. *Handbook of Multivalued Analysis: Volume I: Theory*. Mathematics and Its Applications. Springer US, 1997.

[66] H. Huang, C. S. Adjiman, and N. Shah. "Quantitative Framework for Reliable Safety Analysis". In: *AIChE Journal* 48.1 (Jan. 2002), pp. 78–96.

[67] D. G. Hull. "Conversion of Optimal Control Problems into Parameter Optimization Problems". In: *Journal of Guidance, Control, and Dynamics* 20.1 (Jan. 1997), pp. 57–60.

[68] W. R. Huster, D. Bongartz, and A. Mitsos. "Deterministic Global Optimization of the Design of a Geothermal Organic Rankine Cycle". In: *Energy Procedia*. 4th International Seminar on ORC Power SystemsSeptember 13-15th

2017POLITECNICO DI MILANOBOVISA CAMPUSMILANO, ITALY 129 (Sept. 2017), pp. 50–57.

[69] L. Jaulin. "Nonlinear Bounded-Error State Estimation of Continuous-Time Systems". In: *Automatica* 38.6 (June 2002), pp. 1079–1082.

[70] K. A. Khan. "Subtangent-Based Approaches for Dynamic Set Propagation". In: *2018 IEEE Conference on Decision and Control (CDC)*. Miami Beach, FL: IEEE, Dec. 2018, pp. 3050–3055.

[71] K. A. Khan. "Whitney Differentiability of Optimal-Value Functions for Bound-Constrained Convex Programming Problems". In: *Optimization* 68.2-3 (Mar. 2019), pp. 691–711.

[72] K. A. Khan, H. A. J. Watson, and P. I. Barton. "Differentiable McCormick Relaxations". In: *Journal of Global Optimization* 67.4 (Apr. 2017), pp. 687–729.

[73] K. A. Khan, M. Wilhelm, M. D. Stuber, H. Cao, H. A. J. Watson, and P. I. Barton. "Corrections to: Differentiable McCormick Relaxations". In: *Journal of Global Optimization* 70.3 (Mar. 2018), pp. 705–706.

[74] K. A. Khan. "Sensitivity Analysis for Nonsmooth Dynamic Systems". PhD thesis. Massachusetts Institute of Technology, Feb. 2015.

[75] M. Kieffer, E. Walter, and I. Simeonov. "Guaranteed Nonlinear Parameter Estimation for Continuous-Time Dynamical Models". In: *IFAC Proceedings Volumes* 39.1 (2006), pp. 843–848.

[76] E. Kofman, H. Haimovich, and M. M. Seron. "A Systematic Method to Obtain Ultimate Bounds for Perturbed Systems". In: *International Journal of Control* 80.2 (Feb. 2007), pp. 167–178.

[77] C. Kontoravdi, E. N. Pistikopoulos, and A. Mantalaris. "Systematic Development of Predictive Mathematical Models for Animal Cell Cultures". In: *Computers & Chemical Engineering* 34.8 (Aug. 2010), pp. 1192–1198.

[78] W. Kühn. "Rigorously Computed Orbits of Dynamical Systems without the Wrapping Effect". In: *Computing* 61.1 (1998), pp. 47–67.

[79] A. Kurzhanski and P. Varaiya. "On Ellipsoidal Techniques for Reachability Analysis. Part I: External Approximations". In: *Optimization Methods and Software* 17.2 (Jan. 2002), pp. 177–206.

[80] D. J. Lacks. "Real-time Optimization in Nonlinear Chemical Processes: Need for Global Optimizer". In: *AIChE Journal* 49.11 (2003), pp. 2980–2983.

[81] Y. Lin and M. A. Stadtherr. "Deterministic Global Optimization of Nonlinear Dynamic Systems". In: *AIChE Journal* 53.4 (Apr. 2007), pp. 866–875.

[82] Y. Lin and M. A. Stadtherr. "Fault Detection in Nonlinear Continuous-Time Systems with Uncertain Parameters". In: *AIChE Journal* 54.9 (Sept. 2008), pp. 2335–2345.

[83] Y. Lin and M. A. Stadtherr. "Validated Solution of ODEs with Parametric Uncertainties". In: *Computer Aided Chemical Engineering*. Vol. 21. 16th ESCAPE and 9th PSE Symposium. Elsevier, Jan. 2006, pp. 167–172.

[84] Y. Lin and M. A. Stadtherr. "Validated Solutions of Initial Value Problems for Parametric ODEs". In: *Applied Numerical Mathematics* 57.10 (Oct. 2007), pp. 1145–1162.

[85] J. Löber. *Optimal Trajectory Tracking of Nonlinear Dynamical Systems*. Springer, Dec. 2016.

[86] C. Long, P. Polisetty, and E. Gatzke. "Nonlinear Model Predictive Control Using Deterministic Global Optimization". In: *Journal of Process Control* 16.6 (July 2006), pp. 635–643.

[87] R. Luus and D. E. Cormack. "Multiplicity of Solutions Resulting from the Use of Variational Methods in Optimal Control Problems". In: *The Canadian Journal of Chemical Engineering* 50.2 (Apr. 1972), pp. 309–311.

[88] K. Makino and M. Berz. "Remainder Differential Algebras and Their Applications". In: *Computational Differentiation: Techniques, Applications, and Tools*. SIAM, 1996, p. 13.

[89] G. P. McCormick. "Computability of Global Solutions to Factorable Nonconvex Programs: Part I–Convex Underestimating Problems". In: *Mathematical Programming* 10.1 (1976), pp. 147–175.

[90] R. Misener and C. A. Floudas. "ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations". In: *Journal of Global Optimization* 59.2-3 (July 2014), pp. 503–526.

[91] I. Mitchell, A. Bayen, and C. Tomlin. "A Time-Dependent Hamilton-Jacobi Formulation of Reachable Sets for Continuous Dynamic Games". In: *IEEE Transactions on Automatic Control* 50.7 (July 2005), pp. 947–957.

[92] A. Mitsos, B. Chachuat, and P. I. Barton. "McCormick-Based Relaxations of Algorithms". In: *SIAM Journal on Optimization* 20.2 (Jan. 2009), pp. 573–601.

[93] A. Mitsos, N. Asprion, C. A. Floudas, M. Bortz, M. Baldea, D. Bonvin, A. Caspari, and P. Schäfer. "Challenges in Process Optimization for New Feedstocks and Energy Sources". In: *Computers & Chemical Engineering* 113 (May 2018), pp. 209–221.

[94] R. E. Moore, R. B. Kearfott, and M. J. Cloud. *Introduction to Interval Analysis*. SIAM, Apr. 2009.

[95] C. Navasca and A. Krener. "Solution of Hamilton Jacobi Bellman Equations". In: *Proceedings of the 39th IEEE Conference on Decision and Control*. Vol. 1. Sydney, NSW, Australia: IEEE, 2000, pp. 570–574.

[96] N. S. Nedialkov, K. R. Jackson, and G. F. Corliss. "Validated Solutions of Initial Value Problems for Ordinary Differential Equations". In: *Applied Mathematics and Computation* (1999), p. 48.

[97] Y. Nesterov. "Nonsmooth Convex Optimization". In: *Lectures on Convex Optimization*. Springer Optimization and Its Applications. Cham: Springer International Publishing, 2018, pp. 139–240.

[98] A. Neumaier. *Interval Methods for Systems of Equations*. Encyclopedia of Mathematics and Its Applications. Cambridge: Cambridge University Press, 1991.

[99] U. D. o. L. OHSA. *Process Safety Management of Highly Hazardous Chemicals*. 1992.

[100] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Classics in Applied Mathematics. New York: Society for Industrial and Applied Mathematics, Jan. 2000.

[101] I. Papamichail and C. S. Adjiman. "A Rigorous Global Optimization Algorithm for Problems with Ordinary Differential Equations". In: *Journal of Global Optimization* 24.1 (2002), pp. 1–33.

[102] I. Pappas, N. A. Diangelakis, and E. N. Pistikopoulos. "The Exact Solution of Multiparametric Quadratically Constrained Quadratic Programming Problems". In: *Journal of Global Optimization* 79.1 (Jan. 2021), pp. 59–85.

[103] E. N. Pistikopoulos. "Perspectives in Multiparametric Programming and Explicit Model Predictive Control". In: *AIChE Journal* 55.8 (Aug. 2009), pp. 1918–1925.

[104] C. Rackauckas and Q. Nie. "DifferentialEquations.Jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia". In: *Journal of Open Research Software* 5.1 (May 2017), p. 15.

[105] D. M. Raimondo, G. Roberto Marseglia, R. D. Braatz, and J. K. Scott. "Closed-Loop Input Design for Guaranteed Fault Diagnosis Using Set-Valued Observers". In: *Automatica* 74 (Dec. 2016), pp. 107–117.

[106] T. Raïssi, N. Ramdani, and Y. Candau. "Set Membership State and Parameter Estimation for Systems Described by Nonlinear Differential Equations". In: *Automatica* 40.10 (Oct. 2004), pp. 1771–1777.

[107] D. Ralph and S. Dempe. "Directional Derivatives of the Solution of a Parametric Nonlinear Program". In: *Mathematical Programming* 70.1-3 (Oct. 1995), pp. 159–172.

[108] N. Ramdani, N. Meslem, and Y. Candau. "A Hybrid Bounding Method for Computing an Over-Approximation for the Reachable Set of Uncertain

Nonlinear Systems". In: *IEEE Transactions on Automatic Control* 54.10 (Oct. 2009), pp. 2352–2364.

[109] A. Rapaport and D. Dochain. "Interval Observers for Biochemical Processes with Uncertain Kinetics and Inputs". In: *Mathematical Biosciences* 193.2 (Feb. 2005), pp. 235–253.

[110] S. Ratschan and Z. She. "Constraints for Continuous Reachability in the Verification of Hybrid Systems". In: *Proceedings of the 8th International Conference on Artificial Intelligence and Symbolic Computation*. AISC'06. Berlin, Heidelberg: Springer-Verlag, Sept. 2006, pp. 196–210.

[111] A. M. Sahlodin and B. Chachuat. "Discretize-Then-Relax Approach for Convex/Concave Relaxations of the Solutions of Parametric ODEs". In: *Applied Numerical Mathematics* 61.7 (July 2011), pp. 803–820.

[112] A. Sahlodin and B. Chachuat. "Convex/Concave Relaxations of Parametric ODEs Using Taylor Models". In: *Computers & Chemical Engineering* 35.5 (May 2011), pp. 844–857.

[113] 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". In: *Journal of Global Optimization* (Aug. 2018).

[114] H. Schichl. "Global Optimization in the COCONUT Project". In: *Numerical Software with Result Verification*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2004, pp. 243–249.

[115] S. Scholtes. *Introduction to Piecewise Differentiable Equations*. SpringerBriefs in Optimization. New York: Springer-Verlag, 2012.

[116] J. Schumacher. "Complementarity Systems in Optimization". In: *Mathematical Programming* 101.1 (Sept. 2004), pp. 263–295.

[117] J. K. Scott. "Reachability Analysis and Deterministic Global Optimization of Differential-Algebraic Systems". PhD thesis. Massachusetts Institute of Technology, 2012.

[118] J. K. Scott and P. I. Barton. "Bounds on the Reachable Sets of Nonlinear Control Systems". In: *Automatica* 49.1 (Jan. 2013), pp. 93–100.

[119] J. K. Scott and P. I. Barton. "Convex Relaxations for Nonconvex Optimal Control Problems". In: *IEEE Conference on Decision and Control and European Control Conference*. Orlando, FL, USA: IEEE, Dec. 2011, pp. 1042–1047.

[120] J. K. Scott and P. I. Barton. "Improved Relaxations for the Parametric Solutions of ODEs Using Differential Inequalities". In: *Journal of Global Optimization* 57.1 (Sept. 2013), pp. 143–176.

[121] J. K. Scott and P. I. Barton. "Tight, Efficient Bounds on the Solutions of Chemical Kinetics Models". In: *Computers & Chemical Engineering* 34.5 (May 2010), pp. 717–731.

[122] J. K. Scott, B. Chachuat, and P. I. Barton. "Nonlinear Convex and Concave Relaxations for the Solutions of Parametric ODEs". In: *Optimal Control Applications and Methods* 34.2 (2013), pp. 145–163.

[123] J. K. Scott, M. D. Stuber, and P. I. Barton. "Generalized Mccormick Relaxations". In: *Journal of Global Optimization* 51.4 (Dec. 2011), pp. 569–606.

[124] K. Shen and J. K. Scott. "Exploiting Nonlinear Invariants and Path Constraints to Achieve Tighter Reachable Set Enclosures Using Differential Inequalities". In: *Mathematics of Control, Signals, and Systems* 32.1 (Mar. 2020), pp. 101–127.

[125] K. Shen and J. K. Scott. "Mean Value Form Enclosures for Nonlinear Reachability Analysis". In: *2018 IEEE Conference on Decision and Control (CDC)*. Miami Beach, FL: IEEE, Dec. 2018, pp. 7112–7117.

[126] K. Shen and J. K. Scott. "Rapid and Accurate Reachability Analysis for Nonlinear Dynamic Systems by Exploiting Model Redundancy". In: *Computers & Chemical Engineering* 106 (Nov. 2017), pp. 596–608.

[127] K. Shen and J. K. Scott. "Tight Reachability Bounds for Nonlinear Systems Using Nonlinear and Uncertain Solution Invariants". In: *2018 Annual American Control Conference (ACC)*. Milwaukee, WI: IEEE, June 2018, pp. 6236–6241.

[128] A. B. Singer and P. I. Barton. "Bounding the Solutions of Parameter Dependent Nonlinear Ordinary Differential Equations". In: *SIAM Journal on Scientific Computing* 27.6 (Jan. 2006), pp. 2167–2182.

[129] A. B. Singer, J. W. Taylor, P. I. Barton, and W. H. Green. "Global Dynamic Optimization for Parameter Estimation in Chemical Kinetics". In: *The Journal of Physical Chemistry A* 110.3 (Jan. 2006), pp. 971–976.

[130] Y. Song and K. A. Khan. "Comparing Solutions of Related Ordinary Differential Equations Using New Differential Inequalities". In: *Under review* (2021).

[131] Y. Song and K. A. Khan. "Optimization-Based Convex Relaxations for Nonconvex Parametric Systems of Ordinary Differential Equations". In: *Mathematical Programming* (Apr. 2021).

[132] Y. Song and K. A. Khan. "Subgradient Propagation for ODE Relaxations Using Differential Inequalities". In: (in preparation).

[133] Y. Song, H. Cao, C. Mehta, and K. A. Khan. "Bounding Convex Relaxations of Process Models from below by Tractable Black-Box Sampling". In: *Computers & Chemical Engineering* 153 (Oct. 2021), p. 107413.

[134] M. Stuber and M. Wilhelm. "Easy Advanced Global Optimization (Eago): An Open-Source Platform for Robust and Global Optimization". In: Oct. 2017.

[135] M. D. Stuber and P. I. Barton. "Semi-Infinite Optimization with Implicit Functions". In: *Industrial & Engineering Chemistry Research* 54.1 (Jan. 2015), pp. 307–317.

[136] M. D. Stuber, J. K. Scott, and P. I. Barton. "Convex and Concave Relaxations of Implicit Functions". In: *Optimization Methods and Software* 30.3 (May 2015), pp. 424–460.

[137] J. Szarski. *Differential Inequalities*. PWN, 1965.

[138] M. Tawarmalani and N. Sahinidis. *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications*. Nonconvex Optimization and Its Applications. Springer US, 2002.

[139] M. Tawarmalani and N. V. Sahinidis. "A Polyhedral Branch-and-Cut Approach to Global Optimization". In: *Mathematical Programming* 103.2 (June 2005), pp. 225–249.

[140] J. W. Taylor, G. Ehlker, H.-H. Carstensen, L. Ruslen, R. W. Field, and W. H. Green. "Direct Measurement of the Fast, Reversible Addition of Oxygen to Cyclohexadienyl Radicals in Nonpolar Solvents". In: *The Journal of Physical Chemistry A* 108.35 (Sept. 2004), pp. 7193–7203.

[141] I. B. Tjoa and L. T. Biegler. "Simultaneous Solution and Optimization Strategies for Parameter Estimation of Differential-Algebraic Equation Systems". In: *Industrial & Engineering Chemistry Research* 30.2 (Feb. 1991), pp. 376–385.

[142] C. Tomlin, I. Mitchell, A. Bayen, and M. Oishi. "Computational Techniques for the Verification of Hybrid Systems". In: *Proceedings of the IEEE* 91.7 (July 2003), pp. 986–1001.

[143] T. H. Tsang, D. M. Himmelblau, and T. F. Edgar. "Optimal Control via Collocation and Non-Linear Programming". In: *International Journal of Control* (Mar. 2007).

[144] A. Tsoukalas and A. Mitsos. "Multivariate McCormick Relaxations". In: *Journal of Global Optimization* 59.2-3 (July 2014), pp. 633–662.

[145] A. Tulsyan and P. I. Barton. "Reachability-Based Fault Detection Method for Uncertain Chemical Flow Reactors". In: *IFAC-PapersOnLine* 49.7 (2016), pp. 1–6.

[146] M. E. Villanueva, B. Houska, and B. Chachuat. "Unified Framework for the Propagation of Continuous-Time Enclosures for Parametric Nonlinear Odes". In: *Journal of Global Optimization* 62.3 (July 2015), pp. 575–613.

[147] A. Wächter and L. T. Biegler. "On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming". In: *Mathematical Programming* 106.1 (Mar. 2006), pp. 25–57.

[148] W. Walter. *Differential and Integral Inequalities*. Ergebnisse Der Mathematik Und Ihrer Grenzgebiete. 2. Folge. Berlin Heidelberg: Springer-Verlag, 1970.

[149] Z. Wang, M. S. Escotet-Espinoza, and M. Ierapetritou. "Process Analysis and Optimization of Continuous Pharmaceutical Manufacturing Using Flowsheet Models". In: *Computers & Chemical Engineering*. In Honor of Professor Rafiqul Gani 107 (Dec. 2017), pp. 77–91.

[150] A. Wechsung, S. D. Schaber, and P. I. Barton. "The Cluster Problem Revisited". In: *Journal of Global Optimization* 58.3 (Mar. 2014), pp. 429–438.

[151] A. Wechsung, J. K. Scott, H. A. J. Watson, and P. I. Barton. "Reverse Propagation of McCormick Relaxations". In: *Journal of Global Optimization* 63.1 (Sept. 2015), pp. 1–36.

[152] M. E. Wilhelm and M. D. Stuber. "EAGO.Jl: Easy Advanced Global Optimization in Julia". In: *Optimization Methods and Software* (Aug. 2020), pp. 1–26.

[153] M. E. Wilhelm, A. V. Le, and M. D. Stuber. "Global Optimization of Stiff Dynamical Systems". In: *AIChE Journal* 65.12 (Dec. 2019).

[154] D. Wolpert and W. Macready. "No Free Lunch Theorems for Optimization". In: *IEEE Transactions on Evolutionary Computation* 1.1 (Apr. 1997), pp. 67–82.

[155] X. Yang and J. K. Scott. "A Comparison of Zonotope Order Reduction Techniques". In: *Automatica* 95 (Sept. 2018), pp. 378–384.

[156] J. Zhang, L. Xie, and S. Wang. "Particle Swarm for the Dynamic Optimization of Biochemical Processes". In: *Computer Aided Chemical Engineering*. Vol. 21. Elsevier, 2006, pp. 497–502.