Imperial College London

Set-Theoretic Methods for Analysis, Estimation and Control of Nonlinear Systems

Mario Eduardo Villanueva

Centre for Process Systems Engineering
Department of Chemical Engineering
Imperial College London

This dissertation is submitted in part fulfilment
of the requirements for the degree:
Doctor of Philosophy In Chemical Engineering
and Diploma of Imperial College

March 2016

To my mothers Maria and Soledad, two exceptional women to whom I owe everything . . .
To the memory of Toyita and Nacho, my loving grandparents. . .

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation contains fewer than 100,000 words including appendices, bibliography, footnotes, tables and equations.

Mario Eduardo Villanueva
March 2016

"One tube to rule them all, One tube to imply them
One tube to control them all and in the uncertainty bound them
In the realm of sets, where enclosures lie."

–M.E. Villanueva and H.L. Sant'Ana Pereira,
based on J.R.R. Tolkien's epigraph to The Lord of The Rings

Acknowledgements

Someone once told me that if you did not spend at least as much time writing the acknowledgements as you did writing the thesis, then you definitely did something wrong. Well I guess I did something right, since I have spent about a month and a half writing this part of my thesis and I still believe I have not properly acknowledged everyone I wanted to.

First, I would like to express my most sincere gratitude to Dr. Benoît Chachuat –my thesis supervisor. I definitely owe him greatly for the time and attention he devoted to my development as a researcher. Benoît was always available whenever I knocked on his door to ask him 'a quick question'–which always ended up turning into a full-length discussion. I could list all of the qualities that make Benoît a great supervisor, instead I will only say that if I could go back to that moment where he offered me the opportunity to work with him I would definitely say yes again.Benoít: It is a great honour to have worked with you and call you a mentor and most importantly my friend.

I would also like to extend my gratitude to Prof. Efstratios Pistikopoulos and Prof. Rolf Findeisen for their comments and corrections to this thesis, I am deeply grateful to have had you as my examiners. To Dr. Boris Houska for providing me with his mathematical insight and for all those tricky proof methods he taught me during our coffee breaks; without him this thesis would not be what it is today. To Dr. Radoslav Paulen for making complicated concepts seem like child's play. To all the members of the OMEGA group: it was great to work and collaborate with you. Thanks also to Prof. Moritz Diehl and the Systems Theory group at the Institute for Microsystems Technology (IMTEK) at the University of Freiburg for hosting me and for the fruitful collaboration we shared.

Thanks to the Mexican Council for Research and Development (CONACyT) for providing me with a doctoral scholarship. I would also like to thank staff at Imperial College London for their help. In particular to Mrs. Susi Underwood for her continuous help and advice, specially when I needed to get 'something done'.

To Andreas, Channarong (K), Hugo, Jai, Naveed, Susie, Nasim and Niki–the legendary 509 crew– for turning the everyday at the office into an unique experience, there has never nor will ever be a better crew at that office. To Matt, Pedro, Jose, Eneritz, Javier, Tomas,

Thomas, Zili and Xiangyi –the mostly spanish friday night beer crew– for everyday making me wish it was friday, for the amazing dinners we shared and in general for making me forget about work on fridays after seven. To my friend Benito, for his encouragement and his help in my quest to formalize my knowledge of mathematics. To my friends Tine and Emilio for their support. To my friend Marco for always helping me and being there whenever I needed. To my flatmates Shane, Shelly, Keeran (and for a brief time Naveed, Omar, Matt and Laura), I know its not easy living with me but you managed to do it while always providing me with your friendship and support. Special thanks to my friend and brother Naveed: I have no words to thank you for all the personal support you gave me.

To my Portuguese family and friends in particular to Tio Luis, Tia Ana, Tomás, Fils and Hugo for their love and support, for receiving me in your home and making me feel like a member of your family. Also to Tio Finzi, Tia Paula, Tia Tete, Tio Kaka, Tio Kiki, Tia Carla, Nuno, Mafi, Rui, Rita, Joanna, Buga, Miguel, Jimmy, Pi and Gonçalo for all the laughs, the talks, the wine the food, the pool, the new-year cigars, the Benfica matches (Carrega Benfica!) and in general all the experiences we shared that made me feel like an honorary Portuguese. Without those visits to Lisbon I don't know if I would have made it. Special thanks also to my friend and brother-from-another-mother Hugo Sant'Ana Pereira, for all the moments we shared, for his support in my times of need, for his help with this thesis, for all the amazing conversations we had about work, life, the future, science, ...; for everything I thank you my friend.

To my Mexican friends and family for their love and continuous support throughout my life and career. In particular I would like to thank my uncles, aunts and cousins, I always and will always have you by my side. To Faby and Paola, thanks for your love and respect. To my grandparents, thank you for always believing in me. To my crew Daniel, Omar, Beto and Amado –it is unbelievable that we still remain as close friends as we were 20 years ago.

Thanks to my teachers, mentors and all the people who have helped shaping me into the man that I am today.

Finally, none of this would have been possible without two amazing women: Maria and Maria Soledad Villanueva Infanzón –my loving mothers and the two most important women in my life. To them all my love, admiration and gratitude. Amadas mamás, nunca terminaré de pagarles todo el amor que me han dado y todo lo que han hecho por mi. Gracias su ejemplo soy el hombre que soy ahora y este logro es tan mio como de ustedes. Le agradezco a Dios haberme puesto en nuestra familia. Las amo y una vez más: Gracias.

Gracias totales,
Mario Eduardo Villanueva.

Abstract

This thesis is concerned with the application of set-theoretical methods to problems in analysis, estimation and control of nonlinear systems. Set-theoretical concepts are often used in the formulation of various problems in science and engineering. One of the key enablers for the successful application of set-theoretical methods is the ability to enclose the image set of nonlinear multivariate systems, which is the focus of the main body of this thesis.

Chapter 2 concentrates on bounding the image of factorable, vector-valued functions. To this aim, a framework is developed which enables the analysis of existing set-valued arithmetics—such as interval and polynomial model arithmetics—and the construction of new ones e.g. an ellipsoidal arithmetic for vector-valued nonlinear factorable functions. This framework also allows to study on a unified way the convergence (in the Hausdorff sense) of the enclosures to the exact image as the domain of the function shrinks.

Chapters 3 and 4 are concerned with set propagation through dynamic systems (reachability analysis) defined by parametric ordinary differential equations (ODEs). Computing enclosures for the reachable set is not straightforward, since it is the image of a function which is not factorable, but it is defined implicitly by the ODEs. Nevertheless, computational methods for reachability analysis can take advantage the factorable structure of the ODE right-hand side. The focus on Chapter 3 is on discrete-time set propagation, i.e. methods where the integration horizon is discretized into finite steps, and then propagating the enclosure through each of these steps. Classical methods rely on Taylor expansions of the ODE solution and proceed in two phases. First, an a step-size and an a priori enclosure for the reachable set over the current step are determined. Then, the enclosure is tightened at the end of the step. The algorithm presented in this chapter is also based on Taylor expansions of the solution, but the order of the phases is reversed. This construction leads to a natural step-size control mechanism and eliminates the need for tightening the enclosure at the end of the time-step. Furthermore, sufficient conditions are then derived for the algorithm to be locally asymptotically stable in the neighborhood of a locally asymptotically stable periodic orbit or equilibrium point. The key requirement for stability is that the affine-set extensions used in the propagation have quadratic Hausdorff convergence order.

On the other hand, Chapter 4 deals with continuous-time set propagation methods. These class of methods rely on the construction of an auxiliary system of ODEs, whose solution is guaranteed to enclose the reachable set of the original ODEs. Here, a unified framework for the construction of continuous time methods is presented. It is based on a generalized differential inequality (GDI), whose solutions describe the support function of time-varying enclosures for the reachable set. This GDI contains as special cases known continuous-time reachability methods, such as differential inequalities and ellipsoidal set propagation techniques. Although the GDI is based on the support function characterization of convex sets, an extension for nonconvex sets is provided using polynomial models with convex remainders. The framework also provides a means for analyzing the Hausdorff convergence properties of continuous-time enclosure methods. A nontrivial extension of the GDI in the form of a min-max differential inequality is also introduced for the characterization of robust forward invariant tubes. This min-max DI provides as a by product a semi-explicit nonlinear feedback control law, which can also be exploited in robust optimal control and tube-based robust model predictive control.

Chapter 5 is concerned with the characterization of sets defined implicitly by systems of constraints. These problems are addressed from a set-theoretical perspective, by adopting the use of complete-search based constraint projection methods. The chapter presents a branch-and-prune algorithm which is enhanced by the use of higher order bounding strategies based on polynomial models. The use of optimization-based domain reduction strategies inspired by developments in branch-and-bound algorithms for complete-search global optimization is also studied. We also introduce a CPU time reduction strategy for polynomial models, which allows reusing the computed bounds whenever they have converged. For constraint systems that include undetermined systems of equations a domain reduction strategy in reduced space is presented. This strategy relies on the use of polynomial models in order to characterize the boundary of the set and makes use of state-of-the-art Newton-like methods for the solution of systems of nonlinear implicit algebraic equations. The algorithm is applied to two different problems: guaranteed parameter estimation and guaranteed asymptotic analysis.

The methods in this thesis have been implemented in CRONOS (https://bitbucket.org/omega-icl/cronos), a C++ library that builds upon the MC++ library (https://bitbucket.org/omega-icl/mcpp) for bounding factorable functions. These algorithms have also been tested in a variety of case studies drawn from Chemical Engineering and Systems biology.

Table of contents

Notation

Without recalling standard mathematical notation, we collect in this section some of the syntax used in this thesis. This list is not comprehensive and all of the concepts which are not defined here are defined throughout the different chapters.

The set of compact subsets of $\mathbb{R}^n$ is denoted by $\mathbb{K}^n$, and the subset of compact convex subsets of $\mathbb{K}^n$, by $\mathbb{K}^n_{\mathrm{C}}$. We define the unit ball $\mathscr{B}^n \in \mathbb{K}^n_{\mathrm{C}}$ as $\mathscr{B}^n := \{z \in \mathbb{R}^n \mid \|z\|_2 \leq 1\}$. The diameter $\operatorname{diam}(Z)$ of a set $Z \in \mathbb{K}^n$ is defined as

$$\operatorname{diam}(Z) := \max_{z_1, z_2 \in Z} \|z_1 - z_2\|,$$

and the support function $V[Z] : \mathbb{R}^n \to \mathbb{R}$ of Z as

$$\forall c \in \mathbb{R}^n, \quad V[Z](c) \;:=\; \max_{z} \; \left\{ c^{\top} z \mid z \in Z \right\}.$$

The Minkowski sum of two compact sets $W, Z \in \mathbb{K}^n$ is denoted by

$$W \oplus Z := \{w + z \mid w \in W, z \in Z\},$$

and the Haussdorf distance between W and Z is given by

$$d_{\mathrm{H}}(W,Z) \;:=\; \max \left\{ \max_{w \in W} \min_{z \in Z} \|w - z\| \,, \max_{z \in Z} \min_{w \in W} \|w - z\| \right\}. \tag{1}$$

In particular, the Hausdorff distance of any two convex compact sets $W, Z \in \mathbb{K}^n_{\mathrm{C}}$ is bounded by the maximum difference of their support functions

$$d_{\mathrm{H}}(W,Z) \;\leq \max_{c \in \mathbb{R}^n, c^T c = 1} |V[W](c) - V[Z](c)| \,. \tag{2}$$

This result follows readily from (1), by noting that the Hausdorff distance $d_{\mathrm{H}}(W,Z) = \|w^* - z^*\|$ at a min-max point [resp. a max-min point] $(w^*, z^*) \in W \times Z$ is bounded by the dif-

ference $|V[W](c) - V[Z](c)|$ in the direction $c = w^* - z^*$ [resp. in the direction $c = z^* - w^*$]. If $W \subseteq Z$, we have

$$d_{\mathrm{H}}(W,Z) = \max_{z\in Z} \min_{w\in W} \|w - z\| \ .$$

Moreover, by a small abouse of notation, we denote the Hausdorff distance between a compact set $Z \in \mathbb{K}^n$ and the origin by

$$\|Z\|_{\mathrm{H}} := d_{\mathrm{H}}(Z,\{0\}) = \max_{z\in Z} \|z\| \ ,$$

although the function $\|\cdot\|_{\mathrm{H}}$ does not define a norm in general. The set of n-dimensional interval vectors is denoted by $\mathbb{IR}^n$. The midpoint and radius of an interval vector $P := \left[p^{\mathrm{L}}, p^{\mathrm{U}}\right] \in \mathbb{IR}^n$ are defined as $\mathrm{mid}\,(P) := \frac{1}{2}(p^{\mathrm{U}} + p^{\mathrm{L}})$ and $\mathrm{rad}(P) := \frac{1}{2}(p^{\mathrm{U}} - p^{\mathrm{L}})$, respectively. The n-by-n matrix $\mathrm{diag}\,\mathrm{rad}\,(P) \in \mathbb{R}^{n\times n}$ is a diagonal matrix whose elements are the components of $\mathrm{rad}(P)$.

The set of n-dimensional positive semi-definite symmetric matrices is denoted by $\mathbb{S}^n_+$, while the set of n-dimensional positive definite matrices is denoted by $\mathbb{S}^n_{++}$. For a matrix $M \in \mathbb{S}^n_+$, $M^{\frac{1}{2}}$ denotes its symmetric square-root. An ellipsoid with center $c \in \mathbb{R}^n$ and shape matrix $M \in \mathbb{S}^n_+$ is denoted by

$$\mathcal{E}(c,M) := \left\{ c + M^{\frac{1}{2}} v \mid v \in \mathbb{R}^n : v^{\mathsf{T}} v \leq 1 \right\}$$

or, simply, $\mathcal{E}(M)$ if the ellipsoid is centered at the origin.

he Moore-Penrose pseudoinverse of a rectangular matrix $M \in \mathbb{R}^{m\times n}$ is denoted by $M^{\dagger}$ and its Frobenius norm by $\|M\|_F := \sqrt{\mathrm{Tr}(A^{\mathsf{T}}A)}$.

A multi-index γ is a vector in $\mathbb{N}^n$, $n > 0$. The order of γ is $|\gamma| := \sum_{i=1}^n \gamma_i$. Given a point $p \in \mathbb{R}^n$, p^γ is a shorthand notation for the expression $\prod_{i=1}^n p_i^{\gamma_i}$. Moreover, given a function $f : \mathbb{R}^n \to \mathbb{R}$, $\partial^\gamma f$ is a shorthand notation for the partial derivative $\frac{\partial^{|\gamma|} f}{\partial p_1^{\gamma_1} \cdots \partial p_n^{\gamma_n}}$.

The set of n-dimensional Lebesgue-integrable functions on the interval $I \subseteq \mathbb{R}$ is denoted by $\mathbb{L}(I)^n$. Likewise, $\mathbb{L}^n$ denotes the set of Lebesgue integrable functions on $\mathbb{R}$. Unless otherwise stated, Lebesgue integration is understood with respect to the time variable.

Chapter 1

Introduction

A great variety of natural, engineered and industrial systems are inherently dynamic, including batch processes and processes operating at a cyclic steady state. Other such systems are operated in a transient manner on purpose, for instance in order to improve their versatility or robustness. All these systems have their evolution governed by differential equations and are often uncertain due to the presence of time-varying disturbances or due to incomplete/inaccurate knowledge of certain parameters or initial conditions.

Uncertain dynamic systems in the form of ordinary differential equations appear in design and control problems. These problems can in turn be formulated, analyzed or solved in a set-theoretic framework. By set-theoretic we refer here to any method which exploits properties of suitably chosen sets or constructed sets in the state space [18]. In designing a control system for instance, the constraints, uncertainties and design specifications all together are described naturally in terms of sets; and in measuring the effect of a disturbance on a system's response or in bounding the error of an estimation algorithm likewise, sets play a central role. A number of key set-theoretic concepts have been proposed in the early 1970s, but their systematic applications were not possible until enough computational capability became widely available.

Today, many such methods and tools are available for the estimation and control of linear systems – Witness for instance the popularity of Matlab's Multi-Parametric Toolbox [48]. To name but a few features, they support the construction and a variety of operations on convex sets, the synthesis and implementation of explicit model predictive control (MPC) for linear time invariant or piecewise affine systems, the construction of maximal invariant sets or Lyapunov functions for piecewise affine systems, *etc* [see, e.g., 69, 18].

Despite enormous progress in recent years, set-theoretic methods and tools for nonlinear systems are not as developed as their linear counterparts, and they still constitute a widely

open fields of research when it comes to computational efficiency; see, e.g., Streif et al. [141] for a recent survey (with applications to biochemical networks).

The main aim of this thesis is to develop methods inspired by set-theory for the solution of problems in estimation and control of nonlinear systems. The thesis can roughly be divided in three main parts. The first part is concerned with computing enclosures for the image set of nonlinear vector-valued factorable functions—informally, those nonlinear vector-valued functions that can be represented explicitly by a computer code. The second part is concerned with the construction of enclosures for the image set of functions defined by ordinary differential equations, which are not factorable. The third part is concerned with the application of the above enclosure methods to higher-level set-theoretical algorithms, such as constraint projection for the characterization of implicitly defined sets. Through this thesis we approach problems from a unified theoretical perspective, but never loosing sight of their computational implementation.

The remainder of this chapter presents a review of the literature relevant to the topics addressed in this thesis. Section 1.1 reviews the literature dealing with the construction of enclosures for factorable functions. Section 1.2 presents a road map of the contributions made by different communities to set-propagation methods for dynamic systems. Section 1.3 presents an overview of methods for robust model predictive control, with emphasis on the set-theoretical formalism known as tube-base MPC. In Section 1.4 we review the literature concerned with the characterization of implicitly defined sets. In this section we emphasize contributions to the problem of guaranteed parameter estimation and the global solution of systems of nonlinear algebraic equations. Finally, Section 1.5 presents the contributions and the general organization of this thesis.

1.1 Set Propagation Through Factorable Functions

A key enabler for set-theoretic methods is the ability to enclose the range of nonlinear multivariate systems, and the class of factorable functions—namely, those functions which can be represented by means of a finite computational graph—has attracted much attention. Since the invention of interval analysis by Moore more than 50 years ago, many computational techniques have been developed to construct tight enclosures for the range of factorable functions. For real-valued, factorable functions, interval arithmetic [97] provides a natural way of computing such enclosures, since the image set of any continuous function is itself an interval. In practice, an interval enclosure can be obtained by traversing a directed acyclic graph (DAG) of the function in order to bound its atom operations recursively and

as tightly as possible [129]. Although simple, this approach suffers two main limitations, namely the dependency problem and the wrapping effect. The former happens when multiple occurrences of the same variable, e.g. in a complicated function expression, are treated as if they were independent from each other. The latter is due to the fact that the image of an interval vector under a vector-valued function, even a linear linear function, is generally not an interval vector itself, thus leading to overestimation in enclosing that image set with an interval vector [82].

The inherent problems of interval arithmetic have lead to the development of alternative methods to bound the image set of a function, such as the construction of so called convex/concave enclosures. Convex/concave enclosures are of particular interest in global optimization, since their aim is to construct under and over estimators for the graph of a function. One of such approaches is the α-based enclosures for general twice-continuously differentiable functions [86]. The general idea behind this method is to subtract (resp. add) from the original function a quadratic term —which depends on a $\alpha \geq 0$ in such a way that the resulting function is convex (resp. concave) in the domain of interest. Intuitively we can visualize the negative quadratic term as a way of 'pulling down' the function until it becomes convex. The main challenge in constructing α-based estimators is in calculating a value of the α parameter which will guarantee the convexity of the underestimator, while being small enough for the enclosure to remain tight. Methods to calculate the α parameter rely on the observation that the enclosure is convex if and only if its Hessian matrix is positive semidefinite in the domain of interest, thus computing the α parameter often requires bounds on the eigenvalues of the Hessian matrix on the domain. Bounds on the eigenvalues of the Hessian matrix can be computed using interval arithmetic and Gersgorin's theorem [50] or the so called eigenvalue arithmetic for twice-continuously differentiable functions [94].

Another approach to construct convex and concave enclosures was defined by McCormick, in his seminal 1976 paper [89]. McCormick's composition technique provides a way to relax the original function by the application of rules for the relaxation of univariate composition, binary multiplication and binary addition. This approach requires the use of interval arithmetic and the availability of convex and concave enclosures for univariate functions. McCormick relaxations have also been generalized to account for the case of composition of functions where the the outter function is factorable but the inner function is not [134]. One of the key properties of McCormick and generalized McCormick relaxations is that they are at least as tight as their underlying interval bounds [128]. In general, these relaxations are nonsmooth, thus they can not be used directly with gradient-based optimization methods. Approaches to handle this problem rely on the use of nonsmooth optimization

which requires the construction of subgradients of the enclosure. A systematic method for the construction of subgradients has been presented in [93]. Analogously to the construction of McCormick relaxation, the propagation of subgradients is carried out by a recursive application of rules for binary addition and multiplication and composition with univariate functions.

It is worth noticing that α-based enclosures and McCormick relaxations are convex when we consider them as enclosures of the the graph of the function, i.e. the enclosures are convex with respect to its variables. When considering them as enclosures of the image of the function they are only intervals. Nevertheless, convexity of the enclosure with respect to its variables is a higly desirable property in optimization, since any local optimization method for a convex problem will yield a global optimum. For the problem of bounding the image of a vector-valued function, convexity may not be such a desirable property since the image of a vector-valued function is in general nonconvex. Taylor models [15] provide a way of constructing nonconvex enclosures using multivariate polynomials. Informally a Taylor model of a function is a pair consisting of a multivariate polynomial of a given order matching the Taylor expansion of the function at a point and an interval remainder bounding the difference between the expansion and the function [103]. Taylor models can be constructed and propagated using arithmetic rules much like McCormick relaxations. Computationally, the polynomial is propagated by symbolic computations whenever its possible while specialised rules to handle the interval remainder exist [15, 13, 103]. Taylor polynomials provide an approximation of a function around a given point, and thus are somehow 'local' approximations. Since, we are often looking for global properties of a function it makes sense to look for 'global approximations'. Recently a polynomial arithmetic based on Chebyshev polynomial basis has been introduced and analysed [113, 34]. This arithmetic has the same convergence properties of Taylor models while providing the global approximation characteristics of Chebyshev expansions and Chebyshev interpolating polynomials. As mentioned Taylor and Chebyshev models are nonconvex both in the function space and with respect to its variables, but methods to obtain convex bounds based on Taylor models have also been developed. One way of obtaininig convex/concave bounds its to compute an interval extension of the polynomial and add the interval remainder. Although straightforward, the bounds obtained through a naive interval computation can be rather loose, this has lead to several improvements such as computing exact interval bounds for the first and second order terms of the polynomial [85, 78] and expressing the polynomial in a Bernstein basis [77]. Another technique for computing convex/concave enclosures involves the use of McCormick relaxations for Taylor models [93].

Polytopes can also be used to enclose the image of a function. Polyhedral relaxations have been widely used in the context of global optimisation [142]. These relaxations can be constructed using the decomposition/relaxation/outer-approximation technique [139]. These technique proceeds by decomposing nonlinear terms into unary and binary operations with the aid of intermediate variables, then bilinear and fractional terms are replaced by polyhedral envelopes. In the last step, nonlinear relaxations of univariate functions are outer approximated using affine relaxations at well chosen points. Other techniques for the construction of polyhedral relaxations are the reformulation-linearization technique [136, 135] and affine relaxations (via subgradients) of McCormick relaxations [93].

Among the alternatives to enclose the range of functions or sets defined by equalities and/or inequalities, mention should also be made of the sum-of-squares (SOS) approaches which provide nested sequences outer-approximations as hierarchies of linear matrix inequality (LMI) relaxations [73]; see also Parrilo [107].

An important property of methods for constructing enclosures of the image of a function is how quickly these enclosures converge to the exact image as the size of the domain vanishes. This property is studied through the convergence rate of these enclosures in the Hausdorff sense. This is specially relevant in global optimization where it has been shown that the convergence order can be related to the so-called cluster effect [33, 155]. The convergence properties of McCormick relaxations, Taylor models and α-based estimators as applied to factorable functions have been analyzed in [19, 20]. In particular it was shown that McCormick relaxations and α-based estimators converge quadratically while the convergence of Taylor models is equal to the degree of the Taylor polynomial plus one.

The approaches taken for the construction of enclosures for the image-set of factorable functions are various, but they seem to have several properties in common. Now, Can these similarities be formalized in any way?, Can we converge to a general theory for the construction and analysis of enclosure methods for factorable functions?. Typically these methods are also analyzed in the context of multivariate but real-valued functions. When extending these methods to vector-valued functions, the analysis is carried componentwise. This raises another question, Do the properties of these methods trivially extend to vector-valued functions?. These questions will be addressed in Chapter 2As it turns out, we can construct a general arithmetic which contains most of the methods presented in this review. This framework was designed specially for vector-valued functions and allows to study concepts like convergence in general way. Moreover the framework provides means to construct new set-valued arithmetics.

1.2 Set Propagation Through Dynamic Systems

The next level of complexity of set propagation is the construction of enclosures for the set of states that can be reached by a dynamic system when it is subject to uncertainties either in the initial conditions or in the form of uncertain parameters. The aforementioned set is called reachable set and computing enclosures for the reachable set is motivated for example by applications in the field of dynamic optimization. In the context of global optimization, an over-approximation of the reachable set can be used to construct relaxations of the feasible region and the objective function, which are pivotal in branch-and-bound search and its variants; [see, e.g., 106, 24, 78, 126]. In robust optimization likewise, the enclosure of a system's reachable set can be used to ensure feasibility of that system's response despite the presence of uncertainty [56]. Other applications of reachability analysis are for problems in the fields of robust MPC [76], guaranteed state and parameter estimation [57, 62, 3], and system verification and fault detection [79, 145, 144].

Advances in the field of reachability analysis originate from various fields. In viability theory [8], uncertain dynamics are expressed as differential inclusions. In contrast, comparison principles consider auxiliary sets of ODEs that do not depend on the uncertain variables to enclose the reachable set of an uncertain dynamic system. One such comparison principle is provided by the theory of differential inequalities [71, 154], where the right-hand side of the auxiliary ODEs under/overestimate the minimal/maximal value of the right-hand side of the uncertain dynamic system. Other comparison principles relate the reachable sets of uncertain ODEs to the solutions of Hamilton-Jacobi-Bellman (HJB) and Hamilton-Jacobi-Isaacs (HJI) equations [67, 91, 92, 83].

Contributions to reachability analysis are also found in the field of reliable computing. Verified ODE integration techniques were first developed to compute guaranteed solutions for ODEs with rounding errors as the sole source of uncertainty. In the classical approach, the integration horizon is discretized into finite steps and a two-phase algorithm is applied at each step [35, 100]. Phase I is concerned with the computation of an *a priori* enclosure and a step-size, such that existence and uniqueness of the solution can be ascertained on that domain. This typically relies on a high-order Taylor series expansion of the ODE solutions, in combination with interval analysis [29, 99]. Phase II propagates a tightened enclosure until the end of a given step, for instance based on an interval Taylor series [81] or a more involved interval Hermite-Obreschkoff method [100]. Further developments aiming to reduce conservatism in Phase II include wrapping mitigation strategies [81] as well as the use of Taylor models [84, 103] to mitigate the dependency problem [16, 14, 80, 101]. A hybrid

approach combining differential inequalities with the two-phase approach is proposed in [119], which provides a verified implementation of the method of differential inequalities. Another verified implementation featuring a simple Runge-Kutta method in combination with a Piccard iteration scheme is described in [120].

A well-developed theory for enclosing the solution set of linear and hybrid uncertain systems has emanated from the field of control theory. These methods commonly proceed by propagating sets forward in time too [18]. One example is the ellipsoidal bounding approach [70], first developed for linear systems under ellipsoidal uncertainty sets, and later extended to address nonlinear systems as well [56]. This method involves constructing vector- and matrix-valued ODEs, whose solutions are, respectively, the center and the shape matrix of ellipsoids that enclose the reachable set of the uncertain dynamic system, pointwise in time [147]. Other set representations used for bounding reachable sets include polytopes [26] and zonotopes [6].

In the field of global optimization, the construction of convex lower-bounding problems for nonlinear dynamic optimization problems is pivotal to the development of efficient complete search methods. Considerable effort has thus been devoted to obtaining convex estimators for the solution of parametric nonlinear ODEs with respect to the uncertain parameters. For linear dynamic systems, a method was developed in [137] which provides affine bounds in the parameter space via the solution of an auxiliary set of ODEs. This method was later extended to propagate affine bounds [138] as well as a pair of convex and concave bounds [131, 132] for the solutions of parametric nonlinear ODEs based on differential inequalities and (generalized) McCormick relaxations [89, 134]. The refinement of reachable set enclosures by accounting for *a priori* or physical information about a given system was also investigated in [137, 133]. Another way of computing pointwise-in-time, convex under-estimators and concave over-estimators using the α-BB approach [87, 2] has been described in [106, 22, 105]. Discretization-based methods adapted from the field of reliable computing too have been considered in the context of global optimization. An extension of the two-phase approach based on Taylor models for handling parametric uncertainty was considered in [80, 78], and similar extensions for the propagation of McCormick relaxations and McCormick-Taylor models [20] are described in [127, 128].

Other approaches involve constructing outer-approximations of the reachable set of (possible controlled and constrained) dynamic systems as hierarchies of linear matrix inequality (LMI) relaxations [47]. Moreover, so-called indirect approaches have also been developed, which formulate reachability questions as optimal control (or game theory) problems and determine reachable-set outer-approximations by Hamilton-Jacobi projections [91, 83]. Re-

lated to reachability analysis, the problems of computing the region of attraction of a target set or the maximum invariant set are also essential and long-standing challenges in dynamic system and control theory [47, 65].

In sum, existing methods for constructing enclosures of the reachable set of nonlinear ODEs are scattered across different disciplines. A possible classification for all of these methods is whether the enclosures are propagated continuously in time or on a time grid after discretization of the integration horizon.

The concept of Hausdorff convergence order can also be extended to the enclosures of reachable sets. Although no systematic analysis has been carried out for the propagation of the convergence order when propagating sets through ODEs it is not hard to contrive examples whereby the enclosures computed based on differential inequalities would only enjoy first-order Hausdorff convergence. One of the main problems that these method suffer, is the so-called bound explosion which occurs when the size of the enclosures of reachable sets grows without a bound in a finite time horizon. This problem is analogous to the finite escape time occurring in ODEs whose right-hand sides are not Lipschitz continuous. Although no rigorous analysis of this problem has been carried out, empirical evidence points out to the accumulation of overestimation through the set propagation method.

Again, the first question that can be asked at this point is: What do these methods have in common?. In particular, Is there any link betweeen continuous-time set propagation methods such as the ellipsoidal method or differential inequalities? and Can we link continuous and discrete-time set propagation methods?. Furthermore the last paragraph shows the need for a systematic analysis of these methods. In Chapters 3 and 4 these questions are addressed. In particular in Chapter 3 we provide a method for discrete-time set propagation using affine-set parameterizations and also analyze its stability properties. As it turns out, the problem of stability in set-propagation through dynamic systems is intimately related to the convergence properties of the chosen arithmetic. In Chapter 4 a unifying framework for the construction of continuous-time set propagation methods in the form of a generalized differential inequality (GDI) for support functions is presented. In particular the methods of differential inequalities and ellipsoidal calculus are special case of this GDI. Furthermore, this framework provides means for analyzing the convergence properties of such enclosures.

1.3 Robust Model Predictive Control

Model predictive control (MPC) refers to a class of feedback controllers, which proceed by solving, at each time step, an optimal control problem predicting the future behavior of

a dynamic system on a finite, receding time-horizon, using the current state measurement as initial condition [122]. The predicted optimal control trajectory is applied to the actual system until the next measurement becomes available, and the process is then repeated. The implementation of such controllers is based on a certainty-equivalence principle, whereby the future of the system is optimized as if neither external disturbances nor model mismatch were present, despite the fact that such disturbances and mismatch are the reason why feedback is needed in the first place.

The main advantage of certainty-equivalence in MPC is that the resulting optimization problems can often be solved efficiently, in real time [32, 51]. This approach works well in many practical applications, and it often exhibits a certain robustness due its inherent ability to reject disturbances [104, 156]. However, the constraints may become violated when large disturbances occur, since uncertainty is not taken into account in optimizing the predicted state trajectories. In such cases, robust MPC schemes can be used to mitigate these optimistic, certainty-equivalence-based predictions [122]. Nonetheless, a rigorous formulation of robust MPC calls for the solution, at each sampling time, of an optimization problem whose decision variables are the future control policies, that is, functions mapping the state measurements onto the control actions. Such optimization problems are hard to solve in general, and only for very short time-horizons can brute-force approximations, e.g. based on scenario trees [30, 36], be used currently. Because scenario-tree approaches scale exponentially with the length of the time-horizon, they may even be worse than robust dynamic programming approaches [12, 31, 122], which scale linearly with the length of the prediction horizon, yet exponentially with the state dimension.

Convex formulations of robust MPC have been derived for certain classes of problems, for instance when the dynamic system is jointly affine in the state, control and uncertainty and the feedback control law is itself affine in the disturbance [42, 41]. There, the number of the (matrix-valued) optimization variables scales quadratically with the length of the prediction horizon. The conservatism introduced by an affine parameterization of the control law is discussed in [146]. In this context, we also refer to [157], whereby real-time variants of robust MPC based on certain affine feedback laws are analyzed. Other convex formulations can be obtained by reformulating the semi-infinite constraints arising in robust MPC as linear matrix inequalities (LMIs). One such LMI reformulation for bounding the worst-case performance of linear systems under additive bounded uncertainty using constant state-feedback control laws was derived in [66]. Another approach was presented in [75], whereby the future model variations are bounded by a family of polytopes expressed as LMI constraints.

Other state-of-the-art approaches in robust MPC follow a set-theoretic perspective. These methods find their origins in viability theory [7, 68, 69] or, more specifically, in set-theoretic methods for control [17, 18]. Robust MPC schemes based on these parametric set-propagation methods are also known collectively under the name tube-based MPC. There, the predicted trajectory is replaced by a robust forward invariant tube (RFIT) in the state-space, namely a tube that encloses all possible state trajectories under a given feedback control law, which is independent of the uncertainty realization [72]. Tube-based approaches are typically analyzed under the assumption that exact state measurements are available [116], or that the equations of a parameterized state estimator, e.g. a linear filter, can be added to the system dynamics so that standard tube-based methods transfer readily [88].

A parameterized tube-based MPC formulation for linear discrete-time systems with affine uncertainty has been proposed in [117]. This formulation allows for the simultaneous optimization of tubes and control laws that are nonlinear in the state measurements, resulting in a computationally tractable, linear programming (LP) formulation, whose decision variables and constraints scale quadratically with the prediction horizon. A generalization handling more general cost functions is considered in [114], and a way of reducing the online complexity of this approach to linear complexity via offline computations is further presented in [118]. Tube-based methods have also been developed for linear systems with multiplicative uncertainty, for example by using polytopic tubes with quadratic cost, which leads to a quadratic programming (QP) formulation [37]. Regarding nonlinear dynamics, a possible tube-based approach involves linearizing the system around a feasible, but suboptimal, trajectory and computing the tube by regarding the linearization errors as additional uncertainty. This idea was used in [74] with polytopic tubes and affine feedback laws. A similar approach was developed by [21] in the case of quadratic cost terms and ellipsoidal tubes.

In Chapter 4 (cf. Section 4.7) we provide a characterization of robust forward invariant tubes developed to be used in tube-based MPC and also suited for robust optimal control problems. These tubes are characterized in terms of a min-max differential inequality, a nontrivial extension of the GDI. Furthermore, as a bi product of the min-max DI formulation we can compute (semi) explicit nonlinear feedback control laws.

1.4 Implicitly Defined Sets and Constraint Projection

Mathematical modelling has become an integral part of modern process design methodologies as well as in control system design and operations optimization. Often, these model-based methodologies involve the characterization of sets defined implicitly via systems of mixed equality and inequality constraints. In the analysis of dynamic models two problems that can be expressed in terms of implicitly defined sets are: guaranteed parameter estimation and guaranteed asymptotic analysis.

Given a model structure (in the form of parametric ODEs), parameter estimation —often referred to as model fitting— normally proceeds by determining parameter values for which the model predictions closely match the available process measurements. Failure to find an acceptable agreement calls for a revision of the model structure, before repeating the parameter estimation.

Most commonly, the parameter estimation problem is posed as an optimization problem that determines the parameter values minimizing the gap between the measurements and the model predictions, for instance in the least-square sense. Nonetheless, several factors can impair a successful and reliable estimation procedure. First of all, structural model mismatch is inherent to the modeling exercise, and it is illusive to look for the 'true' parameter values in this context. Even in the absence of model mismatch, fitting a set of experimental data exactly is generally not possible due to various sources of uncertainty. A measurement's accuracy is always tied to the resolution of the corresponding apparatus. Moreover, measured data are typically corrupted with noise, for instance Gaussian white noise or more generally colored noise.

Among the available approaches to account for uncertainty in parameter estimation, the focus in this thesis is on *guaranteed* parameter estimation [153], namely the determination of *all* parameter values—referred to as the solution set subsequently—that are consistent with the measurements under given uncertainty scenarios. Specifically, we consider the case that the uncertainty enters the estimation problem in the form of bounded measurement errors. An inherent advantage of this approach over more traditional parameter estimation is that no (consistent) solution to the problem will be lost, and this can help detect problems arising due to lack of identifiability. Moreover, the estimation process does not rely on a particular statistical description of the uncertainty, as is typically the case when applying maximum likelihood or Bayesian techniques. On the downside nonetheless, performing guaranteed parameter estimation turns out to be a very challenging and demanding task from a computational standpoint.

The second problem addressed in this thesis is that of guaranteed asymptotic analysis of a dynamic model. By asymptotic analysis we mean studying the long term behaviour of a dynamic system, in particular its equilibrium behaviour. This problem involves computing the solution manifold of a system of —generally nonlinear— algebraic equations. And by guaranteed —similarly to guaranteed parameter estimation— we mean every possible solution of the system in a given domain.

In the case that the algebraic system is fully determined— e.g. when all parameters are fixed and we are only interested in the equilibrim states— the solution set is a discrete space (0-manifold) [10]. In fact since this set is compact and discrete it is also finite. This means that the solution manifold is composed of a finite number of locally isolated solutions to the algebraic system. In this case, solutions can be approximated using iterative methods such as Newton's method and its variants. The success of Newton-type methods depends on the choice of initial guess and no guarantee can be made that all solutions have been found.

In the overdetermined case —e.g. when we are interested in the equilibrium solution of the model with respect to both state variables and parameters—numerical continuation methods [4, 5] can be used. The basic idea behind continuation-type methods is to consider some variables (typically state variables) as a function of some independent variables (parameters). The procedure starts from a known (fixed) point in the equilibrium manifold and as it name suggests approximately continuing the solution curve in the neighbourhood of that point. Computationally this can be achieved in a two-phase procedure. The first phase is the predictor step, where an approximate solution is found by advancing the initial fixed point by a small perturbation. In the second phase (corrector step) the algebraic condition is enforced using Newton-type methods with the predictor point as an initial guess. This, so-called natural parameter continuation tends to fail in the presence of singularities or turns in the solution curve which has lead to the development of refinements such as piecewise-linear (or simplicial) continuation, arc length and pseudo-arc length continuation (see e.g. [123] for an overview).

It is worth noticing that the solution set over a domain of interest is typically the union of possibly disconnected manifolds. In this case if one wishes to approximate every solution curve using continuation-type methods, it would be necessary to have an initial point for every curve.

In general, an exact characterization of implicitly defined sets (such as the solution set of guaranteed parameter estimation and guaranteed asymptotic analysis) is not possible and thus we have to resort to approximate methods. Constraint projection provides a framework for the construction of rigorous enclosures for implicitly defined sets. In fact, constraint

projection is concerned with the exhaustion of the set of values satisfying the system of implicit constraints and as a by product allows for lower-dimensional projections of this set [103].

Complete approaches to solve the constraint problem rely on exhaustive search algorithms e.g. branch and prune (B&P). Much as branch and bound methods for complete-search global optimization, B&P algorithms the computational domain is partitioned into progressively smaller domains where inclusion tests are performed to decide whether the elements in the domain satisfy the implicit constraints, else is removed from the search.

For nonlinear algebraic equations, the problem of approximating the solution set by a box partition, at an arbitrary precision, has been shown to be tractable using branch and prune with interval analysis [95], for instance using the set-inversion algorithm SIVIA [59].

In the context of guaranteed parameter estimation, the branch and prune approach has been extended to dynamic systems using ODE bounding techniques [e.g., 57, 111]. In a recent paper, Kieffer and Walter [61] have identified the main computational bottlenecks of set-inversion algorithms for guaranteed parameter estimation in dynamic systems to be: (i) the need for tight bounds on the solutions of the dynamic system; and, (ii) the need for efficient domain-reduction strategies as part of the exclusion tests. In this context, the use of ODE bounding techniques based on Taylor models has been investigated by Lin and Stadtherr [80] and Kletting et al. [63] using discrete-time bounding techniques and, more recently, by Paulen et al. [108] using a continuous-time approach. These authors have reported significant improvements in the convergence speed of the set-inversion algorithm compared to classical approaches based on interval enclosures. In principle, the higher the Taylor expansion order of the ODE solutions with respect to the uncertain parameters, the smaller the number of iterations required by the set-inversion algorithm to converge. Nonetheless, a higher-order expansion can incur a significant computational overhead, thereby defining a trade-off in terms of the overall computational burden with regards to the expansion order.

Another approach to enhancing the convergence of branch-and-prune algorithms for constraint projection involves applying contractors to the parameter boxes in order to reduce their width. Contractors based on optimality tests were derived by Jaulin et al. [58] using interval analysis, and later applied to dynamic system, e.g., by Kieffer and Walter [61]. Besides enabling higher-order convergence, the use of Taylor models to enclose the ODE solutions provides an explicit representation of parameter dependencies via the multivariate polynomial part. Lin and Stadtherr [80] and Kletting et al. [63] took advantage of this representation and used a constraint-propagation strategy in order to contract the parameter boxes.

1.5 Contribution and Overview of the Thesis

The next four chapters of this thesis present novel contributions in the area of estimation and control of uncertain dynamic systems.

Chapter 2 is concerned with methods to bound the image set of nonlinear vector-valued functions. In particular, a novel framework for the construction and analysis of set-valued arithmetics is presented. This framework relies on the use of so-called affine-set parameterizations, which allow the representation of sets such as intervals, ellipsoids, and polytopes as affine transformations of simpler sets such as unit norm-balls. Moreover, the affine-set parameterization framework allows for the construction of nonconvex set-valued arithmetics such as polynomial models with both interval and ellipsoidal remainders. In this chapter, we also provide sufficient conditions to analyze the convergence properties of set-valued arithmetics. Moreover, conditions to construct quadratically convergent arithmetics have been provided. The theoretical developments of this chapter have been used to construct an ellipsoidal arithmetic for vector-valued functions. Moreover, it has been shown that this ellipsoidal arithmetic is quadratically convergent. The material in this chapter has appeared in the following peer-reviewed publications:

[54] Houska, B., Villanueva, M. and Chachuat, B. (2015b). "Stable set-valued integration of nonlinear dynamic systems using affine set-parameterizations". *SIAM Journal on Numerical Analysis*, vol. 53 (5), pp. 2307–2328.

[151] Villanueva, M.E., Rajyaguru, J., Houska, B. and Chachuat, B. (2015). "Ellipsoidal arithmetic for multivariate systems". *12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering*, *Computer Aided Chemical Engineering*, vol. 37 (J.K.H. Krist V. Gernaey and R. Gani, eds.). Elsevier, pp. 767 – 772.

Chapter 3 deals with discrete-time methods for set propagation through parametric dynamic systems. In particular this chapter presents a novel algorithm that propagates generic affine-set parameterizations and whose images are guaranteed to enclose the reachable set of parametric ODEs. The algorithm reverses the classical two-phase approach of validated ODE integration by first constructing a predictor of the reachable set and then determines a step-size for which this predictor yields a valid enclosure. This reversed approach leads to a natural step-size control mechanism, which no longer relies on the availability of an a priori enclosure. Since the algorithm is generic, it encompasses the propagation of intervals, ellipsoids and polynomial models. Furthermore, an stability analysis of the enclosures computed

with the algorithm is presented. The main contribution of this chapter is to prove that, for a certain class of assymptotically stable parametric ODEs, locally stable and convergent enclosures can be obtained on infinite time horizons when the underlying affine set aritymetic converges quadratically in the Hausdorff sense. Another contribution of this chapter is on the use of ODE invariants in order to tighten the enclosures of the reachable set. The properties of the algorithm are illustrated with some numerical case studies. The material in this chapter has appeared in the following peer-reviewed publications:

[53] Houska, B., Villanueva, M.E. and Chachuat, B. (2013). "A validated integration algorithm for nonlinear ODEs using Taylor models and ellipsoidal calculus". *2013 IEEE 52nd Annual Conference on Decision and Control (CDC)*. pp. 484–489.

[54] Houska, B., Villanueva, M. and Chachuat, B. (2015b). "Stable set-valued integration of nonlinear dynamic systems using affine set-parameterizations". *SIAM Journal on Numerical Analysis*, vol. 53 (5), pp. 2307–2328.

[148] Villanueva, M.E., Houska, B. and Chachuat, B. (2014). "On the stability of set-valued integration for parametric nonlinear {ODEs}". *24th European Symposium on Computer Aided Process Engineering*, *Computer Aided Chemical Engineering*, vol. 33 (J. Klemes, P. Varbanov and P.Y. Liew, eds.). Elsevier, pp. 595 – 600.

[150] Houska, B., Villanueva, M. and Chachuat, B. (2015b). "Stable set-valued integration of nonlinear dynamic systems using affine set-parameterizations". *SIAM Journal on Numerical Analysis*, vol. 53 (5), pp. 2307–2328.

In Chapter 4 a framework for the construction and analysis of enclosures for the reachable set of nonlinear parametric ODEs using continuous-time methods is presented. This framework is based on the support function representation of convex sets. In particular a generalized differential inequality, which provides sufficient conditions for a time-varying support function to describe a convex enclosure for the reachable set is introduced. We show that this GDI encompases results from both the classical theory of differential inequalities and ellipsoidal set-propagation techniques as special cases. We also extend our construction to nonconvex sets characterized as the minkowski sum of a polynomial and a convex remainder (in particular polynomial models with interval and ellipsoidal remainders). The GDI is then extended to control systems with time-varying disturbances, and a minmax GDI is introduced to characterize robust forward invariant tubes for control systems with particular emphasis on tube-based MPC methods. As a by product of the minmax GDI,

feedback control laws can be derived in a semi-explicit manner. This minmax GDI is also used to construct practical tube-mpc approaches using robust forward invariant tubes with ellipsoidal cross sections. In this chapter we also analyze the convergence properties of continuous-time enclosures constructed using the GDI. Here again, we recover the classical linear and quadratic convergence results for differential inequalities and ellipsoidal set propagation respectively. Moreover, the convergence rate of nonconvex set propagation methods based on Taylor models is analyzed, showing that the Hausdorff convergence order of these enclosures is $q+1$ when qth order Taylor models are considered. The material in this chapter has appeared in (or has been submitted to) the following peer-reviewed publications (or journals):

[25] Chachuat, B. and Villanueva, M.E. (2012). "Bounding the solutions of parametric ODEs: when taylor models meet differential inequalities". *22nd European Symposium on Computer Aided Process Engineering*, *Computer Aided Chemical Engineering*, vol. 30 (I.D.L. Bogle and M. Fairweather, eds.). Elsevier, pp. 1307–1311.

[152] Villanueva, M.E., Paulen, R., Houska, B. and Chachuat, B. (2013). "Enclosing the reachable set of parametric ODEs using taylor models and ellipsoidal calculus". *23rd European Symposium on Computer Aided Process Engineering*, *Computer Aided Chemical Engineering*, vol. 32 (A.K. Turunen and Ilkka, eds.). Elsevier, pp. 979–984.

[149] Villanueva, M.E., Houska, B. and Chachuat, B. (2015a). "Unified framework for the propagation of continuous-time enclosures for parametric nonlinear ODEs". *J. Global Optim.*, vol. 62 (3), pp. 575–613.

[150] Villanueva, M.E., Quirynen, R., Diehl, M., Chachuat, B. and Houska B. (2016). "Robust MPC via Min-Max Differential Inequalities". *Automatica*, (Submitted).

Chapter 5 presents a constraint projection algorithm for computing enclosures of implicitly defined sets. This algorithm is based on branch and prune and thus belongs to the class of complete methods for global search. The algorithm is applied to two particular problems that occur in the analysis of dynamic systems: guaranteed parameter estimation and guaranteed asymptotic analysis. The constraint projection algorithm presented in this chapter is enhanced by the use of higher order bounding strategies such as the ones presented in Chapter 4 when dealing with guaranteed parameter estimation problems or the ones presented in Chapter 2 when dealing with guaranteed asymptotic analysis. The effect

of domain and CPU-time reduction strategies based on polyhedral relaxation of the polynomial model enclosures is also investigated. Finally, we propose a novel approach to mitigate the clustering effect in constraint projection problems whose defining constraints contain a system of overdetermined algebraic equations. This approach relies on the use of rigorous Newton-like methods and allows to express the boundary of a constraint set as the union of polynomial models. The material in this chapter has appeared (or has been accepted) in the following peer-reviewed publications:

[23] Chachuat, B., Houska, B., Paulen, R., Peric, N., Rajyaguru, J. and Villanueva, M.E. (2015). "Set-theoretic approaches in analysis, estimation and control of nonlinear systems." *9th {IFAC} Symposium on Advanced Control of Chemical Processes {ADCHEM}*, vol. 48. Whistler, Canada, pp. 981 – 995.

[108] Paulen, R., Villanueva, M. E., Fikar, M. and Chachuat, B. (2013). "Guaranteed parameter estimation in nonlinear dynamic systems using improved bounding techniques". *Proceedings of the 2013 European Control Conference (ECC'13)*. Zurich, Switzerland, pp. 4514–4519.

[109] Paulen, R., Villanueva, M.E. and Chachuat, B. (2013b). "Optimization-based domain reduction in guaranteed parameter estimation of nonlinear dynamic systems". *9th IFAC Symposium on Nonlinear Control Systems (NOLCOS)*, *Nonlinear Control Systems*, vol. 9. Toulouse, France, pp. 564 – 569.

[110] Paulen, R., Villanueva, M.E. and Chachuat, B. (2015). "Guaranteed parameter estimation of non-linear dynamic systems using high-order bounding techniques with domain and CPU-time reduction strategies". *IMA Journal of Mathematical Control and Information*, (Accepted).

Finally, although a conclusion is presented in every chapter of this thesis, Chapter 6 presents an overall conclusion of the work. Here, we emphasize the unifying aspect of this work by presenting some of the connections between the methods developed in each chapter. Some future research directions to advance set-theoretical methods are also proposed.

Chapter 2

Affine-Set Arithmetics

This Chapter deals with methods for the construction of enclosures for the image of a given host-set under a nonlinear vector-valued factorable function. The main aim of this Chapter is to provide a general framework for the construction and analysis of such methods. One of main issues we address in this chapter is how quickly these enclosures converge to the exact image-set as the host-set is shrinks.

The chapter is organized as follows, Section 2.1 introduces the building blocks of these framework: affine-set parameterizations and how typical enclosures such as intervals and polynomial models can be casted within this framework. Section 2.2 introduces factorable functions and methods to compute their affine-set extensions. This section also introduces the concepts related to the convergence of these extensions. Section 2.3 presents a method to construct quadratically convergent affine-set extensions for vector-valued function. In Section 2.4 presents a practical application of the affine-set parameterization formalism in the form of an ellipsoidal arithmetic. This arithmetic is shown to be quadratically convergent under mild assumptions. Finally, Section 2.5 concludes the Chapter. An implementation in MC++ of the concepts presented in this chapter can be found in Appendix D.

2.1 Affine-Set Parameterizations

There are certainly many ways to store a set in a computer. However, throughout this chapter the focus is on a particular class of computer represenable sets that are based on so-called affine set parameterizations.

Definition 2.1. *Let $\mathbb{E}_\ell \subseteq \mathbb{R}^\ell$ and $\mathbb{D}_{n,\ell} \subseteq \mathbb{R}^{n\times(\ell+1)}$, with $\ell \geq 1$ and $n \geq 1$. For any $Q \in \mathbb{D}_{n,\ell}$, we define the image of $\mathbb{E}_\ell$ under the affine map $\xi \mapsto Q(\xi^\mathsf{T}, 1)^\mathsf{T}$ as*

$$\mathrm{Im}_{\mathbb{E}_\ell}(Q) \;:=\; \{\, Q(\xi^\mathsf{T}, 1)^\mathsf{T} \mid \xi \in \mathbb{E}_\ell \,\} \;\subset\; \mathbb{R}^n.$$

In this set-representation, Q is referred to as the parameterization*, and the pair $(\mathbb{E}_\ell, \mathbb{D}_{n,\ell})$ is called an* affine set-parameterization*, with $\mathbb{E}_\ell$ and $\mathbb{D}_{n,\ell}$ the* basis set *and the* domain set*, respectively.*

The main practical motivation for this definition is that we can store the finite dimensional parameter $Q \in \mathbb{D}_{n,\ell}$ rather than the associated set $\mathrm{Im}_{\mathbb{E}_\ell}(Q)$ itself. Usual families of convex sets such as intervals, ellipsoids, zonotopes or polytopes can all be represented using affine set-parameterizations with convex basis sets.

Example 2.1. *Every ℓ-dimensional ellipsoid in $\mathbb{R}^n$ can be represented using an affine set-parameterization with the basis set*

$$\mathbb{E}_\ell^{\mathrm{ball}} \;:=\; \left\{ \xi \in \mathbb{R}^\ell \,\middle|\, \|\xi\|_2 \leq 1 \right\},$$

and the associated domain set $\mathbb{R}^{n\times(\ell+1)}$. Likewise, every polytope and every zonotope in $\mathbb{R}^n$ can be represented using affine set-parameterizations with the basis sets

$$\mathbb{E}_\ell^{\mathrm{simplex}} \;:=\; \left\{ \xi \in \mathbb{R}_+^\ell \,\middle|\, \|\xi\|_1 \leq 1 \right\} \quad \textit{and} \quad \mathbb{E}_\ell^{\mathrm{box}} \;:=\; \left\{ \xi \in \mathbb{R}^\ell \,\middle|\, \|\xi\|_\infty \leq 1 \right\},$$

respectively, and the same domain set $\mathbb{R}^{n\times(\ell+1)}$. The latter basis set $\mathbb{E}_\ell^{\mathrm{box}}$ can also be associated with the domain

$$\mathbb{D}_\ell^{\mathrm{interval}} \;:=\; \left\{ (\mathrm{diag}\,(r), c) \,\middle|\, r \in \mathbb{R}_+^\ell, c \in \mathbb{R}^\ell \right\}$$

in order to describe interval boxes in $\mathbb{R}^\ell$ with radius r and centered at c. ◇

Affine set-parameterizations can also be used to describe certain classes of nonconvex sets by considering nonconvex basis sets.

Example 2.2. *The affine set-parameterization $(\mathbb{E}_\ell^{\mathrm{pol}(q)}, \mathbb{R}^{n\times(\alpha_\ell^{(q)}+1)})$ with*

$$\mathbb{E}_\ell^{\mathrm{pol}(q)} \;:=\; \left\{ M_{\ell,q}(\xi) \,\middle|\, \xi \in [-1,1]^\ell \right\},$$

where $M_{\ell,q}(\xi) \in \mathbb{R}^{\alpha_\ell^{(q)}}$ *is the vector containing the first* $\alpha_\ell^{(q)}$ *monomials in* ξ *in lexicographic order,*

$$M_{\ell,q}(\xi) := (\,\xi_1, \xi_2, \ldots, \xi_\ell, \xi_1^2, \xi_1\xi_2, \ldots, \xi_\ell^2, \xi_1^3, \ldots, \xi_\ell^q\,)^\top\,, \tag{2.1}$$

describes nonconvex sets in $\mathbb{R}^n$*—this parameterization shall be referred to as the class of* qth-order polynomial models *subsequently. Clearly, other bases than the monomial basis* $M_{\ell,q}(\xi)$ *can be used in* (2.1)*, such as the Legendre basis or the Chebyshev basis. More generally, any qth-order polynomial model can be combined with a convex set, e.g., in the manner of the remainder term in a Taylor model [103] or Chebyshev model [113] to account for higher-order terms. Considering interval boxes as the convex set, such a combination leads to the parameterization* $(\mathbb{E}_\ell^{\mathrm{pol}(q)} \times \mathbb{E}_n^{\mathrm{box}}, \mathbb{R}^{n\times\alpha_\ell^{(q)}} \times \mathbb{D}_n^{\mathrm{interval}})$*; with ellipsoids likewise, it leads to the affine set-parameterization* $(\mathbb{E}_\ell^{\mathrm{pol}(q)} \times \mathbb{E}_n^{\mathrm{ball}}, \mathbb{R}^{n\times(\alpha_\ell^{(q)}+n+1)})$*. In the latter construct, the first n-by-*$\alpha_\ell^{(q)}$ *block of the parameter* $Q \in \mathbb{R}^{n\times(\alpha_\ell^{(q)}+n+1)}$ *comprises the multivariate polynomial coefficients; the following n-by-n block, the shape matrix coefficients of the ellipsoid; and the last n-by-*1 *block, the center of the ellipsoid.* ◇

Of special interest for the stability analysis conducted in the next chapter of this thesis is the concept of invariance under affine transformation, as defined next.

Definition 2.2. *An affine set-parameterization* $(\mathbb{E}_\ell, \mathbb{D}_{n,\ell})$ *is said to be* invariant under affine transformation *if for every affine map* $x \mapsto Ax+b$ *with* $(A,b) \in \mathbb{R}^{m\times(n+1)}$ *and every* $Q \in \mathbb{D}_{n,\ell}$*,*

$$\exists Q' \in \mathbb{D}_{m,\ell}: \quad \mathrm{Im}_{\mathbb{E}_\ell}\left(Q'\right) \;=\; \left\{Ax+b \;\middle|\; x \in \mathrm{Im}_{\mathbb{E}_\ell}\left(Q\right)\right\} \;=\; \mathrm{Im}_{\mathbb{E}_\ell}\left(AQ\right) \oplus \{b\}\,.$$

Remark 2.1. *An immediate consequence of the property of invariance under affine transformation is that the image of* $\mathbb{E}_\ell$ *under the composition of two affine transformations, say given by the parameterizations* $Q \in \mathbb{D}_{n,\ell}$ *and* $P \in \mathbb{R}^{m\times(n+1)}$*, can* always *be represented exactly as the image of* $\mathbb{E}_\ell$ *under a third affine transformation with* $Q' \in \mathbb{D}_{m,\ell}$ *such that*

$$\mathrm{Im}_{\mathbb{E}_\ell}\left(Q'\right) \;=\; \mathrm{Im}_{\mathrm{Im}_{\mathbb{E}_\ell}(Q)}\left(P\right)\,.$$

Among the convex and nonconvex set representations considered in Examples 2.1 and 2.2, some, but not all, are invariant under affine transformation.

Example 2.3. *Given an ellipsoid* $\mathrm{Im}_{\mathbb{E}_\ell^{\mathrm{ball}}}(Q)$ *with* $Q \in \mathbb{R}^{n\times(\ell+1)}$*, the application of any affine transformation* $x \mapsto Ax+b$ *with* $(A,b) \in \mathbb{R}^{m\times(n+1)}$ *yields another ellipsoid* $\mathrm{Im}_{\mathbb{E}_\ell^{\mathrm{ball}}}(Q')$ *with*

$$Q' := AQ + (0_{m\times\ell}, b) \in \mathbb{R}^{m\times(\ell+1)}.$$

Therefore, the class of ellipsoids is invariant under affine transformation. The same parameters Q' *can be used to show that the classes of polytopes, zonotopes and polynomial models—as well as any finite combination of these parameterizations—are also invariant under affine transformation.* ◇

Example 2.4. *The rotation of an interval box in* $\mathbb{R}^n$ *may yield another interval box whose edges are no longer aligned with the original axes, in which case the transformed box cannot be represented* exactly *in terms of the affine parameterization* $(\mathbb{E}_n^{\mathrm{box}}, \mathbb{D}_n^{\mathrm{interval}})$ *anymore. In other words, the class of interval boxes is not invariant under affine transformation, which is one of the main sources of the wrapping effect in interval analysis. It also follows that any affine set-parameterizations obtained from the combination with interval boxes, e.g. polynomial models with interval boxes as in Example 2.2, will fail to be invariant under affine transformation.* ◇

2.2 Factorable Functions and their Affine-Set Extensions

Before introducing the main results of this section, We need to define what a factorable function is. This concept is formalized in the following definition.

Definition 2.3. *A function* $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ *is called factorable if it can be decomposed into a finite recursive composition of a given finite library of atom operations* $a_i : \mathbb{R}^\kappa \to \mathbb{R}$*, where* $\kappa \in \{1,2\}$*. In particular, the atom operations* a_i *considered are bivariate sum, bivariate product and univariate operations.*

Originally introduced by McCormick [89] for the development of a convex/concave relaxation arithmetic, factorable functions cover an extremely inclusive class of functions which can be represented finitely on a computer by means of a code list or a computational graph involving atom operations. These are typically unary and binary operations within a library of atom operators, which can be based for example on the C-code library `math.h`. Besides convex/concave relaxations, factorable functions find applications in automatic differentiation (AD) [43, 98] as well as in interval analysis [97] and Taylor model arithmetic [103].

A factorable function $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ can be evaluated by a computer code taking a vector $x \in \mathbb{R}^n$, setting $u^0(x) = x$ and applying the recursive rule

$$\forall i \in \{1,\ldots,N\}, \quad u^i(x) \;=\; g_i(u^{i-1}(x)) \;:=\; \begin{pmatrix} u^{i-1}(x) \\ a_i(u^{i-1}(x)) \end{pmatrix}. \tag{2.2}$$

Then, we have

$$\varphi(x) = Pu^N(x) = P[g_N \circ g_{N-1} \circ \ldots \circ g_1](x) \tag{2.3}$$

where $P \in \mathbb{R}^{n_f \times (n_x+N)}$ is a projection matrix selecting the appropriate components. The j-th component of $u^i(x)$ is denoted as $u^i_j(x)$ subsequently.

Definition 2.4. *Consider a function $\varphi : \mathbb{R}^n \to \mathbb{R}^m$, and let $(\mathbb{E}_1, \mathbb{D}_1)$ and $(\mathbb{E}_2, \mathbb{D}_2)$ be two affine set-parameterizations. The function $\varphi^{\mathbb{E}_1,\mathbb{E}_2} : \mathbb{D}_1 \to \mathbb{D}_2$ is called an* extension of φ from $\mathbb{D}_1$ to $\mathbb{D}_2$ *if*

$$\forall Q \in \mathbb{D}_1, \quad \mathrm{Im}_{\mathbb{E}_2}\left(\varphi^{\mathbb{E}_1,\mathbb{E}_2}(Q)\right) \;\supseteq\; \overline{\varphi}^{\mathbb{E}_1}(Q),$$

where $\overline{\varphi}^{\mathbb{E}_1}(Q) := \left\{\; \varphi(x) \;\middle|\; x \in \mathrm{Im}_{\mathbb{E}_1}(Q) \;\right\}$ denotes the exact image of φ on $\mathrm{Im}_{\mathbb{E}_1}(Q)$. In the special case that $\mathbb{E}_1 = \mathbb{E}_2 =: \mathbb{E}$, we use the shorthand notation $\varphi^{\mathbb{E}}$.

A special notation is also defined for extensions of the binary addition operation for convenience. We shall only use this notation when it is clear from the context what the basis and domain of the affine set-parameterization are.

Definition 2.5. *Let $(\mathbb{E}, \mathbb{D})$ be an affine set-parameterization. An extension of the binary addition is a function $\uplus : \mathbb{D} \times \mathbb{D} \to \mathbb{D}$ such that*

$$\forall Q, Q' \in \mathbb{D}, \quad \mathrm{Im}_{\mathbb{E}}\left(Q \uplus Q'\right) \;\supseteq\; \left\{\; x + x' \;\middle|\; x \in \mathrm{Im}_{\mathbb{E}}(Q)\;,\; x' \in \mathrm{Im}_{\mathbb{E}}\left(Q'\right) \;\right\}.$$

Moreover, $\uplus$ is said to be a regular *addition extension if*

$$\forall Q, Q' \in \mathbb{D}, \quad d_{\mathrm{H}}\left(\mathrm{Im}_{\mathbb{E}}\left(Q \uplus Q'\right), \mathrm{Im}_{\mathbb{E}}(Q)\right) \;=\; \mathbf{O}\left(\left\|\mathrm{Im}_{\mathbb{E}}\left(Q'\right)\right\|_H\right). \tag{2.4}$$

Note that (2.4) only imposes a mild regularity condition, which is automatically satisfied when using either interval arithmetic [97] or Polynomial model arithmetic [15, 13, 103, 113, 34].

The regularity condition imposed by Equation (2.4) can also be satisfied with ellipsoidal parameterizations. In order to see this consider ellipsoidal set-parameterizations of the form $\left(\mathbb{E}^{\mathrm{ball}}_{\ell}, \mathbb{R}^{n\times(\ell+1)}\right)$, with $\ell = n$ in order to represent ellipsoids with non-zero volume in $\mathbb{R}^n$—

see Example 2.1. Given any two set-parameterizations $(R_1, r_1), (R_2, r_2) \in \mathbb{R}^{n\times(n+1)}$, a regular addition extension $\uplus$ is given by

$$(R_1, r_1) \uplus (R_2, r_2) \;:=\; \left(\left[\frac{1}{\lambda_1} R_1^\top R_1 + \frac{1}{\lambda_2} R_2^\top R_2 \right]^{1/2}, r_1 + r_2 \right), \tag{2.5}$$

where $[\cdot]^{1/2}$ denotes any (e.g., the symmetric) matrix square-root, and

$$\lambda_1 = 1 - \lambda_2 = \frac{\sqrt{\operatorname{tr} R_1^\top R_1 + \varepsilon}}{\sqrt{\operatorname{tr} R_1^\top R_1 + \varepsilon} + \sqrt{\operatorname{tr} R_2^\top R_2 + \varepsilon}},$$

with $\varepsilon > 0$ a small numerical regularization to prevent division by zero in degenerate cases; see, e.g., [56, 69] for a proof.

Other set-valued functions of interest are the so-called image bounding functions. These functions allow the construction of sets enclosing the original set.

Definition 2.6. *The image bounding function (or image bounder) from a parameterization* $(\mathbb{E}_1, \mathbb{D}_1)$ *to a parameterization* $(\mathbb{E}_2, \mathbb{D}_2)$ *is a function* $\mathcal{B}^{\mathbb{E}_1,\mathbb{E}_2} : \mathbb{D}_1 \to \mathbb{D}_2$ *such that*

$$\forall Q \in \mathbb{D}_1 : \quad \operatorname{Im}_{\mathbb{E}_1}(Q) \subseteq \operatorname{Im}_{\mathbb{E}_2}\left(\mathcal{B}^{\mathbb{E}_1,\mathbb{E}_2}(Q) \right)$$

Example 2.5. *The interval hull of a set is a well known and widely used image bounder. Consider the parameterizations* $(\mathbb{E}_n^{\text{ball}}, \mathbb{D}_n^{\text{ball}})$, $(\mathbb{E}_n^{\text{box}}, \mathbb{D}_n^{\text{box}})$ *and* $Q \in \mathbb{D}_n^{\text{ball}}$. *The interval hull of the ellipsoid parameterized by* Q *is given by*

$$\mathcal{B}^{\mathbb{E}_n^{\text{ball}},\mathbb{E}_n^{\text{box}}}(Q) = Q' := \operatorname{diag}\left(\sqrt{Q_{i,i}} \right)_{1\le i\le n}. \tag{2.6}$$

Clearly we have $\operatorname{Im}_{\mathbb{E}_n^{\text{ball}}}(Q) \subseteq \operatorname{Im}_{\mathbb{E}_n^{\text{box}}}(Q')$.

A key property of the affine set-parameterization extension of a function is how much overestimation it carries with respect to the actual image set of that function. Especially relevant for the methods developed in this thesis, is the Hausdorff convergence order.

Definition 2.7. *The extension* $\varphi^{\mathbb{E}_1,\mathbb{E}_2} : \mathbb{D}_1 \to \mathbb{D}_2$ *of a function* $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ *is said to have* Haussdorf convergence order $q \ge 1$ *if*

$$\forall Q \in \mathbb{D}_1, \quad d_{\text{H}}\left(\operatorname{Im}_{\mathbb{E}_2}\left(\varphi^{\mathbb{E}_1,\mathbb{E}_2}(Q) \right), \overline{\varphi}^{\mathbb{E}_1}(Q) \right) \;=\; \mathbf{O}\left(\operatorname{diam}\left(\operatorname{Im}_{\mathbb{E}_1}(Q) \right)^q \right).$$

In case an affine set-parameterization is *not* invariant under affine transformation, constructing extensions that have Hausdorff convergence order two or higher is not possible in general. This is even so for extensions of simple linear functions, as illustrated below in the case of interval analysis.

Example 2.6. *Consider the affine set-parameterization* $(\mathbb{E}_2^{\text{box}}, \mathbb{D}_2^{\text{interval}})$ *and let the matrix* $Q_\delta := (\text{diag}(\delta\,,\,\delta), (0\,,\,0)^\mathsf{T}) \in \mathbb{D}_2^{\text{interval}}$. *Notice that* $\text{Im}_{\mathbb{E}_2^{\text{box}}}(Q_\delta)$ *describes a 2d-box centered at the origin and of radius* δ*, the diameter of which is thus in* $\mathbf{O}(\delta)$. *Now, consider the linear transformation* $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ *such that*

$$\forall x \in \mathbb{R}^2, \quad \varphi(x) := \frac{1}{\sqrt{2}} \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix} .$$

The exact image of Q_δ *under* φ *is given by*

$$\overline{\varphi}^{\mathbb{E}_2^{\text{box}}}(Q_\delta) := \{Ax \mid x \in [-\delta, \delta] \times [-\delta, \delta]\} \quad \textit{with} \quad A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} ,$$

which is a 45° *rotation of* $\text{Im}_{\mathbb{E}_2^{\text{box}}}(Q_\delta)$ *in the plane, around the origin. On the other hand, the natural interval extension* $\varphi^{\mathbb{E}_2^{\text{box}}}$ *of* φ *at* Q_δ *gives*

$$\varphi^{\mathbb{E}_2^{\text{box}}}(Q_\delta) := \sqrt{2}\, \text{Im}_{\mathbb{E}_2^{\text{box}}}(Q_\delta) ,$$

which is also the interval hull of $\overline{\varphi}^{\mathbb{E}_2^{\text{box}}}(Q_\delta)$. *Nonetheless, the Hausdorff distance between the sets* $\varphi^{\mathbb{E}_2^{\text{box}}}(Q_\delta)$ *and* $\overline{\varphi}^{\mathbb{E}_2^{\text{box}}}(Q_\delta)$ *is such that*

$$\max_{x \in \varphi^{\mathbb{E}_2^{\text{box}}}(Q_\delta)} \; \min_{y \in \overline{\varphi}^{\mathbb{E}_2^{\text{box}}}(Q_\delta)} \|x - y\|_\infty \;=\; (\sqrt{2} - 1)\, \delta \;=\; \mathbf{O}(\delta) ,$$

that is, the natural interval extension of φ *has Hausdorff convergence order 1.* ◇

Remark 2.2. *The affine set-parameterization* $(\mathbb{E}_\ell^{\text{pol}(q)} \times \mathbb{E}_n^{\text{box}}, \mathbb{R}^{n \times \alpha_\ell^{(q)}} \times \mathbb{D}_n^{\text{interval}})$*—namely the class of polynomial models combined with interval boxes—is not invariant under affine transformation for the same reason as in Example 2.6 before. Therefore, extensions of certain functions may only have Hausdorff convergence order 1 for such parameterization, regardless of the polynomial order q. This is the case, for instance, when using Taylor model arithmetic [103] to construct an extension* $\varphi^{\mathbb{E}_\ell^{\text{pol}(q)} \times \mathbb{E}_n^{\text{box}}, \mathbb{E}_\ell^{\text{pol}(q)} \times \mathbb{E}_m^{\text{box}}}$ *from* $\mathbb{R}^{m \times \alpha_\ell^{(q)}} \times \mathbb{D}_n^{\text{interval}}$ *to* $\mathbb{R}^{m \times \alpha_\ell^{(q)}} \times \mathbb{D}_m^{\text{interval}}$ *of a given factorable and* $(q+1)$*-times continuously differentiable function* $\varphi : \mathbb{R}^n \to \mathbb{R}^m$. *It is worth mentioning that the foregoing observation is not in contradic-*

tion with the convergence analysis in [20], which shows that an extension $\varphi^{\mathbb{E}_\ell^{\mathrm{pol}(q)},\mathbb{E}_\ell^{\mathrm{pol}(q)}\times\mathbb{E}_m^{\mathrm{box}}}$ *from* $\mathbb{R}^{n\times\alpha_\ell^{(q)}}$ *to* $\mathbb{R}^{m\times\alpha_\ell^{(q)}}\times\mathbb{D}_m^{\mathrm{interval}}$ *of* φ *in the form of a qth-order Taylor model will have Hausdorff convergence order* $q+1$*. In the present, more general, context, it is indeed the wrapping of the interval box by the function that causes the loss of convergence order.*

The following result is given for future reference, as it will be used in the stability analysis of set-valued integrators.

Proposition 2.1. *Let the extension* $\varphi^{\mathbb{E}_1,\mathbb{E}_2}:\mathbb{D}_1\to\mathbb{D}_2$ *of a continuously-differentiable function* $\varphi:\mathbb{R}^n\to\mathbb{R}^m$ *have quadratic Hausdorff convergence. Then,*

$$d_{\mathrm{H}}\left(\ \mathrm{Im}_{\mathbb{E}_2}\left(\varphi^{\mathbb{E}_1,\mathbb{E}_2}(Q)\right)\ ,\ \varphi(\xi)+\frac{\partial\varphi}{\partial x}(\xi)\cdot[\mathrm{Im}_{\mathbb{E}_1}(Q)-\xi]\ \right) = \mathbf{O}\left(\|\mathrm{Im}_{\mathbb{E}_1}(Q)-\xi\|_{\mathrm{H}}^2\right)\ , \tag{2.7}$$

for all $Q\subseteq\mathbb{D}_1$ *and all* $\xi\in\mathbb{R}^n$ *with* $\|\mathrm{Im}_{\mathbb{E}_1}(Q)-\xi\|_{\mathrm{H}}$ *sufficiently small.*

Proof. Without loss of generality, we assume $\xi=0$ as we can always shift the sets by a constant offset. The main idea of the proof is to exploit the triangle inequality for the Hausdorff metric, which yields

$$\begin{aligned} &d_{\mathrm{H}}\left(\ \mathrm{Im}_{\mathbb{E}_2}\left(\varphi^{\mathbb{E}_1,\mathbb{E}_2}(Q)\right)\ ,\ \varphi(0)+\frac{\partial\varphi}{\partial x}(0)\cdot[\mathrm{Im}_{\mathbb{E}_1}(Q)]\ \right)\\ &\quad\le\ d_{\mathrm{H}}\left(\ \mathrm{Im}_{\mathbb{E}_2}\left(\varphi^{\mathbb{E}_1,\mathbb{E}_2}(Q)\right)\ ,\ \overline{\varphi}^{\mathbb{E}_1}(Q)\right)+d_{\mathrm{H}}\left(\ \overline{\varphi}^{\mathbb{E}_1}(Q)\ ,\ \varphi(0)+\frac{\partial\varphi}{\partial x}(0)\cdot[\mathrm{Im}_{\mathbb{E}_1}(Q)]\ \right)\ . \end{aligned}$$

By assumption, we have

$$d_{\mathrm{H}}\left(\ \mathrm{Im}_{\mathbb{E}_2}\left(\varphi^{\mathbb{E}_1,\mathbb{E}_2}(Q)\right)\ ,\ \overline{\varphi}^{\mathbb{E}_1}(Q)\right)\ =\ \mathbf{O}\left(\|\mathrm{Im}_{\mathbb{E}_1}(Q)\|_{\mathrm{H}}^2\right)\ ,$$

and, by Taylor's theorem, we also have

$$d_{\mathrm{H}}\left(\ \overline{\varphi}^{\mathbb{E}_1}(Q)\ ,\ \varphi(0)+\frac{\partial\varphi}{\partial x}(0)\cdot[\mathrm{Im}_{\mathbb{E}_1}(Q)]\ \right)\ =\ \mathbf{O}\left(\|\mathrm{Im}_{\mathbb{E}_1}(Q)\|_{\mathrm{H}}^2\right)\ ,$$

which completes the proof □

2.2.1 Affine-Set Extensions Using Polynomial Models

The construction of extensions can be automated for factorable functions using a variety of arithmetics, which can be conveniently implemented in computer programs. Unlike inter-

val arithmetic, Taylor and Chebyshev model arithmetics can be used to construct extension functions that enjoy higher-order Hausdorff convergence. The idea is to propagate the polynomial part (expressed either in monomial or Chebyshev basis) by symbolic calculations wherever possible, and processing the remainder term as well as the higher-order terms according to the rules of a specific arithmetic.

First, we will consider real-valued functions. Given a $(q+1)$-times continuously-differentiable function $\varphi : \mathbb{R}^n \to \mathbb{R}$ on a set $X \in \mathbb{IR}^n$, a qth-order Taylor model of φ on X at a point $\hat{x} \in X$ is the pair $\left(\mathcal{P}^q_{\varphi,X}, \mathcal{R}^q_{\varphi,X}\right)$ of a qth-order multivariate polynomial $\mathcal{P}_{\varphi,X} : \mathbb{R}^n \to \mathbb{R}$ with an interval remainder $\mathcal{R}^q_{\varphi,X} \in \mathbb{IR}$ satisfying

$$\forall x \in X : \quad \varphi(x) - \mathcal{P}_{\varphi,X}(p) \in \mathcal{R}^q_{\varphi,X} \quad \text{and} \,, \qquad \mathcal{P}^q_{\varphi,X}(x) = \sum_{\substack{\gamma \in \mathbb{N}^n, \\ |\gamma| \leq q}} \frac{\partial^\gamma \varphi(\hat{x})}{\gamma!} (x - \hat{x})^\gamma \,.$$

Likewise $\left(\mathcal{P}^q_{\varphi,X}, \mathcal{R}^q_{\varphi,X}\right)$ is called a qth order Chebyshev model of φ on X if it satisfies

$$\forall x \in X : \quad \varphi(x) - \mathcal{P}_{\varphi,X}(p) \in \mathcal{R}^q_{\varphi,X} \quad \text{and} \,, \qquad \mathcal{P}^q_{\varphi,X}(x) = \sum_{\substack{\gamma \in \mathbb{N}^n, \\ |\gamma| \leq q}} c_\gamma(X) T_\gamma(x^s) \,,$$

with $\left\| c_\gamma(X) \right\|_{\mathrm{H}} = \mathbf{O}\left(\operatorname{diam}(X)^{|\gamma|}\right)$ and $x_i^s := \frac{x - \operatorname{mid}()}{\operatorname{rad}(X_i)}$.

Rules for binary sum, binary product and univariate composition between Taylor models or Chebyshev models have been described and analyzed, e.g., in Makino and Berz [85], Bompadre et al. [20], Dzetkulič [34], Rajyaguru et al. [113]. As well as enabling the computation of Taylor and Chebyshev models for factorable functions, these rules guarantee high-order convergence of the remainder term to zero with the diameter of the parameter host set P as

$$\left\| \mathcal{R}^q_{\varphi,X} \right\|_{\mathrm{H}} = \mathbf{O}\left(\operatorname{diam}(X)^{q+1}\right) . \tag{2.8}$$

Naturally, the same approach applies to vector-valued functions $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ by treating each function component separately. In the sense of the affine set-arithmetic formalism presented in this chapter, the extension of a vector-valued function retains the convergence property (2.8) when the extension is of the form $\varphi^{\mathbb{E}_n^{\mathrm{pol}(q)}, \mathbb{E}_m^{\mathrm{pol}(q)} \times \mathbb{E}_m^{\mathrm{box}}}$.

In connection to the affine set-parameterization formalism introduced in Sect. 2.2, Taylor and Chebyshev model arithmetics support the constructions of extensions for a vector-

valued function $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ in the form $\varphi^{\mathbb{E}_\ell^{\mathrm{pol}(q)} \times \mathbb{E}_m^{\mathrm{box}}} : \mathbb{R}^{n \times \alpha_\ell^{(q)}} \times \mathbb{D}_n^{\mathrm{interval}} \to \mathbb{R}^{m \times \alpha_\ell^{(q)}} \times \mathbb{D}_m^{\mathrm{interval}}$. In this sense it is also important to recall Remark 2.2, since the extension may only have Hausdorff convergence order 1, regardless of the order of the polynomial.

Finally, because the image set of a multivariate polynomial of order 2 or higher is non-convex in general, applications of Taylor/Chebyshev models often require a bounding function as given by Definition 2.6. Bounding functions can be constructed using constant/affine bounds or convex/polyhedral enclosures.

- Tight interval bounds can be obtained using LMI methods [73]. Other ways of deriving rigorous interval bounds involve exact bounding of the polynomial's first- and second-order terms [85, 80] or expressing the polynomial in Bernstein bases [77].

- Affine bounds can be obtained likewise by retaining the first-order term, while bounding all of the other terms using one of the foregoing approaches.

- Polyhedral enclosures can be obtained on the application of the reformulation-linearization technique (RLT) developed by Sherali and Fraticelli [136] and Sherali et al. [135]. Other approaches to convex/polyhedral enclosures include the decomposition/relaxation/outer-approximation technique [139, 142] as well as McCormick's relaxation technique [89, 93].

2.2.2 Construction of Affine-Set Extensions of Factorable Functions via Lifting

Another way of constructing set extensions is by exploiting the recursive method for evaluating a factorable function, given by Equations (2.2), (2.3). Given a parameterization $(\mathbb{E}_n^1, \mathbb{D}_n^1)$ a factorable function $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ decomposed into $\varphi = P[g_N \circ g_{N-1} \circ \ldots \circ g_1]$ with factors $g_1, \ldots, g_N$ and parameterizations $(\mathbb{E}_{n+1}^1, \mathbb{D}_{n+1}^1), \ldots (\mathbb{E}_{n+N}^N, \mathbb{D}_{n+N}^N)$ we aim compute an $\mathbb{E}_n^1, \mathbb{E}_m^2$-extension of φ on the parameter $Q \in \mathbb{D}_n^1$. This can be achieved by recursively evaluating

$$\forall i \in \{1, \ldots, N\} : \quad Q^i = g_i^{\mathbb{E}_{n+i-1}^{i-1}, \mathbb{E}_{n+i}^i}\left(Q^{i-1}\right) \tag{2.9}$$

with $Q^0 := Q$, $\mathbb{E}_n^0 := \mathbb{E}_n^1$, $\mathbb{D}_n^0 := \mathbb{D}_n^1$ and $Q^i \in \mathbb{D}_{n+i}^i$. Setting $\mathbb{E}_m^2 = \mathbb{E}_{n+N}^N$ and $\mathbb{D}_m^2 = \mathbb{D}_{n+N}^N$ the $\mathbb{E}_n^1, \mathbb{E}_m^2$-extension of φ on Q can then be obtained as

$$\varphi^{\mathbb{E}_n^1, \mathbb{E}_m^2} := PQ^N \, . \tag{2.10}$$

An important property of the evaluation method given by Equations (2.9), (2.10) is that the $\mathbb{E}_n^1, \mathbb{E}_m^2$-extension of a factorable function inherits the image Hausdorff convergence order of the $\mathbb{E}_{n+i-1}^{i-1}, \mathbb{E}_{n+i}^{i}$-extension of its factors as elaborated in the following Lemma.

Lemma 2.1. *If all atom operations a_i of a factorable function φ are Lipschitz continuous and if the $\mathbb{E}_{n+i-1}^{i-1}, \mathbb{E}_{n+i}^{i}$-extensions $g_i^{\mathbb{E}_{n+i-1}^{i-1}, \mathbb{E}_{n+i}^{i}}$ of all the factors g_i have Hausdorff convergence order q, then the $\mathbb{E}_n^1, \mathbb{E}_m^2$-extension $\varphi^{\mathbb{E}_n^1, \mathbb{E}_m^2}$ implemented given by Equations* (2.9)*,* (2.10) *has also Hausdorff convergence order q.*

Proof. The proof of this Lemma follows by induction over the atom operations showing recursively that the function

$$u^i := g_i \circ g_{i-1} \circ \ldots \circ g_1$$

and its associated $\mathbb{E}$-extension

$$u^{i,(\mathbb{E}_n^0, \mathbb{E}_{n+i}^i)} := g_i^{\mathbb{E}_{n+i-1}^{i-1}, \mathbb{E}_{n+i}^i} \circ g_{i-1}^{\mathbb{E}_{n+i-2}^{i-2}, \mathbb{E}_{n+i-1}^{i-1}} \circ \ldots \circ g_1^{\mathbb{E}_n^0, \mathbb{E}_{n+1}^1}$$

have Haussdorff convergence order q. Notice that if a_{i-1} is Lipschitz continuous, g_i is also Lipschitz continuous. Here, the induction start is trivial, as the function $u^{1,(\mathbb{E}_n^0, \mathbb{E}_{n+1}^1)} = g_1^{\mathbb{E}_n^0, \mathbb{E}_{n+1}^1}$ has by assumption Haussdorff convergence order q. Next, the induction step can be verified by using the triangular inequality for the Haussdorff metric, which yields

$$\begin{aligned}
&d_{\mathrm{H}}\left(\mathrm{Im}_{\mathbb{E}_{n+i+1}^{i+1}}\left(u^{i+1,(\mathbb{E}_n^0, \mathbb{E}_{n+i+1}^{i+1})}(Q)\right), u^{i+1}\left(\mathrm{Im}_{E_n^0}(Q)\right)\right) \\
&\quad = d_{\mathrm{H}}\left(\mathrm{Im}_{\mathbb{E}_{n+i+1}^{i+1}}\left(g_{i+1}^{\mathbb{E}_{n+i-1}^{i}, \mathbb{E}_{n+i+1}^{i+1}}\left(u^{i,(\mathbb{E}_n^0, \mathbb{E}_{n+i}^i)}(Q)\right)\right), g_{i+1}\left(u^i\left(\mathrm{Im}_{\mathbb{E}_n^0}(Q)\right)\right)\right) \\
&\quad \leq d_{\mathrm{H}}\left(\mathrm{Im}_{\mathbb{E}_{n+i+1}^{i+1}}\left(g_{i+1}^{\mathbb{E}_{n+i}^{i}, \mathbb{E}_{n+i+1}^{i+1}}\left(u^{i,(\mathbb{E}_n^0, \mathbb{E}_{n+i}^i)}(Q)\right)\right), g_{i+1}\left(\mathrm{Im}_{\mathbb{E}_{n+i}^{i}}\left(u^{i,(\mathbb{E}_n^0, \mathbb{E}_{n+i}^i)}(Q)\right)\right)\right) \\
&\qquad + d_{\mathrm{H}}\left(g_{i+1}\left(\mathrm{Im}_{\mathbb{E}_{n+i}^{i}}\left(u^{i,(\mathbb{E}_n^0, \mathbb{E}_{n+i}^i)}(Q)\right)\right), g_{i+1}\left(u^{i+1}\left(\mathrm{Im}_{\mathbb{E}_n^0}(Q)\right)\right)\right) \\
&\quad = \mathbf{O}\left(\mathrm{diam}\left(\mathrm{Im}_{\mathbb{E}_{n+i}^{i}}\left(u^{i,(\mathbb{E}_n^0, \mathbb{E}_{n+i}^i)}(Q)\right)\right)\right) \\
&\qquad + L d_{\mathrm{H}}\left(\mathrm{Im}_{\mathbb{E}_{n+i}^{i}}\left(u^{i,(\mathbb{E}_n^0, \mathbb{E}_{n+i}^i)}(Q)\right), u^i\left(\mathrm{Im}_{\mathbb{E}_n^0}(Q)\right)\right),
\end{aligned}$$

where $L < \infty$ denotes an upper bound on the Lipschitz constant of the function g_{i+1}. Thus, it follows from the induction assumption,

$$d_{\mathrm{H}}\left(\mathrm{Im}_{\mathbb{E}_{n+i}^{i}}\left(u^{i,(\mathbb{E}_n^0, \mathbb{E}_{n+i}^i)}(Q)\right), u^i\left(\mathrm{Im}_{\mathbb{E}_n^0}(Q)\right)\right) = \mathbf{O}\left(\mathrm{diam}\left(\mathrm{Im}_{\mathbb{E}_n^0}(Q)\right)^q\right),$$

that we have

$$\begin{aligned}\operatorname{diam}\left(\operatorname{Im}_{\mathbb{E}^i_{n+i}}\left(u^{i,(\mathbb{E}^0_n,\mathbb{E}^i_{n+i})}(Q)\right)\right)^q &\leq \left[\operatorname{diam}\left(u^i\left(\operatorname{Im}_{\mathbb{E}^0_n}(Q)\right)\right)+\mathbf{O}\left(\left(\operatorname{Im}_{\mathbb{E}^0_n}(Q)\right)^q\right)\right]^q\\ &\leq \left[L_i\operatorname{diam}\left(\operatorname{Im}_{\mathbb{E}^0_n}(Q)\right)+\mathbf{O}\left(\left(\operatorname{Im}_{\mathbb{E}^0_n}(Q)\right)^q\right)\right]^q\\ &= \mathbf{O}\left(\operatorname{diam}\left(\operatorname{Im}_{\mathbb{E}^0_n}(Q)\right)^q\right),\end{aligned}$$

where L_i denotes an upper bound on the Lipschitz constant of the function u_i. Clearly. these inequalities imply

$$d_{\mathrm{H}}\left(\operatorname{Im}_{\mathbb{E}^{i+1}_{n+i+1}}\left(u^{i+1,(\mathbb{E}^0_n,\mathbb{E}^{i+1}_{n+i+1})}(Q)\right),u^{i+1}\left(\operatorname{Im}_{\mathbb{E}^0_n}(Q)\right)\right)=\mathbf{O}\left(\operatorname{diam}\left(\operatorname{Im}_{\mathbb{E}^0_n}(Q)\right)^q\right)$$

which completes our induction proof. □

2.3 Quadratically Convergent Affine-Set Extensions of Vector-Valued Factorable Functions

This section presents a systematic approach for constructing quadratically Hausdorff convergent extensions of twice continuously-differentiable factorable functions when the underlying affine set-parameterization is invariant under affine transformation—see Definition 2.2.

For simplicity of presentation, we only discuss the construction of an extension $\varphi^{\mathbb{E}_\ell}:\mathbb{D}_{n,\ell}\to\mathbb{D}_{m,\ell}$ of a given vector function $\varphi:\mathbb{R}^n\to\mathbb{R}^m$ subsequently, the construction of more general extensions acting on different basis sets being analogous.

The construction starts with a first-order Taylor expansion of the function φ at a point $x^*\in\mathbb{R}^n$,

$$\varphi(x)=\varphi(x^*)+A(x-x^*)+R(\xi_1,\ldots,\xi_m,x-x^*), \tag{2.11}$$

where $R(\xi_1,\ldots,\xi_m,x-x^*):=(r_1(\xi_1,x-x^*),\ldots,r_m(\xi_m,x-x^*))^{\mathsf{T}}$ and with the shorthand notations

$$A:=\frac{\partial\varphi}{\partial x}(x^*),\quad\text{and}\quad\forall a,b\in\mathbb{R}^n,i\in\{1,\ldots,m\},\ r_i(a,b):=\left[\frac{\partial\varphi_i}{\partial x}(a)-A_i\right]b.$$

Points $\xi_i\in\mathbf{conv}(\{x,x^*\})$, $i\in\{1,\ldots,m\}$, such that (2.11) holds, where $\mathbf{conv}(\{x,x^*\})$ denotes the convex hull of the points x and x^* in $\mathbb{R}^n$, are guaranteed to exist by the mean-value

theorem. Moreover, φ being a factorable and continuously-differentiable function, the forward mode of automatic differentiation can be applied and the residual function R is itself a factorable function. At this point, it is worth noting that there is no unique way of choosing a suitable expansion point x^*. In the analysis that follows, we shall assume that x^* is contained in the original set $\mathrm{Im}_{\mathbb{E}_\ell}(Q)$; for instance, the center of that set if $\mathrm{Im}_{\mathbb{E}_\ell}(Q)$ is an ellipsoid. Likewise, there are multiple ways of choosing A besides the Jacobian of φ at x^*, including the Jacobian of φ at other points or the midpoint of an interval enclosure of the Jacobian of φ on $\mathrm{Im}_{\mathbb{E}_\ell}(Q)$.

Provided that an extension $R^{\mathbb{E}_\ell} : \mathbb{D}_{n,\ell} \to \mathbb{D}_{m,\ell}$ of the remainder function R is available such that

$$\left\{R(\xi_1,\ldots,\xi_m,x-x^*) \mid \xi_1,\ldots,\xi_m,x \in \mathrm{Im}_{\mathbb{E}_\ell}(Q)\right\} \subseteq \mathrm{Im}_{\mathbb{E}_\ell}\left(R^{\mathbb{E}_\ell}(Q)\right),$$

an extension of the function φ can be obtained in the form

$$\varphi^{\mathbb{E}_\ell}(Q) := (AQ + (0_{m\times\ell}, \varphi(x^*))) \uplus R^{\mathbb{E}_\ell}(Q). \tag{2.12}$$

The following theorem provides conditions under which such an extension has quadratic Hausdorff convergence.

Theorem 2.1. *Let the affine set-parameterization $(\mathbb{E}_\ell, \mathbb{D}_{n,\ell})$ be invariant under affine transformation and such that the addition extension $\uplus$ is regular. Assume that the extension $R^{\mathbb{E}_\ell}$ of the residual function R in* (2.11) *is locally Lipschitz continuous, so that for all $Q \in \mathbb{D}_{n,\ell}$ with sufficiently small* $\mathrm{diam}\left(\mathrm{Im}_{\mathbb{E}_\ell}(Q)\right)$*, there exists a constant $L < \infty$ such that*

$$\left\|R^{\mathbb{E}_\ell}(Q)\right\|_H \leq L \left\|\left\{R(\xi_1,\ldots,\xi_m,x-x^*) \mid x,\xi_1,\ldots,\xi_m \in \mathrm{Im}_{\mathbb{E}_\ell}(Q)\right\}\right\|_H,$$

with $x^* \in \mathrm{Im}_{\mathbb{E}_\ell}(Q)$. Then, the extension $\varphi^{\mathbb{E}_\ell}$ in (2.12) *has Hausdorff convergence order* 2.

Proof. Let $x^* \in \mathrm{Im}_{\mathbb{E}_\ell}(Q)$. By invariance of $(\mathbb{E}_\ell, \mathbb{D}_{n,\ell})$ under affine transformation, the set $AQ + (0_{m\times\ell}, \varphi(x^*))$ corresponds to the exact image of the affine approximation $\varphi(x^*) + A(x-x^*)$ of φ on $\mathrm{Im}_{\mathbb{E}_\ell}(Q)$. Moreover, φ being twice continuously differentiable, we have

$$\max_{x,\xi_1,\ldots,\xi_m \in \mathrm{Im}_{\mathbb{E}_\ell}(Q)} \|R(\xi_1,\ldots,\xi_m,x-x^*)\| = \mathbf{O}\left(\left\|\mathrm{Im}_{\mathbb{E}_\ell}(Q) - x^*\right\|_H^2\right),$$

and it follows from local Lipschitz-continuity of $R^{\mathbb{E}_\ell}$ that

$$\left\| R^{\mathbb{E}_\ell}(Q) \right\|_H = \mathbf{O}\left(\left\| \mathrm{Im}_{\mathbb{E}_\ell}(Q) - x^* \right\|_H^2 \right) ,$$

for any $Q \in \mathbb{D}_{n,\ell}$ with sufficiently small diam$\left(\mathrm{Im}_{\mathbb{E}_\ell}(Q)\right)$. The result follows by noting that the addition extension $\uplus$ is regular. □

Although the proof of Theorem 2.1 is quite straightforward from a mathematical standpoint, the construct (2.12) proves especially useful to compute practical extensions of factorable functions with quadratic Hausdorff convergence. The key step involves constructing a Lipschitz-continuous range bounder of the residual function R, yet this poses no particular problem as simple interval arithmetics [96] can be used for this purpose.

A quadratically convergent arithmetic can be constructed using a parameterization of the form $(\mathbb{E}_\ell^{\mathrm{pol}(q)} \times \mathbb{E}_n^{\mathrm{ball}}, \mathbb{R}^{n\times(\alpha_\ell^{(q)}+n+1)})$, namely polynomial models with ellipsoidal remainders. The arithmetic requires the availability of a regular extension of the bivariate sum, which can be defined in a straight forward manner using the regular addition extension for ellipsoids given by Equation (2.5) and noticing that the sum of two qth-order polynomials is itself a qth-order polynomial. The quadratic convergence of such arithmetic thus, follows from Theorem 2.1.

2.4 Ellipsoidal Arithmetic for Factorable Functions

In this section the method presented in Section 2.2.1 is used for the construction of an ellipsoidal arithmetic for factorable functions. In particular, the arithmetic is based on the propagation of ellipsoids using Equations (2.9) and (2.10) with parameterizations of the form $(\mathbb{E}_n^{\mathrm{ball}}, \mathbb{R}^{n\times(n+1)})$.

In order to present the method in a more concrete manner, the notation for affine set arithmetics will not be used. Instead, an n-dimensional ellipsoid will be denoted by $\mathcal{E}(q,Q) = \{q + Q^{\frac{1}{2}}v \,|\, \|v\| \leq 1\}$, where $q \in \mathbb{R}^n$ and $Q \in \mathbb{S}_+^n$ denote the center and shape matrix of the ellipsoid. In this section we will also denote the set of n-dimensional ellipsoids in $\mathbb{R}^n$ by $\mathbb{E}^n$. Notice that this definition is not in conflict with the definition given in Example 2.1, since in the affine-set arithmetic formalism, the same ellipsoid is given by taking the base set $\mathbb{E}_n^{\mathrm{ball}}$ and choosing the parameter $(Q^{\frac{1}{2}}, q) \in \mathbb{R}^{n\times(n+1)}$.

Consider a factorable function $f : \mathbb{R}^{n_x} \to \mathbb{R}^{n_f}$ and its decomposition $f = P[g_N \circ g_{N-1} \circ \cdots \circ g_1]$ into factors g_i with a projection matrix $P \in \mathbb{R}^{n_f \times (n_x+N)}$. Given an ellipsoidal domain

$\mathcal{E}(q_x, Q_x) \in \mathbb{E}^{n_x}$ the exact image of $\mathcal{E}(q_x, Q_x)$ under f is denoted by

$$f(\mathcal{E}(q_x, Q_x)) := \{f(x) \,|\, x \in \mathcal{E}(q_x, Q_x)\}. \tag{2.13}$$

The main focus of this section is on computing an enclosure of the image set of f in the form of an ellipsoid $\mathcal{E}(q_f, Q_f)$; that is, $\mathcal{E}(q_f, Q_f) \supseteq f(\mathcal{E}(q_x, Q_x))$ with $(q_f, Q_f) \in \mathbb{R}^{n_f} \times \mathbb{S}_+^{n_f}$.

The method starts from an ellipsoid $\mathcal{E}(q_{i-1}, Q_{i-1}) \in \mathbb{E}^{i-1}$ for some $i \in n_x + 1, \ldots, n_f$, an ellipsoidal extension of the i-th factor $g_i^{\mathbb{E}}(\mathcal{E}(q_{i-1}, Q_{i-1}))$ is obtained as:

$$q_i := \begin{pmatrix} q_{i-1} \\ l^i \end{pmatrix} \quad \text{and} \quad Q_i := \begin{pmatrix} I^{i-1} \\ L^i \end{pmatrix} Q_{i-1} \operatorname{tr} \begin{pmatrix} I^{i-1} \\ L^i \end{pmatrix} \tag{2.14}$$

where I^{i-1} is an $(i-1)$-dimensional identity matrix, $l \in \mathbb{R}$ and $L \in \mathbb{R}^{1 \times i}$ are the lifting parameters. These lifting parameters depend on the functional form of the atom operations a_i.

The desired enclosure $\mathcal{E}(q_f, Q_f)$ (and thus the ellipsoidal extension of f) can then be recovered by projecting the lifted ellipsoid into $\mathbb{R}^m$, i.e. $f^{\mathbb{E}}(q_x, Q_x) = \mathcal{E}(q_f, Q_f) := P\mathcal{E}(q_N, Q_N)$.

In the following we provide ways to construct the lifting parameters for different atom operations.

2.4.1 Affine atom operations

Here we are concerned with operations $u_i(x) := g_i(u_{i-1}(x))$ where the atom operation a_i is an affine function of at most two preceding factors $u_k(x), u_m(x)$ with $k, m \leq i-1$. This corresponds to the case of addition $(u_k(x) + u_m(x))$, subtraction $(u_k(x) - u_m(x))$, scaling $(bu_k(x))$ and shifting $(u_k(x) + b)$.

Addition and subtraction $(u_k(x) \pm u_m(x))$ The lifting parameters are given by:

$$l := q_k \pm q_m \quad \text{and} \quad L_j := \begin{cases} 1 & \text{if } j = k \\ \pm 1 & \text{if } j = m \\ 0 & \text{otherwise} \end{cases}. \tag{2.15}$$

Scaling ($bu_k(x)$) Geometrically speaking, multiplying an ellipsoid by a scalar corresponds to scaling such ellipsoid. In this case, the lifting parameters are given by:

$$l := q_k \quad \text{and} \quad L_j := \begin{cases} b & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}. \tag{2.16}$$

Shifting ($u_k(x) + b$) Shifting an ellipsoid is an operation that affects only the center of the ellipsoid. In this case, the lifting parameters are given by:

$$l := q_k + b \quad \text{and} \quad L_j := \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}. \tag{2.17}$$

2.4.2 Nonlinear univariate atom operations

In this section we are concerned with operations $u_i(x) = g_i(u_{i-1}(x))$ where the atom operation a_i is a possible nonlinear function of at most one factor $u_k(x)$ with $k \leq i-1$. Moreover we assume that a_i is at least twice continuously differentiable. The domain $D_i \subseteq \mathbb{R}$ of a_i is given by the projection of the ellipsoid $\mathcal{E}(q_{i-1}, Q_{i-1})$ onto the k-th component of the $(i-1)$-th dimensional space $D_i := (q_{i-1})_k + \sqrt{(Q_{i-1})_{k,k}}\,[-1,1]$.

The approach to handle nonlinear functions g_i is to expand it at a point $v \in \mathcal{E}(q_{i-1}, Q_{i-1})$:

$$g_i(u_{i-1}(x)) = g(v) + A \begin{pmatrix} u_{i-1}(x) - v \\ u_k(x) - v_k \end{pmatrix} + R\left(\begin{pmatrix} \xi \\ \xi_k \end{pmatrix}, \begin{pmatrix} u_{i-1}(x) - v \\ u_k(x) - v_k \end{pmatrix} \right) \tag{2.18}$$

where A, is the Jacobian matrix of g_i evaluated at v, R is the residual function and $\xi \in \mathcal{E}(q_{i-1}, Q_{i-1})$. The construction of the ellipsoidal extension requires the availability of validated linear approximations of the atom operations a_i on D_i, i.e. constants c_0, c_1 and an interval $[-\eta, \eta]$ such that $\mathrm{Im}\,(a(D_i)) \subseteq \{c_1 y + c_0 \mid y \in D_i\} \oplus [-\eta, \eta]$. These approximations can be computed in several ways, for example using Taylor or Chebyshev model arithmetic.

The lifting operation is a two-step procedure, in the first one, the linear approximation is absorbed into the a lifted ellipsoid $\mathcal{E}(q_i, \tilde{Q}_i)$ in a procedure that is similar to the one used for shifting and scaling i.e. apply Equation (2.14) with the parameters

$$l := c_0 \quad \text{and} \quad L_j := \begin{cases} \frac{c_1}{\mathrm{rad}(D_i)} & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}. \tag{2.19}$$

In the second step the interval $[-\eta,\eta]$ which encloses the nonlinearity is added (In the Minkowski sense) to the i-th component of the lifted ellipsoid. Since the set of n-dimensional ellipsoids is not closed under Minkowski addition, we require an ellipsoidal extension of the Minkowski sum $\mathcal{E}(q_i,Q_i) := \mathcal{E}(q_i,\tilde{Q}_i) \oplus^{\mathbb{E}} \mathcal{E}(Q_\eta)$, where $Q_\eta := \eta^2(e_i e_i^\top)$ and e_i is the euclidean i-th basis vector. Notice that in the affine-set formalism, an extension of the Minkowski sum is equivalent to an extension of the addition operation between parameterization. With this observation, we can use the regular addition operation given in Equation (2.5). The shape matrix Q_i is then computed as:

$$Q_i := \frac{1}{\lambda_0}\tilde{Q}_i + \frac{1}{\lambda_1}Q_\eta \quad \text{with} \quad \lambda_0 = 1-\lambda_1 = \frac{\sqrt{\operatorname{tr}\tilde{Q}_{i-1}+\varepsilon}}{\sqrt{\operatorname{tr}\tilde{Q}_{i-1}+\varepsilon}+\sqrt{\operatorname{tr}Q_\eta+\varepsilon}}\,. \tag{2.20}$$

2.4.3 Bivariate product

There are many different ways for enclosing the product of two factors $u_k(x)$, $u_m(x)$ with $k,m \leq i-1$. In here we use a DC-decomposition approach

$$u_k(x)u_m(x) = \frac{1}{4}\left((u_k(x)+u_m(x))^2 - (u_k(x)+u_m(x))^2\right) \tag{2.21}$$

and then apply recursively the rules for bivariate operations and univariate compositions given in the previous sections.

2.4.4 Convergence Analysis

After explaining the construction of the ellipsoidal arithmetic, we proceed to show its convergence properties. Recall that an ellipsoidal extension $f^{\mathbb{E}}$ of f on a domain $\mathcal{E}(q_x,Q_x) \in \mathbb{E}^{n_x}$ has Hausdorff convergence order $q \geq 1$ if

$$d_{\mathrm{H}}\left(f^{\mathbb{E}}(\mathcal{E}(q_x,Q_x)), \mathrm{Im}(f(\mathcal{E}(q_x,Q_x)))\right) = \mathbf{O}(\operatorname{diam}(\mathcal{E}(q_x,Q_x))^q)\,. \tag{2.22}$$

We now present the main result of this section.

Theorem 2.2. *The construction of an ellipsoidal extension $f^{\mathbb{E}}$ of a function f, as given in Section 2.4 has Hausdorff Image Convergence order 2.*

Proof. First, we will show that the ellipsoidal extension of each atom operation converges at least quadratically. The class of ellipsoids is closed under affine transformations, as a consequence the procedure described in Section 2.4.1 does not introduce any overestimation

or wrapping and hence. For bivariate products, the procedure is reduced to the application of affine functions and univariate nonlinear compositions. Consider now a factorable update $g_i(u_{i-1}(x))$ and assume without loss of generality that all the ellipsoids are centered at zero.By construction of $g_i^{\mathbb{E}}$ and invariance of ellipsoids under affine transformations the ellipsoidal approximation $\mathcal{E}(\tilde{Q}_i)$ matches the exact image of the linear approximation of g:

$$\mathcal{E}(\tilde{Q}_i) = \left\{ A \begin{pmatrix} u_{i-1}(x) \\ u_k(x) \end{pmatrix} \middle| v, u_{i-1}(x) \in \mathcal{E}(Q_{i-1}) \right\}.$$

Let $\mathcal{R}$ be the exact image of the residual function R on $\mathcal{E}(Q_{i-1})$, applying the triangular inequality

$$\begin{aligned} d_{\mathrm{H}}\left(\mathcal{E}(Q_i), g_i(\mathcal{E}(Q_{i-1}))\right) &= d_{\mathrm{H}}\left(\mathcal{E}(\tilde{Q}_i) \oplus^{\mathbb{E}} \mathcal{E}(Q_\eta), \mathcal{E}(\tilde{Q}_i) \oplus \mathcal{R}\right) \\ &\leq d_{\mathrm{H}}\left(\mathcal{E}(\tilde{Q}_i) \oplus^{\mathbb{E}} \mathcal{E}(Q_\eta), \mathcal{E}(\tilde{Q}_i)\right) + d_{\mathrm{H}}\left(\mathcal{E}(\tilde{Q}_i), \mathcal{E}(\tilde{Q}_i) \oplus \mathcal{R})\right) \\ &\leq d_{\mathrm{H}}\left(\mathcal{E}(\tilde{Q}_i) \oplus^{\mathbb{E}} \mathcal{E}(Q_\eta), \mathcal{E}(\tilde{Q}_i)\right) + \mathbf{O}\left(\operatorname{diam}(\mathcal{R})\right). \end{aligned} \tag{2.23}$$

It can be shown from Equation (2.20) that the ellipsoidal extension of the Minkowski sum satisfies

$$d_{\mathrm{H}}\left(\mathcal{E}(\tilde{Q}_i) \oplus^{\mathbb{E}} \mathcal{E}(Q_\eta)\right) \leq \mathbf{O}\left(\operatorname{diam}(\mathcal{E}(Q_{i-1}))\right) \tag{2.24}$$

Since g_i is twice continuously differentiable and by the local Lipschitz-continuity of interval enclosures we respectively get:

$$\operatorname{diam}(\mathcal{R}) \leq \mathbf{O}\left(\operatorname{diam}(\mathcal{E}(Q_{i-1}))^2\right) \quad \text{and} \quad \operatorname{diam}(\mathcal{E}(Q_\eta)) \leq \mathbf{O}\left(\operatorname{diam}(\mathcal{E}(Q_{i-1}))^2\right). \tag{2.25}$$

Combining Inequalities (2.23), (2.24) and (2.25) proves the quadratic convergence of the ellipsoidal update for nonlinear functions. Application of Lemma 2.1 with $q = 2$ gives the result. □

2.4.5 Numerical Implementation and Case Studies

Although illustrative, the matrix multiplication approach given in Section 2.4 is not the most effective way to implement the arithmetic. Firstly, the number of atom operations N participating in a factorable function is known a priori, which allows the preallocation of the center and shape matrix. The second observation is that the lifting operation defined in Equation (2.14) corresponds to an update in the i-th row of the shape matrix, for example

for the addition and substraction $(u_k(x) \pm u_m(x))$ for each $j \in \{0,\ldots,i\}$:

$$(Q_N)_{i,j} = \begin{cases} (Q_N)_{k,k} \pm 2(Q_N)_{k,m} + (Q_N)_{m,m} & \text{if} \quad j = 1 \\ (Q_N)_{k,j} \pm (Q_N)_{m,j} & \text{otherwise} \end{cases}$$

and for scaling $(bu_k(x))$:

$$(Q_N)_{i,j} = \begin{cases} b^2(Q_N)_{k,k} & \text{if} \quad j = 1 \\ b(Q_N)_{k,j} & \text{otherwise} \end{cases}.$$

An implementation of the ellipsoidal arithmetic is available as part of the software package MC++ (https://projects.coin-or.org/MCpp). For efficiency, the (symmetric) shape matrices of the lifted ellipsoid is stored in sparse format and every atom operation corresponds to a sparse update of the shape matrix.

We illustrate the benefits of ellipsoidal arithmetic compared with traditional interval arithmetic for two problems. The left plot in Fig. 2.1 considers the following factorable function with $n_x = n_f = 2$,

$$f_1(x_1,x_2) = \sqrt{x_1 + x_2} + x_1 x_2 \qquad f_2(x_1,x_2) = (x_1 - x_2)^2 + 3x_2\,.$$

The comparison between the actual image set and the corresponding interval and ellipsoidal extensions, here for a variable host set $\mathcal{E}(q_x, Q_x)$ with $q_x = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and $Q_x = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, illustrates the potential of ellipsoidal arithmetic to mitigate the dependency problem. The right plot of Fig. 2.1 shows the solution set of the parametric linear ODE (harmonic oscillator)

$$\dot{x}_1(t) = x_2(t) + p\,, \qquad \dot{x}_2(t) = -x_1(t) + p\,,$$

with joint ellipsoidal host set $\mathcal{E}(q_{x,p}, Q_{x,p})$ for the initial states $x_1(0), x_2(0)$ and the parameter p given by $q_{x,p} = (1 \;\; 0 \;\; 0)^\mathsf{T}$ and $Q_{x,p} = 0.01\, I_3$. The unconditional stability of the reach tube illustrates the ability of ellipsoidal arithmetic to mitigate the wrapping effect, which is due to the property of invariance under affine transformations of ellipsoids.

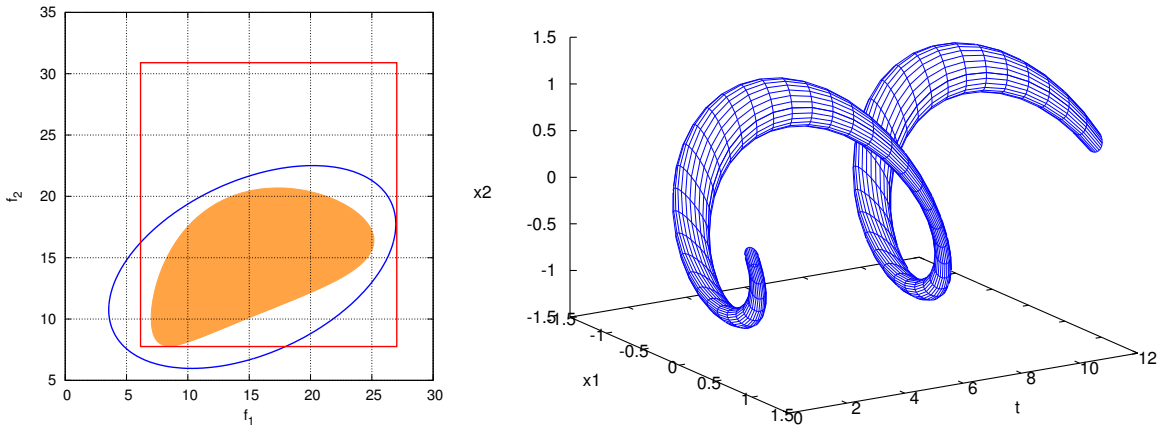

Fig. 2.1 Left plot: Mitigation of the dependency problem using ellipsoidal arithmetic compared with interval arithmetic for a factorable function; Right plot: Mitigation of the wrapping effect by the property of invariance under affine transformation of ellipsoids for a parametric ODE.

2.5 Conclusions

In this chapter, a framework based on the concept of affine-set arithmetics for bounding the image-set of factorable, vector-valued functions was presented. Similarly to natural interval extensions, the concept of affine-set valued extensions and methods to construct those extensions were discussed. The Hausdorff convergence of such extensions was considered, in particular a set of sufficient conditions for the construction of quadratically convergent affine-set extensions was presented.

The general framework of affine-set arithmetics was also used for the construction of an ellipsoidal arithmetic. This arithmetic relies on a lifting operation for every atom operation, which is natural to vector-valued factorable functions. This arithmetic was shown to yield quadratically convergent ellipsoidal extensions under mild assumptions.

In practice, quadratic Hausdorff convergence is a very desirable property in branch-and-bound search for global optimization in order to reduce the cluster effect, as well as in the design of validated integrators for asymptotically stable systems. Further research is needed in order to extend existing interval-based algorithms to ellipsoidal techniques nonetheless.

Chapter 3

Discrete-Time Set Propagation

This Chapter is concerned with set propagation through dynamic systems. In particular, the focus is on discrete-time set propagation methods. In here we propose a novel set-valued integrator using the affine-set arithmetics formalism. One of the main problems of set propagation through dynamic systems is that the diameter of the computed enclosures typically tends to diverge to infinity on finite integration horizons, even thought the ODE trajectories themselves may be asymptotically stable. In this chapter we show that if the chosen affine-set parameterization has quadratic Hausdorff convergence, the enclosures computed with the proposed set-valued integrator remain stable on infinite time horizons, when applied to a dynamic system in the neighborhood of a locally asymptotically stable periodic orbit (or equilibrium point).

The rest of the Chapter is organized as follows. Section 3.1 introduces the problem definition and the preliminary material used throughout the chapter. Section 3.2 presents a brief review of existing two-phase discrete-time propagation approaches in the context of affine-set arithmetics. Section 3.3 introduces a novel two-phase discrete-time set-valued integrator. Due to the reversal of the phases (with respect to existing approaches) this algorithm eliminates the need for an a priori enclosure and also leads to a natural step-size selection strategy, which is also introduced in this section. Section 3.4 presents the stability analysis of the reversed two-phase algorithm. In particular a set of sufficient conditions are provided for the enclosures of asymptotically stable dynamic systems to remain stable on infinite time horizons. Section 3.5 presents a strategy for the use of ODE invariants in order to tighten the enclosures computed using the set-valued integrator. Section 3.6 illustrates the stability of the bounds and the use of invariants in three case studies: a simple cubic oscillator, a two-stage anaerobic digestion model and a model for reversible chemical

reactions. Finally, Section 3.7 concludes the chapter. A computational application of the concepts presented in this chapter can be found in Appendix D.

3.1 Problem Statement

We consider parametric dynamic systems in the form of nonlinear ODEs

$$\forall t \in [0,T], \quad \dot{x}(t,p) = f(t,x(t,p),p) \quad \text{with} \quad x(0,p) = x_0(p)\,. \tag{3.1}$$

In connection to this problem we make the following blanket assumptions:

Assumption 3.1. *The right-hand side function $f : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$ is jointly smooth in (t,x,p) and factorable.*

Assumption 3.2. *The initial-value function $x_0 : \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$ is smooth in p and factorable.*

Assumption 3.3. *The parameter set P is compact and a parameterization $Q_p \in \mathbb{D}_{n_p,\ell}$ is given such that $\mathrm{Im}_{\mathbb{E}_\ell}(Q_p) \supseteq P$, with $\ell \geq 1$ the actual number of degrees of freedom in the parameterization.*

We note that Assumption 3.1 is mainly introduced to keep notation in the chapter as simple as possible, and the algorithm presented later can be readily applied to more general classes of parametric ODEs, whose right-hand side functions are 'sufficiently often continuously-differentiable'. Likewise, Assumption 3.2 makes it possible to compute an enclosure of a (possibly nonconvex) initial value set of the form $X_0 := \{x_0(p) \mid p \in P\}$. In presenting the set-propagation algorithm, we keep our considerations general and do not specialize to a particular affine set-parameterization. Regarding existence and uniqueness of the ODE solutions, we note that a solution $x(\cdot,p) : [0,T] \to \mathbb{R}^{n_x}$ of the initial-value problem (3.1) may not exist for all $p \in P$, but it is guaranteed to be unique on its maximum interval of existence by Assumption 3.1

The state $x : [0,T] \times P \to \mathbb{R}^{n_x}$ is regarded as a function of the uncertain parameter vector $p \in P \subseteq \mathbb{R}^{n_p}$ along the time horizon $[0,T]$. The reachable set of the initial value problem (3.1) is denoted by

$$X(t,P) := \{x(t,p) \mid p \in P\}, \tag{3.2}$$

or simply $X(t)$ when it is clear from the context what the corresponding parameter host set P is.

The focus here is on algorithms that compute a time-varying enclosure $Y(t,P) \supseteq X(t,P)$ for all $t \in [0,T]$ using discretized set-valued integration techniques. Particular emphasis is on analyzing whether, as well as determining under which conditions, one can obtain stable enclosures $Y(t,P)$ on infinite time horizons for sufficiently small (yet finite) parameter host sets P.

A major complication with the solutions $x(t,\cdot)$ of the parametric ODEs (3.1) is that these functions do not have a factorable representation in general, and therefore classical bounding techniques based on interval analysis or Taylor model arithmetic cannot be applied directly. Nonetheless, algorithms for bounding the solution set of parametric ODEs take advantage of the fact that the right-hand side function f and the initial value function x_0 are typically factorable.

3.2 Review of Existing Discretized Set-Valued Algorithms

The focus of discrete-time set propagation is to develop numerical algorithms that constructs a matrix valued function $Q_x : [0,T] \to \mathbb{D}_{n_x,\ell}$ such that

$$\forall t \in [0,T], \quad \mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t)) \supseteq X(t)\,.$$

Existing validated integrators for nonlinear ODEs [100, 101, 127] consider a Taylor series expansion in time of the ODE solutions. Assuming that $x(\cdot,p)$ is the solution of (3.1) up to time $t \in [0,T)$ for a given parameter p, and provided that this solution can be extended until $t+h$ with $h \in (0, T-t]$, the application of Taylor's theorem for an s-th order expansion gives

$$x(t+h,p) = \sum_{i=0}^{s} h^i \phi_i(t,x(t,p),p) + h^{s+1}\phi_{s+1}(\tau,x(\tau,p),p) \tag{3.3}$$

for some $\tau \in [t,t+h]$. Here, $\phi_0, \phi_1, \ldots, \phi_{s+1} : [0,T] \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$ denote the Taylor coefficient functions of the solution, defined recursively as

$$\phi_0(t,x,p) := x \quad \text{and} \quad \phi_i(t,x,p) := \frac{1}{i}\left(\frac{\partial \phi_{i-1}}{\partial x}(t,x,p)\, f(t,x,p) + \frac{\partial \phi_{i-1}}{\partial t}(t,x,p)\right),$$

for $i = 1,\ldots,s+1$ and for all $(t,x,p) \in [0,T] \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_p}$.

State-of-the-art validated integrators [100, 101, 127] proceed in two phases analagous to a predictor-corrector method for integration of ODEs. In the first phase a step size and an a

priori enclosure of the solution trajectory are determined followed by a tightening of this a priori enclosure on the second phase.

Phase I. Given a parameterization $Q_x(t)$ at some $t \in [0,T)$ such that $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t)) \supseteq X(t)$, a step size h and an a priori enclosure $\tilde{Q}_x(t+h)$ of the solution are computed. That is, determine $\tilde{Q}_x(t+h)$ and h such that for some $\tilde{Q}_x^0(t)$ satisfying $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t)) \subset \mathrm{Im}_{\mathbb{E}_\ell}\left(\tilde{Q}_x^0(t)\right)$ we have

$$\mathrm{Im}_{\mathbb{E}_\ell}\left(\tilde{Q}_x(t+h)\right) \subseteq \mathrm{Im}_{\mathbb{E}_\ell}\left(\tilde{Q}_x^0(t)\right) \tag{3.4}$$

$$\tilde{Q}_x(t+h) := \biguplus_{i=0}^{s} [0,h]^i \phi_i^{\mathbb{E}_\ell}([t,t+h], Q_x(t), Q_p) \uplus [0,h]^s \phi_s^{\mathbb{E}_\ell}([t,t+h], \tilde{Q}_x^0(t), Q_p) .$$

Here, $\phi_i^{\mathbb{E}_\ell} : [0,T] \times \mathbb{D}_{n_x,\ell} \times \mathbb{D}_{n_p,\ell} \to \mathbb{D}_{n_x,\ell}$ are extensions of the Taylor coefficient functions ϕ_i for each $i = 0,\ldots,s$ and $\uplus$ stands for the addition extension according to Definition 2.5. Notice that the parameter $\tilde{Q}_x$ can be obtained by recursively increasing $Q^0(t)$ and/or decreasing h until Inclusion (3.4) is satisfied. Typically, heuristics are used in order to determine the step size.

Phase II. Given the parameterization $\tilde{Q}_x(t+h)$ and the step size h computed in Phase I, an enclosure $Q_x(t+h)$ is computed, satisfying $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t+h)) \supseteq X(t+h)$. Notice that the inclusion of the reachable set in the tightened enclosure is only valid at the end of the step, while the a priori enclosure is valid for the whole interval $[t,t+h]$. One approach to compute the tightened enclosure is by evaluating an affine set extension of Equation (3.3) with the remainder term $h^{s+1}\phi_{s+1}(\tau, x(\tau,p), p)$ evaluated at $\tilde{Q}_x(t+h)$. Since validated methods were developed for parameterizations which are not closed under affine operations, intervals and Taylor models with interval remainders, the naive evaluation method tends to induce overestimation due to wrapping and ultimately cause the bounds to escape in finite time (bound explosion). One way of mitigating the wrapping effect is to use the mean-value theorem applied to Equation (3.3) and apply a reformulation in order to propagate rotated enclosures [100].

In the next section we present a novel algorithm for discrete-time propagation of affine-set parameterizations. The algorithm is designed to mitigate (and in some cases eliminate) the wrapping effect, provide continuous output and have –under certain conditions– better stability properties than the traditional two-phase method.

3.3 Discretized Set-Valued Integration Algorithm

In contrast to existing validated methods, the algorithm presented below reverses the order of these two phases, thereby removing the need for an a priori enclosure of the solution and providing a natural mechanism for step-size selection. This procedure is described next for the propagation of a generic affine set-parameterization, as introduced in Chapter 2. It starts with a parameterization $Q_x(0) := x_0^{\mathbb{E}_\ell}(Q_p)$, with $x_0^{\mathbb{E}_\ell} : \mathbb{D}_{n_p,\ell} \to \mathbb{D}_{n_x,\ell}$ an extension of the initial-value function x_0, so that $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(0)) \supseteq X(0)$. Then, the following two steps are applied repeatedly:

Step 1. Given a parameterization $Q_x(t)$ at some $t \in [0,T)$ such that $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t)) \supseteq X(t)$, a predictor $Q_x(t+h)$ of the solution for all $h \in (0, T-t]$ is given by:

$$Q_x(t+h) \;:=\; \biguplus_{i=0}^{s} h^i \phi_i^{\mathbb{E}_\ell}(t, Q_x(t), Q_p) \;\uplus\; h\,\mathrm{TOL}\,Q_{\mathrm{unit}} \tag{3.5}$$

for a pre-specified tolerance $\mathrm{TOL} > 0$ and a given $Q_{\mathrm{unit}} \in \mathbb{D}_{n_x,\ell}$ to be defined below, and where $\phi_i^{\mathbb{E}_\ell} : [0,T] \times \mathbb{D}_{n_x,\ell} \times \mathbb{D}_{n_p,\ell} \to \mathbb{D}_{n_x,\ell}$ are extensions of the Taylor coefficient functions ϕ_i for each $i = 0,\ldots,s$ and $\uplus$ stands for the addition extension according to Definition 2.5. In order to avoid confusion at this point, we note that the predictor $Q_x(t+h)$ does not need to be evaluated for a particular step-size during the propagation. In a practical implementation Q_x can be stored in the form a computational graph, with one of the graph leaves corresponding to the (symbolic) variable h. This way, an enclosure $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t+h)) \supseteq X(t+h)$ can always be obtained *after* the propagation, by evaluating (3.5) for *any* step-size h satisfying the validation condition given next.

Step 2. A step-size $\bar{h}$ is determined, such that the predictor $Q_x(t+h)$ in (3.5) is guaranteed to yield a valid enclosure of the reachable set, $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t+h)) \supseteq X(t+h)$, for all $h \in [0,\bar{h}]$. In particular, any feasible point $\bar{h} > 0$ of the following optimization problem is suitable:

$$\sup_{h>0} \; h \quad \text{s.t.} \quad \begin{cases} \forall \tau \in [t, t+h], \\ \mathrm{Im}_{\mathbb{E}_\ell}\Big((\tau - t)^s \, \phi_{s+1}^{\mathbb{E}_\ell}(t, Q_x(\tau), Q_p) \Big) \;\subseteq\; \mathrm{TOL}\,\mathrm{Im}_{\mathbb{E}_\ell}(Q_{\mathrm{unit}}) \end{cases} \tag{3.6}$$

with $\phi_{s+1}^{\mathbb{E}_\ell} : [0,T] \times \mathbb{D}_{n_x,\ell} \times \mathbb{D}_{n_p,\ell} \to \mathbb{D}_{n_x,\ell}$ an extension of the Taylor coefficient function ϕ_{s+1}.

The validity of this bounding procedure is established in the following theorem.

Theorem 3.1. *Let the function f satisfy the blanket assumption efass::blanket1, and let $Q_x(t) \in \mathbb{D}_{n_x,\ell}$ be such that $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t)) \supseteq X(t)$. If $\bar{h}$ is a feasible point of the step-size selection problem* (3.6)*, then $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t+h)) \supseteq X(t+h)$ for all $h \in [0,\bar{h}]$, with $Q_x(t+h)$ given by* (3.5)*.*

Proof. It follows from the application of the affine set arithmetic formalism to the Taylor series expansion (3.3) that $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t+h)) \supseteq X(t+h)$ whenever the remainder term $(\tau - t)^{s+1}\phi_{s+1}(x(\tau,p))$ is contained in the set $\mathrm{Im}_{\mathbb{E}_\ell}(h\,\mathrm{TOL}\,Q_{\mathrm{unit}})$ for all $\tau \in [t,t+h]$. The semi-infinite constraint in (3.6) ensures that the remainder term satisfies this condition on the interval $[0,\bar{h}]$ by construction. □

For any practical purposes, it is convenient (and sufficient) to solve the step-size selection problem (3.6) approximately. Consider the remainder function $r_t : [0, T-t] \to \mathbb{D}_{n_x}^{\mathrm{interval}}$ given by:

$$r_t(h) := \mathcal{B}^{\mathbb{E}_\ell,\mathbb{E}_{n_x}^{\mathrm{box}}}\left(\phi_{s+1}^{\mathbb{E}_\ell}(t, Q_x(t+h), Q_p)\right), \tag{3.7}$$

where $\mathcal{B}^{\mathbb{E}_\ell,\mathbb{E}_{n_x}^{\mathrm{box}}} : \mathbb{D}_{n_x,\ell} \to \mathbb{D}_{n_x}^{\mathrm{interval}}$ is a range bounder for the chosen affine set-parameterization,

$$\forall Q \in \mathbb{D}_{n_x,\ell}, \quad \mathrm{Im}_{\mathbb{E}^\ell}(Q) \subseteq \mathrm{Im}_{\mathbb{E}_{n_x}^{\mathrm{box}}}(\mathcal{B}_{n_x}(Q)),$$

and let $r_t^{\mathbb{E}_1^{\mathrm{box}},\mathbb{E}_{n_x}^{\mathrm{box}}} : \mathbb{D}_1^{\mathrm{interval}} \to \mathbb{D}_{n_x}^{\mathrm{interval}}$ be an interval extension of r_t. The following corollary of Theorem 3.1 is immediate upon noting that

$$\mathrm{Im}_{\mathbb{E}_\ell}\left(h^s\,\phi_{s+1}^{\mathbb{E}_\ell}(t, Q_x(t+h), Q_p)\right) \subseteq \mathrm{Im}_{\mathbb{E}_{n_x}^{\mathrm{box}}}\left(r_t^{\mathbb{E}_1^{\mathrm{box}},\mathbb{E}_{n_x}^{\mathrm{box}}}(\tfrac{1}{2}h, \tfrac{1}{2}h)\right),$$

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

Corollary 3.1. *Let the function f satisfy the blanket assumption efass::blanket1, let $Q_x(t) \in \mathbb{D}_{n_x,\ell}$ be such that $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t)) \supseteq X(t)$, and let $\sigma \in \mathbb{R}_+^{n_x}$, $\sigma > 0$, be such that $[-\sigma,\sigma] \subseteq \mathrm{Im}_{\mathbb{E}_\ell}(Q_{\mathrm{unit}})$. Suppose that $\bar{h} \in (0, T-t]$ satisfies*

$$\bar{h}^s \left\| \mathrm{diag}(\sigma)^{-1}\,\mathrm{abs}\left(\mathrm{Im}_{\mathbb{E}_{n_x}^{\mathrm{box}}}\left(r_t^{\mathbb{E}_1^{\mathrm{box}},\mathbb{E}_{n_x}^{\mathrm{box}}}(\tfrac{1}{2}h, \tfrac{1}{2}h)\right)\right)\right\|_\infty \leq \mathrm{TOL}\,. \tag{3.8}$$

Then, $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t+h)) \supseteq X(t+h)$ for all $h \in [0,\bar{h}]$, with $Q_x(t+h)$ given by (3.5)*.*

In particular, (3.8) provides a practical condition for step-size validation. The following simple iterative procedure determines a feasible step-size:

1. Consider the following initial guess for the step-size:

$$\bar{h} = \rho \left(\frac{\mathrm{TOL}}{\left\| \mathrm{diag}(\sigma)^{-1} \, \mathrm{abs} \left(\mathrm{Im}_{\mathbb{E}_{n_x}^{\mathrm{box}}} \left(r_t^{\mathbb{E}_1^{\mathrm{box}}, \mathbb{E}_{n_x}^{\mathrm{box}}}(0) \right) \right) \right\|_\infty} \right)^{\frac{1}{s}} , \tag{3.9}$$

with $0 < \rho < 1$ a tuning parameter, e.g., $\rho = 0.8$.

2. While condition (3.8) is not satisfied with $\bar{h}$, reduce the step-size as $\bar{h} \leftarrow \rho \bar{h}$.

Note that this procedure is guaranteed to identify a feasible step-size $\bar{h} > 0$ after finitely many iterations, since the left hand-side expression in (3.8) shrinks with order $O(h^s)$, whereas the right-hand term is constant and has a non-empty interior.

As far as the selection of Q_{unit} and σ is concerned, a practical procedure involves setting σ first—e.g., by accounting for the relative magnitude of the state variables—and determining Q_{unit} accordingly, in such a way that $\mathrm{Im}_{\mathbb{E}_\ell}(Q_{\mathrm{unit}}) \supseteq [-\sigma, \sigma]$. In analogy to standard scaling heuristics used in state-of-the-art (non-validated) ODE solvers [49], σ can be seen as a scaling vector and adjusted dynamically during the integration as follows:

$$\sigma_i := \frac{1}{2} \mathcal{B}_{n_x}(Q_x(t))_{i,i} + \frac{\mathrm{ATOL}}{\mathrm{TOL}} , \quad i = 1, \ldots, n_x , \tag{3.10}$$

where $\mathrm{ATOL} > 0$ is an additional tuning parameter, named absolute tolerance.

The full discretized set-valued integration procedure is summarized in Algorithm 1. Note that this algorithm terminates with an error message as soon as the existence of the reachable set $X(t)$ can no longer be established. Such scenarios cannot be avoided, for instance if a solution trajectory $x(\cdot, p)$ fails to exist over the entire horizon $[0, T]$ for certain parameters $p \in P$ or if the enclosure size blows up due to wrapping effects. On the other hand, not only does Algorithm 1 yield a valid enclosure of the ODE solution upon successful termination, but it also provides a guarantee that the solution trajectories $x(\cdot, p)$ exist for all $p \in P \subseteq \mathrm{Im}_{\mathbb{E}_\ell}(Q_p)$.

Input: ODE (3.1) with factorable right-hand side f and initial value function x_0; affine parameterization $Q_p \in \mathbb{D}_{n_p,\ell}$ such that $\mathrm{Im}_{\mathbb{E}_\ell}(Q_p) \supseteq P$; consistency order $s \geq 1$; tolerances $\mathrm{TOL} \geq \mathrm{ATOL} > 0$; maximum and minimum step-sizes $h_{\max} \geq h_{\min} > 0$; step-size reduction parameter $0 < \rho < 1$.
Initialization:

1. Set $t = 0$, and $Q_x(t) = x_0^{\mathbb{E}_\ell}(Q_p)$.

Loop:

2. Adjust the scaling vector σ using (3.10).
3. Set the predictor $Q_x(t+h)$ for all $h \in [0, T-t]$ as in (3.5).
4. Set the step-size guess $h := \min\{\bar{h}, T-t\}$, with $\bar{h}$ given in (3.9).
   While condition (3.8) is violated, repeat $h \leftarrow \rho h$.
5. If $h < h_{\min}$, return with an error message.
6. Update $t \leftarrow t + h$
7. If $t + h = T$, return with an indication of completion; else return to step 2

Output: Enclosure function $Q_x : [0,t] \to \mathbb{D}_{n_x,\ell}$ such that $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(\tau)) \supseteq X(\tau)$ for all $\tau \in [0,t]$, with $t \leq T$.

Algorithm 1 Set-valued integration of a parametric initial value problem in ODEs for a generic affine set-parameterization.

3.4 Stability Analysis

Conditions under which the discretized set-valued integration algorithm presented above inherits the stability properties of the underlying dynamic system are investigated in this section. Of particular interest are those parametric ODEs having a unique equilibrium point $\overline{x}(p)$ for every $p \in P$. More generally, we shall consider the case that the function $f(\cdot, x, p)$ is periodic and satisfies the following assumption.

Assumption 3.4. *The parametric ODEs* (3.1) *have a unique limit cycle* $\overline{x}(\cdot, p)$ *for every* $p \in P$ *such that*

$$\forall t \in [0,T], \quad \dot{\overline{x}}(t,p) \;=\; f(t, \overline{x}(t,p), p) \quad \textit{and} \quad \overline{x}(0,p) \;=\; \overline{x}(T,p)\,, \tag{3.11}$$

for some $T > 0$.

Implicit to condition (3.11) is the requirement that the cycle time T should be identical for all $p \in P$, thereby restricting the class of periodic dynamic systems that are considered

here. Conditions under which T is independent of the time t, at least locally, are further discussed in Appendix A.

Example 3.1. *Consider the scalar differential equation*

$$\dot{x}(t) \;=\; -p_1 x(t) + \sin(p_2 t) + p_3 \;,$$

with parameters $p_1 > 0$, $p_2 \neq 0$ and $p_3 \in \mathbb{R}$. The following function

$$\overline{x}(t,p) \;:=\; \frac{1}{p_1}\left[\left(1-\frac{p_2^2}{p_1}\right)\left(\sin(p_2 t) - \frac{p_2}{p_1}\cos(p_2 t)\right) + p_3\right]$$

satisfies condition (3.11) *with the cycle time $T = \frac{2\pi}{p_2}$. In particular, the cycle time T is constant if parameter p_2 is fixed. On the other hand, Assumption 3.4 fails to hold when p_2 is allowed to vary.* ◇

Given a compact parameter host set $P \subset \mathbb{R}^{n_p}$, Algorithm 1 propagates a parameterization $Q_x(t,P) \in \mathbb{D}_{n_x,\ell}$, whose image describe a valid enclosure of the reachable parametric initial value problem (3.1), $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) \supseteq X(t,P)$, for $t \geq 0$. In analogy with the definition of local asymptotic stability in dynamic systems, we formalize the concept of local asymptotic stability for a set-valued integration algorithm next.

Definition 3.1. *Let Assumption 3.4 be satisfied, and denote $\overline{X}(t,P) := \{\overline{x}(t,p) \mid p \in P\}$ for $t \geq 0$. Algorithm 1 for set-valued integration with consistency order $s \geq 1$ is said to be* locally asymptotically stable *for the parametric ODE* (3.1) *if the following conditions are satisfied for sufficiently small local tolerance* $\mathrm{TOL} > 0$ *and maximum step size $h_{\max} > 0$:*

(i) *For every $\varepsilon > 0$ and all $t \geq 0$, there exists $\delta > 0$ such that*

$$d_{\mathrm{H}}\left(\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t,P)), \overline{X}(t,P)\right) \;=\; \varepsilon + \mathbf{O}(\mathrm{TOL}) + \mathbf{O}(h_{\max}^s) \;, \tag{3.12}$$

for all $P \subseteq \mathbb{R}^{n_p}$ with $\mathrm{diam}(P) < \delta$, and all $Q_x(0,P)$ with $d_{\mathrm{H}}(\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(0,P)), \overline{X}(0,P)) < \delta$.

(ii) *There exists $\delta > 0$ such that*

$$\limsup_{t\to\infty} \| \mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t,\{p^*\})) - \overline{x}(t,p^*) \|_{\mathrm{H}} \;=\; \mathbf{O}(\mathrm{TOL}) + \mathbf{O}(h_{\max}^s) \;, \tag{3.13}$$

for all $p \in \mathbb{R}^{n_p}$, and all $Q_x(0,\{p^*\})$ with $\| \mathrm{Im}_{\mathbb{E}_\ell}(Q_x(0,\{p^*\})) - \overline{x}(0,p^*) \|_{\mathrm{H}} < \delta$.

A key advantage of a locally asymptotically stable set-valued integrator (in the sense of Definition 3.1) is that the computed reachable set enclosures are guaranteed to remain stable on infinite time horizons when applied to a dynamic system in the neighborhood of a locally asymptotically stable periodic orbit (or locally asymptotically stable equilibrium point). Moreover, in the case that the only parametric uncertainty is via the initial condition, the enclosure $\mathrm{Im}_{\mathbb{E}_\ell}(Q_x(t,\{p^*\}))$ converges to the periodic orbit $\overline{X}(t,\{p^*\})$ for any given parameter p^*, up to a small numerical 'noise' of order $\mathbf{O}(\mathrm{TOL}) + \mathbf{O}(h_{\max}^s)$ that can be made arbitrarily small by adjusting the tuning parameters TOL and $h_{\max}$ in Algorithm 1.

Remark 3.1. *In an alternative definition of local asymptotic stability of a set-valued integrator, the term* $\mathbf{O}(h_{\max}^s)$ *could be dropped in* (3.12) *and* (3.13)*, thereby accounting for terms of order* $\mathbf{O}(\mathrm{TOL})$ *only. From a practical viewpoint, this refinement would be more appropriate for stiff dynamic systems, where a larger* $h_{\max}$ *is desirable in order to not slow down the integration process unnecessarily. The emphasis in this chapter being on explicit integration schemes, the assumption is made that a small* $h_{\max}$ *value compared to the characteristic timescales of the dynamics should not have a major impact on the performance of the integration algorithm, at least for small* TOL *values. We shall revisit this important aspect later on in Remark 3.2, once we have developed a better understanding of the mechanisms that may lead to instability in Algorithm 1.*

Although asymptotic stability of the computed reachable set enclosures would appear to be a natural property for any set-valued integration algorithm to have, all currently available set-valued integrators lack it to the best of our knowledge. This could be attributed to the fact that these integrators rely on interval arithmetics in one way or another and are thus subject to bound explosion in finite time due to wrapping effects, regardless of the size of the uncertainty set. Even state-of-the-art integrators based on Taylor models with interval remainders, such as VSPODE [80], cannot prevent bound explosion in enclosing the reachable set of asymptotically stable dynamic systems, despite the fact that they implement advanced heuristics for rotating the basis of the interval remainder.

In order to determine conditions under which a set-valued integration algorithm is asymptotically stable, we first recall basic stability results for periodic dynamic systems; see, e.g., [143] for more details. Given a parametric ODE of the form (3.1) and a limit cycle $\overline{x}(\cdot,p)$ satisfying (3.11) for $p \in P$, the so called monodromy matrix $\Phi(T,0,p)$ can be obtained as the solution of the variational differential equation for $\tau, t \in [0,T]$:

$$\frac{\partial \Phi(t,\tau,p)}{\partial t} \;=\; \frac{\partial f(t,\overline{x}(t,p),p)}{\partial x}\Phi(t,\tau,p) \quad \text{with} \quad \Phi(\tau,\tau,p) \;=\; I\,. \tag{3.14}$$

It can be shown that the periodic orbit $\overline{x}(\cdot,p)$ is locally asymptotically stable if all the eigenvalues $\lambda_i(G(T,0,p))$ of the monodromy matrix are in the open unit disk,

$$\max_{i\in\{1,\ldots,n_x\}} |\lambda_i(\Phi(T,0,p))| \;<\; 1\;. \tag{3.15}$$

The following theorem establishes quadratic Hausdorff convergence of the underlying affine set-parameterization to be the critical requirement for local asymptotic stability of Algorithm 1.

Theorem 3.2. *Let Assumption 3.4 hold, and consider an affine set-parameterization for which (i) function extensions have Hausdorff convergence order* 2 *(or higher), and (ii) the addition extension is consistent in the sense of Definition 2.5. Then, Algorithm 2 with consistency order* $s \geq 1$ *is locally asymptotically stable for the parametric ODE* (3.1)*.*

Proof. Let $\overline{x}(\cdot,p)$ satisfy condition (3.11) in Assumption 3.4 for a given $p \in P$ and be a locally asymptotically stable orbit. For any $t \in [0,T)$, any $Q_p(P) \in \mathbb{D}_{n_p,\ell}$ with sufficiently small $\left\| \operatorname{Im}_{\mathbb{E}_\ell}(Q_p(P)) - p \right\|_{\mathrm{H}}$, any $Q_x(t,P) \in \mathbb{D}_{n_x,\ell}$ with sufficiently small $\left\| \operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) - \overline{x}(t,p) \right\|_{\mathrm{H}}$, and any step-size h satisfying condition (3.8), we have:

$$\begin{aligned}
&\left\| \operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t+h,P)) - \overline{x}(t+h,p) \right\|_{\mathrm{H}} \\
&\overset{(3.5)}{=} \left\| \operatorname{Im}_{\mathbb{E}_\ell}\left(\biguplus_{i=0}^{s} h^i \phi_i^{\mathbb{E}_\ell}(t, Q_x(t,P), Q_p(P)) \oplus h\,\mathrm{TOL}\, Q_{\mathrm{unit}} \right) - \overline{x}(t+h,p) \right\|_{\mathrm{H}} \\
&\overset{(2.4)}{=} \left\| \operatorname{Im}_{\mathbb{E}_\ell}\left(\biguplus_{i=0}^{s} h^i \phi_i^{\mathbb{E}_\ell}(t, Q_x(t,P), Q_p(P)) \right) - \sum_{i=0}^{s} h^i \phi_i(\overline{x}(t,p)) \right\|_{\mathrm{H}} + h\mathbf{O}(\mathrm{TOL})\,.
\end{aligned}$$

Since the Taylor coefficient functions are continuously-differentiable and their extension have quadratic Hausdorff convergence, Proposition 2.1 gives

$$\begin{aligned}
&\left\| \operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t+h,P)) - \overline{x}(t+h,p) \right\|_{\mathrm{H}} \\
&\overset{(2.7)}{=} \left\| \sum_{i=0}^{s} h^i \frac{\partial \phi_i}{\partial x}(t,\overline{x}(t,p),p) \cdot \left[\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) - \overline{x}(t,p) \right] \right\|_{\mathrm{H}} \\
&\quad + \left\| \sum_{i=0}^{s} h^i \frac{\partial \phi_i}{\partial p}(t,\overline{x}(t,p),p) \cdot \left[\operatorname{Im}_{\mathbb{E}_\ell}(Q_p(P)) - p \right] \right\|_{\mathrm{H}} + h\mathbf{O}(\mathrm{TOL}) \\
&\quad + h\mathbf{O}\left(\left\| \operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) - \overline{x}(t,p) \right\|_{\mathrm{H}}^2 \right) + h\mathbf{O}\left(\left\| \operatorname{Im}_{\mathbb{E}_\ell}(Q_p(P)) - p \right\|_{\mathrm{H}}^2 \right).
\end{aligned} \tag{3.16}$$

Notice the absence of terms of order $\mathbf{O}\left(\left\|\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) - \overline{x}(t,p)\right\|_{\mathrm{H}}^2\right)$ or $\mathbf{O}\left(\left\|\operatorname{Im}_{\mathbb{E}_\ell}(Q_p(P)) - p\right\|_{\mathrm{H}}^2\right)$ in the right-hand side of the previous inequality—Because the Taylor coefficient function ϕ_0 is affine, its extension $\phi_0^{\mathbb{E}_\ell}$ matches the actual image set of ϕ_0 exactly, and therefore all non-trivial terms are of order $\mathbf{O}(h)$. Now, by using a Taylor expansion of the solution Φ of the variational differential equation (3.14), we have:

$$\begin{aligned}
&\left\| \left[\sum_{i=0}^{s} h^i \frac{\partial \phi_i}{\partial x}(t, \overline{x}(t,p), p) - \Phi(t+h, t, \overline{p}) \right] \cdot \left[\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) - \overline{x}(t,p) \right] \right\|_{\mathrm{H}} \\
&\quad = h\,\mathbf{O}\left(\left\|\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) - \overline{x}(t,p)\right\|_{\mathrm{H}}\right) + \mathbf{O}\left(h^{s+1}\right) \\
&\quad = h\,\mathbf{O}(\operatorname{diam}(P)) + h\,\mathbf{O}\left(h_{\max}^s\right) . \qquad (3.17)
\end{aligned}$$

In the last inequality we have used that $\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P))$ is locally Lipschitz continuous in P, which follows trivially from the fact that extensions have quadratic Hausdorff convergence in the chosen set arithmetic. Moreover, we also have:

$$\begin{aligned}
&\left\| \sum_{i=0}^{s} h^i \frac{\partial \phi_i}{\partial p}(t, \overline{x}(t,p), p) \cdot \left[\operatorname{Im}_{\mathbb{E}_\ell}(Q_p(P)) - p \right] \right\|_{\mathrm{H}} \\
&\qquad = h\,\mathbf{O}\left(\left\|\operatorname{Im}_{\mathbb{E}_\ell}(Q_p(P)) - p\right\|_{\mathrm{H}}\right) = h\,\mathbf{O}(\operatorname{diam}(P)) . \qquad (3.18)
\end{aligned}$$

Thus, combining (3.16), (3.17) and (3.18) gives

$$\begin{aligned}
&\left\|\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t+h,P)) - \overline{x}(t+h,p)\right\|_{\mathrm{H}} \\
&\quad = \left\|\Phi(t+h,t,p) \cdot \left[\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) - \overline{x}(t,p)\right]\right\|_{\mathrm{H}} \\
&\qquad + h\,\mathbf{O}\left(\left\|\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) - \overline{x}(t,p)\right\|_{\mathrm{H}}^2\right) + h\,\mathbf{O}\left(\left\|\operatorname{Im}_{\mathbb{E}_\ell}(Q_p(P)) - p\right\|_{\mathrm{H}}^2\right) \\
&\qquad + h\,\mathbf{O}(\operatorname{diam}(P)) + h\,\mathbf{O}(\mathrm{TOL}) + h\,\mathbf{O}\left(h_{\max}^s\right) ,
\end{aligned}$$

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

In a second step, we apply a discrete version of Gronwall's lemma to obtain

$$\begin{aligned}
&\left\|\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t+T,P)) - \overline{x}(t,p)\right\|_{\mathrm{H}} \\
&\quad = \left\|\Phi(t+T,t,p) \cdot \left[\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) - \overline{x}(t,p)\right]\right\|_{\mathrm{H}} \\
&\qquad + \mathbf{O}\left(\left\|\operatorname{Im}_{\mathbb{E}_\ell}(Q_x(t,P)) - \overline{x}(t,p)\right\|_{\mathrm{H}}^2\right) + \mathbf{O}\left(\left\|\operatorname{Im}_{\mathbb{E}_\ell}(Q_p(P)) - p\right\|_{\mathrm{H}}^2\right) \\
&\qquad + \mathbf{O}(\operatorname{diam}(P)) + \mathbf{O}(\mathrm{TOL}) + \mathbf{O}\left(h_{\max}^s\right) ,
\end{aligned}$$

since we can substitute $\overline{x}(t+T,p) = \overline{x}(t,p)$.

In the last step, we use induction over the number N of cycle times together with the fact that $\lim_{N\to\infty} \Phi(t+T,t,p)^N = 0$ for an asymptotically stable orbit in order to show that the above inequality implies

$$\forall t \in [0,T], \quad \lim_{N\to\infty} d_{\mathrm{H}}\left(\mathrm{Im}_{\mathbb{E}_\ell}\left(Q_x(t+NT,P)\right), \overline{X}(t+NT,P)\right) = \mathbf{O}\left(\mathrm{diam}\left(P\right)\right) + \mathbf{O}\left(\mathrm{TOL}\right) + \mathbf{O}\left(h_{\max}^{s+1}\right). \tag{3.19}$$

In particular, (3.19) implies both conditions (i) and (ii) in Definition 3.1. □

Theorem 3.2 sheds light on the most fundamental reason why existing set-valued integrators based on Taylor models (with interval remainders) fail to stabilize the computed reachable set enclosures for small enough uncertainty set, as these affine set-parameterizations have Hausdorff convergence order 1 in general (see Remark 2.2).

We also note that a set-valued integration algorithm may not inherit the stability properties of those dynamic systems that are locally stable, but not locally asymptotically stable. When invariants are known explicitly for such systems though, e.g., based on the underlying conservation laws in the case of physical systems, these invariants can sometimes be used to formulate an equivalent, reduced dynamic system that is locally asymptotically stable.

Finally, it is worth noting that the foregoing stability analysis can be extended to the case that the initial set $X_0 = \{x_0(p) \mid p \in P\}$ is not necessarily close to the periodic orbit. More specifically, if for each $p \in P$ the response trajectory $x(\cdot,p)$ reaches an attractive neighborhood of the periodic limit orbit $\overline{X}(t,P)$ after a finite transition time, then Algorithm 1 shall remain stable on infinite horizons as long as the diameter of the parameter host set P is sufficiently small.

Remark 3.2. *As a final note in connection to Remark 3.1, it is interesting to observe that the term* $\mathbf{O}\left(h_{\max}^{s}\right)$ *is introduced in* (3.18)*. The reason that the discretization error of the variational equation cannot be bounded with a term of order* $\mathbf{O}\left(\mathrm{TOL}\right)$ *is because the step-size control mechanism in Algorithm 1, although it accounts for discretization errors in the nominal state trajectories rigorously, does not account for discretization errors in the associated variational equation on the other hand. In particular, this problem could be resolved by appending the following semi-infinite constraint to the step-size optimization problem* (3.6)*:*

$$\forall \tau \in [t,t+h], \quad \mathrm{Im}_{\mathbb{E}_\ell}\left((\tau-t)^s \frac{\partial \phi_{s+1}}{\partial x}^{\mathbb{E}_\ell}(Q_x(\tau)) \right) \subseteq \mathrm{TOL}\,\mathrm{Im}_{\mathbb{E}_\ell}\left(Q'_{\mathrm{unit}}\right),$$

for a suitable matrix $Q'_{\text{unit}} \in \mathbb{D}_{n_x \times n_x, \ell}$. However, drawbacks of this approach include the need to compute enclosures of the Jacobian matrix of ϕ_{s+1}, which can be computationally demanding in practice, as well as the extra conservatism it would introduce on the selected step-size. A potentially more efficient way of enforcing stability irrespective of the step-size control mechanism would involve developing implicit schemes for set-valued numerical ODE integration, which will be the topic for further research.

3.5 Contraction of the Enclosure using Invariants

In order to improve the stability of the enclosures as well as to tighten them further, suppose now that an invariant for the parametric ODE system is known, namely a function $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}$ such that $h(x(t,p),p) = 0$ for any solution $x(t,p)$ of (3.1). Given an affine-parameterization basis $\mathbb{E}_{n_p}$, it follows that an extension $h^{\mathbb{E}_{n_p}}$ of h should satisfy

$$\forall t \geq 0, \quad \text{Im}_{\mathbb{E}_{n_p}} \left(h^{\mathbb{E}_{n_p}}(Q_x(t), Q_p) \right) = \{0\},$$

for given parameterizations Q_p and $Q_x(t)$ of the parameter and reachable sets.

Consider the special case of qth-order Taylor models with ellipsoidal remainders, such that $[\mathcal{P}_x(t), \mathcal{R}_x(t)] := Q_x(t)$ and $[\mathcal{P}_h(t), \mathcal{R}_h(t)] := h^{\mathbb{E}_{np}^{\text{pol}(q)} \times \mathbb{E}_{np}^{\text{ball}}}(Q_x(t), Q_p)$. The polynomial part $\mathcal{P}_h(t)$ is trivially equal to zero by construction. On application of Algorithm 1 in [53], the ellipsoidal remainder is constructed such that:

$$\text{Im}_{\mathbb{E}_{np}^{\text{ball}}}(\mathcal{R}_h(t)) \supseteq \text{Im}_{\mathbb{E}_{np}^{\text{ball}}} \left(A_h(t)\, \mathcal{R}_x(t)\, A_h(t)^{\mathsf{T}} \right) \oplus N_h(t)\,, \tag{3.20}$$

where $A_h(t)$ is the Jacobian matrix of h evaluated along the solution trajectory for some $\hat{p} \in P$, and $N_h(t)$ is an interval nonlinearity bounder. For linear invariants in particular, we have $N_h(t) = \{0\}$ and the ellipsoidal remainder $\text{Im}_{\mathbb{E}_{np}^{\text{ball}}}(\mathcal{R}_h(t))$ can be thus be safely intersected with the hyperplane $\mathcal{H}_h(t) := \{x \in \mathbb{R}^{n_x} \mid A_h(t)^{\mathsf{T}} x = 0\}$. This intersection is simply repeated multiple times when several invariants are known.

3.6 Numerical Implementation and Case Studies

The main objective of the numerical case study in this final section is merely to illustrate the ability of the developed set-valued ODE integrator to stabilize the reachable set enclosures—not to provide a detailed computational study. Our implementa-

tion of Algorithm 1 comes in the form of a C++ class called `ODEBND_VAL` as part of the CRONOS library, which is made freely available at: http://omega-icl.bitbucket.org/cronos/. It uses the affine set-parameterization $(\mathbb{E}_\ell^{\mathrm{pol}(q)} \times \mathbb{E}_{n_x}^{\mathrm{ball}}, \mathbb{R}^{n_x \times (\alpha_\ell^{(q)} + n_x + 1)})$ of polynomial models combined with ellipsoids, and computes quadratically Hausdorff convergent extensions based on Theorem 2.1; see [53] for more details on the construction of Taylor models with ellipsoidal remainder. In particular, we use the verified library PROFIL (`http://www.ti3.tu-harburg.de/`) for interval analysis and the library MC++ (http://omega-icl.bitbucket.org/mcpp/) [93] for Taylor model arithmetic. The Taylor expansion in time of the ODE solutions are constructed using automatic differentiation in FADBAD++ (http://www.fadbad.com/fadbad.html). Unless otherwise noted, the consistency order is set to $s = 5$, the tolerances to $\mathrm{TOL} = 10^{-7}$ and $\mathrm{ATOL} = 10^{-8}$, and the step-size control parameter to $\rho = 0.8$.

3.6.1 Cubic Oscillator

We consider the following cubic oscillator system:

$$\dot{x}_1(t) = x_2(t) + \frac{1}{10}(1 - x_1(t)^2 - x_2(t)^2)\,x_1(t) \qquad \text{with } x_1(0) = p_1 \tag{3.21}$$

$$\dot{x}_2(t) = -x_1(t) + \frac{1}{10}(1 - x_1(t)^2 - x_2(t)^2)\,x_2(t) - \frac{1}{5}x_2(t) \qquad \text{with } x_2(0) = p_2\,, \tag{3.22}$$

where both parameters p_1 and p_2 are uncertain, given by $p \in P := [1.5, 3] \times [-0.1, 0.1] \subseteq \mathbb{R}^2$.

The results obtained by running the set-valued integrator with 4th-order Taylor models are given in Fig. 3.1, with projections on the state components $x_1(t)$ and $x_2(t)$ shown on the left plot and the right plot, respectively. Both plots illustrate that the computed set $\mathrm{Im}_{\mathbb{E}_\ell^{\mathrm{pol}(4)} \times \mathbb{E}_2^{\mathrm{ball}}}(Q_x(\cdot, P))$ validly enclose the actual reachable set on the time horizon $[0, 20]$, with very small overestimation in this case.

The behavior of the set-valued integrator on the extended time horizon $t \in [0, 400]$ is shown in Fig. 3.2. Also reported on this figure are the bounds computed using VSPODE (v1.4) [80], with the order of the Taylor expansion in time and of the Taylor model set to 5 and 4, respectively, in order to enable direct comparison, and selecting QR-factorization with row permutation as the wrapping mitigation strategy. These results demonstrate the ability of Algorithm 1 to stabilize the reachable set enclosures when propagating Taylor models with ellipsoidal remainders. In contrast, the bounds computed using VSPODE eventually explode around $t = 180$, a behavior attributed to the propagation of Taylor models

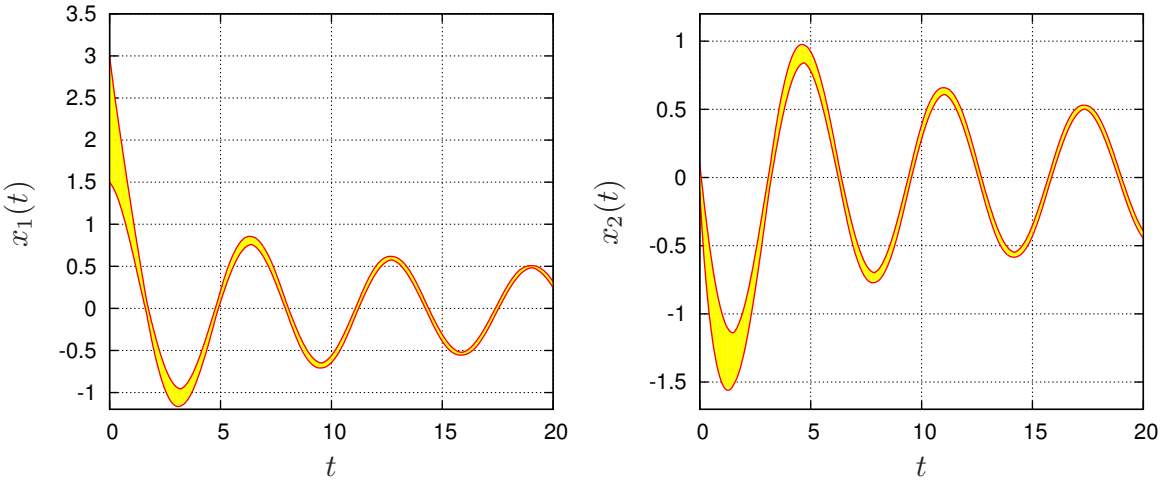

Fig. 3.1 Projections of the exact reachable set $X(t)$ (shaded area) on the state component $x_1(t)$ (left plot) and the state component $x_2(t)$ (right plot). The solid lines show the upper and lower bounds computed with Algorithm 1 over the time horizon $[0,20]$.

with interval remainder, which only enjoy linear Hausdorff convergence in the sense of Definition 2.7 (see also Remark 2.2).

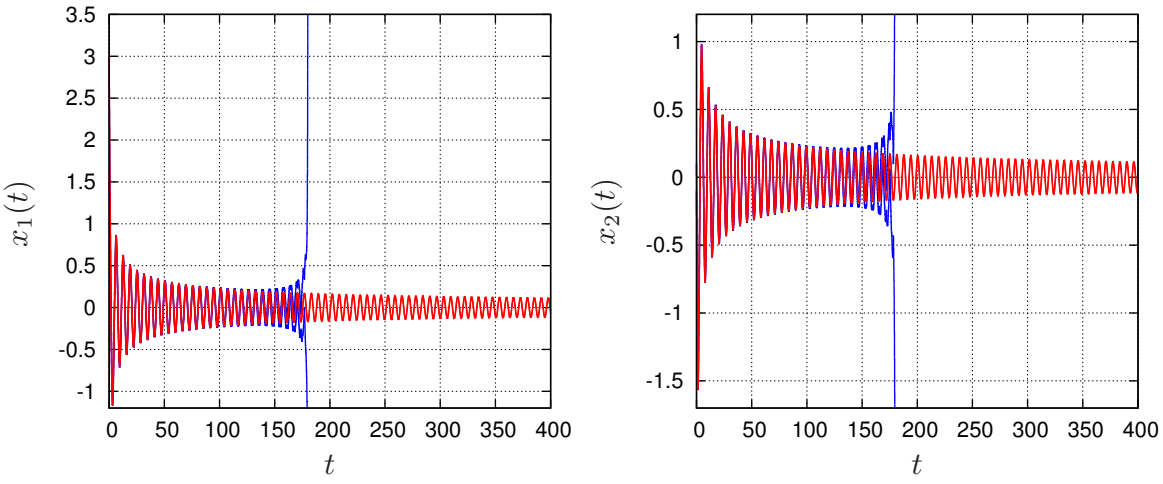

Fig. 3.2 Projections on the state component $x_1(t)$ (left plot) and the state component $x_2(t)$ (right plot) of the upper and lower bounds computed with Algorithm 1 (red lines) and VSPODE (blue lines) over the extended time horizon $[0,400]$.

Closely related to stability is the question how closely the computed enclosure $\mathrm{Im}_{\mathbb{E}_\ell^{\mathrm{pol}(4)}\times\mathbb{E}_2^{\mathrm{ball}}}(Q_x(\cdot,P))$ approximates the actual reachable set $X(t,P)$. This approximation error can be quantified as the diameter of the ellipsoidal remainder $\mathrm{Im}_{\mathbb{E}_2^{\mathrm{ball}}}(Q_x^{\mathrm{rem}}(\cdot,P))$, with $Q_x^{\mathrm{rem}}(t,P)\in\mathbb{R}^{2\times 3}$ such that $[Q_x^{\mathrm{pol}}(t,P),Q_x^{\mathrm{rem}}(t,P)] := Q_x(t,P)$. The left plot on Fig. 3.3 represents the evolution of such a diameter along the time horizon for various Taylor model

expansion orders $q = 2,\ldots,5$. In the case of $q = 2$, Algorithm 1 stops with an error message around $t = 35$, after a dramatic increase in the diameter of the ellipsoidal remainder. This is due to the fact that, in this instance, the inherent stability of the cubic oscillator system (3.21) is over-powered by the wrapping effect inherent to the ellipsoidal remainder term of the Taylor model (for the given uncertainty set P). In contrast, the reachable set enclosures can be stabilized with higher-order expansions $q \geq 3$; that is, Algorithm 1 can in principle propagate these enclosures *ad infinitum*. It is also seen that the approximation error can be reduced by increasing the expansion order q of the Taylor model. While this trend could be confirmed for other examples as well, the question whether or not the approximation error converges to zero in the limit remains open.

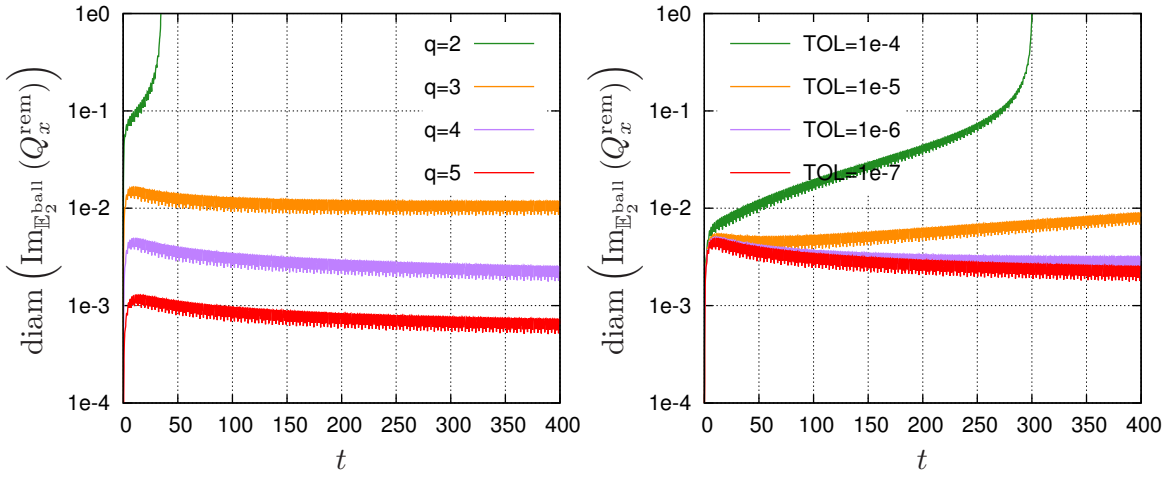

Fig. 3.3 Diameter of the remainder ellipsoid in the parameterization of the reachable set enclosure diam(Im$_{\mathbb{E}_2^{ball}}$ (Q_x^{rem})) computed with the set-valued integrator for various Taylor model orders $q = 2,\ldots,5$ (left plot) and various tolerances TOL $= 10^{-4},\ldots,10^{-7}$ (right plot).

Finally, the right plot on Fig. 3.3 represents the evolution of the diameter of the ellipsoidal remainder for various tolerance values TOL $= 10^{-4},\ldots,10^{-7}$ in the case of 4th-order Taylor models ($q = 4$). These results illustrate the effect of the parameter TOL on the integrator stability. In agreement with the asymptotic stability conditions in Definition 3.1, too large a value for TOL can indeed lead to instability of the computed enclosures and bound explosion in finite time.

3.6.2 Anaerobic Digestion

We consider a six-state model representing the dynamics of an anaerobic digester, as originally proposed by Bernard et al. [11]. Enclosing the solutions of this model in the presence

of parametric uncertainty is challenging due to the presence of complex and liquid-gas transfer and pH self-regulation mechanisms. Moreover, the system exhibits both fast dynamics acting on a time-scale of minutes/hours, and slow dynamics acting on a time-scale of days.

$$\dot{X}_1 = (\mu_1(S_1) - \alpha D) X_1, \tag{3.23a}$$
$$\dot{X}_2 = (\mu_2(S_2) - \alpha D) X_2, \tag{3.23b}$$
$$\dot{S}_1 = D(S_1^{\text{in}} - S_1) - k_1 \mu_1(S_1) X_1, \tag{3.23c}$$
$$\dot{S}_2 = D(S_2^{\text{in}} - S_2) + k_2 \mu_1(S_1) X_1 - k_3 \mu_2(S_2) X_2, \tag{3.23d}$$
$$\dot{Z} = D(Z^{\text{in}} - Z), \tag{3.23e}$$
$$\dot{C} = D(C^{\text{in}} - C) - q_{\text{CO}_2} + k_4 \mu_1(S_1) X_1 + k_5 \mu_2(S_2) X_2 \,. \tag{3.23f}$$

The states X_1 and X_2 stand for the concentrations of acidogenic and methanogenic biomass, respectively; S_1, the organic substrate concentration (COD other than VFA); S_2, the volatile fatty acids (VFA) concentration; Z, the total alkalinity concentration (TALK); and C, the total inorganic carbon concentration (TIC). Moreover, D represents the dilution rate; S_1^{in}, S_2^{in}, Z^{in} and C^{in} are the inlet concentrations of organic substrate, VFA, TALK and TIC, respectively; α is the fraction of biomass in the liquid phase (i.e., not attached to a support); and $k_1, \ldots, k_6$ are pseudo-stoichiometric yield coefficients.

The specific growth rates of acidogenic bacteria, μ_1, and methanogenic bacteria, μ_2, are assumed to follow Michaelis-Menten and Haldane kinetics,

$$\mu_1(S_1) := \bar{\mu}_1 \frac{S_1}{S_1 + K_{S_1}}, \tag{3.23g}$$
$$\mu_2(S_2) := \bar{\mu}_2 \frac{S_2}{S_2 + K_{S_2} + S_2^2/K_{I_2}}, \tag{3.23h}$$

with maximum growth rates $\bar{\mu}_1$ and $\bar{\mu}_2$, half-saturation constants K_{S_1} and K_{S_2}, and inhibition constant K_{I_2} (methanogenic bacteria only). Finally, the molar flowrate of CO_2, q_{CO_2}, is given by

$$q_{\text{CO}_2} := k_{\text{L}} a (C + S_2 - Z - K_{\text{H}} P_{\text{CO}_2}), \tag{3.23i}$$
$$\text{with} \quad P_{\text{CO}_2} := \frac{\phi_{\text{CO}_2} - \sqrt{\phi_{\text{CO}_2}^2 - 4 K_{\text{H}} P_{\text{t}} (C + S_2 - Z)}}{2 K_{\text{H}}} \tag{3.23j}$$
$$\phi_{\text{CO}_2} := C + S_2 - Z + K_{\text{H}} P_{\text{t}} + \frac{k_6}{k_{\text{L}} a} \mu_2(S_2) X_2 \,, \tag{3.23k}$$

where $k_L a$ denotes the liquid-gas transfer constant, K_H is Henry's constant, and P_t is the total pressure. This case study, will be considered repeatedly throughout this thesis, the parameters of the model taken from [11] and summarized in Table 3.1 for the sake of reproducibility.

Table 3.1 Parameters in the anaerobic digestion model (4.62,4.72).

Parameter	Value		Parameter	Value	
$\bar{\mu}_1$	1.2	day^{-1}	k_1	42.14	$\mathrm{g(COD)\,g(cell)}^{-1}$
K_{S_1}	7.1	$\mathrm{g(COD)\,L}^{-1}$	k_2	116.5	$\mathrm{mmol\,g(cell)}^{-1}$
$\bar{\mu}_2$	0.74	day^{-1}	k_3	268.0	$\mathrm{mmol\,g(cell)}^{-1}$
K_{S_2}	9.28	$\mathrm{mmol\,L}^{-1}$	k_4	50.6	$\mathrm{mmol\,g(cell)}^{-1}$
K_{I_2}	256	$\mathrm{mmol\,L}^{-1}$	k_5	343.6	$\mathrm{mmol\,g(cell)}^{-1}$
$k_L a$	19.8	day^{-1}	k_6	453.0	$\mathrm{mmol\,g(cell)}^{-1}$
K_H	16	$\mathrm{mmol\,L}^{-1}\,\mathrm{atm}^{-1}$	S_1^{in}	5	$\mathrm{g(COD)\,L}^{-1}$
P_t	1	atm	S_2^{in}	80	$\mathrm{mmol\,L}^{-1}$
α	0.5	–	Z^{in}	50	$\mathrm{mmol\,L}^{-1}$
D	0.4	day^{-1}	C^{in}	0	$\mathrm{mmol\,L}^{-1}$

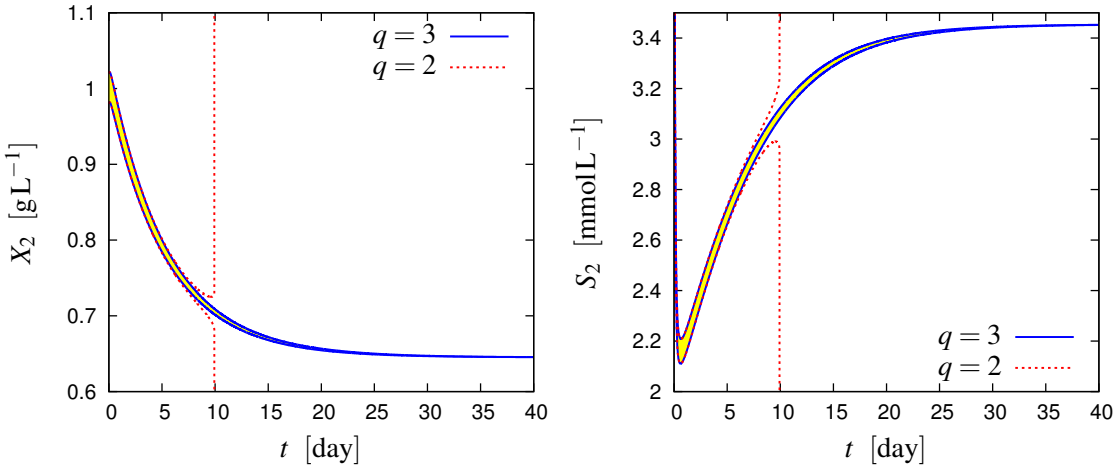

Fig. 3.4. Reachable set enclosure projections for the variables X_2 and S_2.

Enclosures of the reachable set obtained with 2nd- and 3rd-order Taylor models with ellipsoidal remainders are shown in Fig. 3.4, for uncertain initial values given by $X_1(0) \in 0.5 \times [0.98, 1.02]$, $X_2(0) \in [0.98, 1.02]$, $S_1(0) = 1$, $S_2(0) = 5$, $Z(0) = 50$, and $C(0) = 40$. 3rd-order (or higher-order) Taylor models successfully stabilize the reachable set enclosure here, whereas 2nd-order Taylor models fail to do so for this level of uncertainty. In the latter case, stabilizing the enclose would require reducing the uncertainty set further.

3.6.3 Reversible Chemical Reactions

Consider the reversible reactions $\mathsf{A}+\mathsf{B} \rightleftharpoons \mathsf{C}$ and $\mathsf{A}+\mathsf{C} \rightleftharpoons \mathsf{D}$ in a batch reactor, as described by the following dynamic model:

$$\begin{aligned}
\dot{x}_\mathsf{A} &= -r_1(x_\mathsf{A}, x_\mathsf{B}, x_\mathsf{C}) \\
\dot{x}_\mathsf{B} &= -r_1(x_\mathsf{A}, x_\mathsf{B}, x_\mathsf{C}) \\
\dot{x}_\mathsf{C} &= r_1(x_\mathsf{A}, x_\mathsf{B}, x_\mathsf{C}) - r_2(x_\mathsf{A}, x_\mathsf{C}, x_\mathsf{D}) \\
\dot{x}_\mathsf{D} &= r_2(x_\mathsf{A}, x_\mathsf{C}, x_\mathsf{D})
\end{aligned}$$

with

$$\begin{aligned}
r_1(x_\mathsf{A}, x_\mathsf{B}, x_\mathsf{C}) &:= k_1^\mathrm{f} x_\mathsf{A} x_\mathsf{B} - k_1^\mathrm{r} x_\mathsf{C} \\
r_2(x_\mathsf{A}, x_\mathsf{B}, x_\mathsf{C}) &:= k_2^\mathrm{f} x_\mathsf{A} x_\mathsf{B} - k_2^\mathrm{r} x_\mathsf{C}
\end{aligned}$$

Based on mass-conservation considerations, it is not hard to see that the functions $h_1(x) := x_\mathsf{B} + x_\mathsf{C} + x_\mathsf{D}$ and $h_2(x) := x_\mathsf{A} - x_\mathsf{B} + x_\mathsf{D}$ are both linear solution invariants for the ODE system, i.e., $\dot{h}_1(x) = \dot{h}_2(x) = 0$ for all $t \geq 0$. Such invariants are typical in chemical reaction systems [e.g., 140, 130]. Mixed uncertainty in the initial values and kinetic parameters is considered here, with $x_\mathsf{A}(0) = 1$, $x_\mathsf{B}(0) \in [0.95, 1.05]$, $x_\mathsf{C}(0) = x_\mathsf{D}(0) = 0$, $k_1^\mathrm{f} \in [50, 60]$, $k_2^\mathrm{f} = 20$, and $k_1^\mathrm{r} = k_2^\mathrm{r} = 1$.

Enclosures of the reachable set obtained with 3rd-order Taylor models with ellipsoidal remainders are shown in Fig. 3.5, with and without accounting for the invariants. On account of the invariants the set-valued integrator is able to stabilize the reachable set enclosure, whereas it fails to do so for this level of uncertainty when the invariants are ignored.

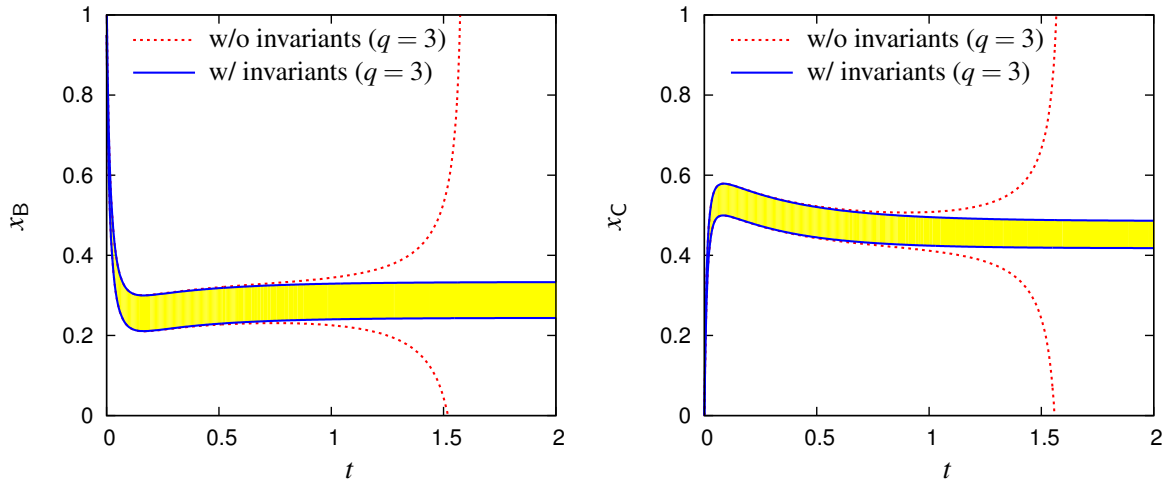

Fig. 3.5. Reachable set enclosure projections for the variables x_A and x_B.

3.7 Conclusion

This chapter has presented a new discretized set-valued integration algorithm for parametric initial value problems in ODEs. This algorithm uses a predictor-validation approach to propagate generic affine set-parameterizations, whose images are guaranteed to enclose the ODE reachable tube by Theorem 3.1. Sufficient conditions have also been derived in Theorem 3.2 for this algorithm to be locally asymptotically stable, in the sense that the computed reachable set enclosures are guaranteed to remain stable on infinite time horizons when applied to a dynamic system in the neighborhood of a locally asymptotically stable periodic orbit (or equilibrium point). These stability properties have been illustrated for a simple cubic oscillator and a more challenging dynamic system representing an anaerobic digestion process. Another contribution has been incorporating ODE invariants in order to further improve the stability of the algorithm, which has been illustrated using a chemical reaction system.

The techniques presented in this chapter have been implemented in an in-house library named CRONOS based on the library MC++, both freely available from: http://omega-icl.bitbucket.org/cronos. Another open-source implementation is available as a sub-package in ACADO Toolkit [55] at: http://www.acadotoolkit.org.

Chapter 4

Continuous-Time Set Propagation

This chapter presents a framework for constructing and analyzing enclosures of the reachable sets of nonlinear ordinary differential equations using continuous-time set propagations methods. Although auxiliary ODEs whose solutions are guaranteed to enclose the reachable set of the dynamic system can be constructed using affine set parameterizations, this framework exploits facet constraints of the system restricting the growth of the reachable set. This is accomplished by focusing on convex enclosures which can be uniquely characterized in terms of their support functions. A generalized differential inequality is introduced, whose solutions describe such support functions for a convex enclosure of the reachable set under mild conditions. This generalized differential inequality is shown to contain as special cases existing continuous-time set propagation strategies such as differential inequalities and ellipsoidal set-propagation. An extension of this frameworks for nonconvex sets using polynomial (Taylor) models with convex remainders is also provided. The framework also provides a means for analyzing the convergence properties of the enclosures constructed using the generalized differential inequality. A non trivial extension for the characterization of robust forward invariant tubes for control-affine nonlinear systems is also presented. This extension is given in terms of a min-max differential inequality for the support function of convex enclosures having smooth boundary with positive curvature. As a by product of this construction nonlinear feedback control functions for the system are constructed by exploiting the structure of the boundary of the set. A tube-based robust model predictive control strategy based on this min-max differential equation is also introduced and a tractable formulation for robust forward invariant tubes is provided.

The remainder of this chapter is organized as follows. In Section 4.1 the problem stated in Chapter 3 is briefly recalled in the context of continuous-time methods, together with the preliminary material needed in this chapter. Section 4.2 presents a brief review of existing

continuous-time convex set propagation methods, namely standard differential inequalities and the ellipsoidal propagation approach. Section 4.3 introduces the main contribution of this chapter, the generalized differential inequality for the propagation of general convex sets through ODEs. Section 4.4 presents some applications of the generalized differential inequality. In particular we show that the results of standard differential inequalities and the ellipsoidal approach are implied by the generalized differential inequality. In this section we also provide an approach for the construction of nonconvex set propagation methods based on polynomial models with convex remainders. Section 4.5.1 presents the convergence analysis of the enclosures constructed using the generalized differential inequality. Section 4.6 discusses the numerical implementation of the bounding methods and illustrates their bounding capabilities and convergence properties using a model for an anaerobic digestion process. In Section 4.7 the characterization of robust forward invariant tubes for control-affine nonlinear system in terms of a min-max differential inequality is introduced. Section 4.8 presents practical conditions for the construction of robust forward invariant tubes with ellipsoidal cross-sections. Section 4.9 introduces a strategy for robust tube-based model predictive control using the min-max differential inequality. In this section an ellipsoidal tube-based MPC procedure is illustrated for a spring-mass-damper system. Finally, Section 4.10 concludes the chapter.

4.1 Problem Definition and Preliminaries

In this chapter we consider again the problem of enclosing the reachable set of parametric ODEs given respectively by Equations (3.1) and (3.2), which are recalled here in order for this chapter to be self-contained,

$$\forall t \in [0,T] \quad \dot{x}(t,p) = f(t,x(t,p),p) \quad \text{with} \quad x(0,p) = x_0(p)\,. \tag{4.1}$$

The reachable set of the initial value problem (4.1) is also denoted by

$$X(t,P) := \{x(t,p) \mid p \in P\}. \tag{4.2}$$

Throughout the paper, we consider solutions to these auxiliary ODEs in the *extended sense*, whereby a solution can be any absolutely continuous function satisfying the initial condition and the ODE almost everywhere on $[0,T]$; see, e.g., [28, Theorem 1.1]. More precisely, we constrain ourselves to the case of locally Lipschitz-continuous functions, which are sufficient for most practical purposes.

The problem addressed in this chapter is the computation of time-varying enclosures $Y(t,P) \supseteq X(t,P)$, for $t \in [0,T]$, using continuous-time set-propagation techniques. Specifically, these enclosures are constructed from the solutions of auxiliary ODEs. In connection to Chapters 2 and 3 we can consider the enclosure $Y(t)$ as a parameterization of the form $Y(t) = \mathrm{Im}_{\mathbb{E}_{n_x,\ell}}(Q_x)$ and propagate the parameterization through ODEs of the form:

$$\dot{Q}_x(t) = F(Q_x(t), Q_p), \tag{4.3}$$

with initial parameterization $Q_x(0) := x_0^{\mathbb{E}_{n_x,\ell}}(Q_p)$. Clearly a possible choice for the right-hand side function F in Eq. (4.3) is the extension $f^{\mathbb{E}_{n_x,\ell}}$ of the original right-hand side f in Eq. (4.1). However, it is useful in practice to account for certain facet constraints that mitigate the growth of the enclosure.

In this chapter, the affine-set formalism for representing sets is dropped. In its place we choose to represent the set $Y(t)$ using support functions. Recall that the support function of a compact set is the maximum extension of a set in a given direction. More precisely, the support function of $Z \in \mathbb{K}^{n_x}$ in the direction $c \in \mathbb{R}^{n_z}$ is given by

$$V[Z](c) := \max_{z \in Z} c^{\mathsf{T}} z \,. \tag{4.4}$$

In particular, there exists positive constants $C_1 \leq C_2 < \infty$ such that

$$\forall Z \in \mathbb{K}^n_{\mathrm{C}}, \quad C_1 \operatorname{diam}(Z) \;\leq\; \max_{c \in \mathbb{R}^n, \|c\| \leq 1} V[Z](c) + V[Z](-c) \;\leq\; C_2 \operatorname{diam}(Z) \,. \tag{4.5}$$

This result follows from the definition of the support function and the diameter of a set, using the equivalence of norms in finite dimensional vector spaces.

The support function is important since it can be used to uniquely characterize a convex and compact set. For example, assuming Z is compact and convex we can reconstruct it using its support function as

$$Z = \bigcap_{c \in \mathbb{R}^{n_z} \setminus \{0\}} \left\{ z \in \mathbb{R}^{n_z} \middle| c^{\mathsf{T}} z \leq V[Z](c) \right\} .$$

In this chapter, we provide a set of sufficient conditions for a set valued function $Y : [0,T] \to \mathbb{K}^{n_x}_C$ to enclose the reachable set $X(t)$ for all $t \in [0,T]$. These conditions are given in terms of the support function of $Y(t)$ and thus they are not amenable for numerical computations —since they need to hold for every direction vector $c \in \mathbb{R}^{n_x} \setminus \{0\}$. Nevertheless,

for some sets the support function has an analytical expression, enabling the construction of conditions which can be satisfied computationally.

When considering sets in $\mathbb{K}_C^n$ we often require some assumptions regarding the regularity of its boundary, which for $Z \in \mathbb{K}_C^n$ is denoted by $\mathrm{bd}\,Z$. Given a set $Z \in \mathbb{K}_C^n$ a point $z \in \mathrm{bd}\,Z$ is called regular if there is an unique supporting hyperplane of Z through that point, otherwise the point is called singular. The set Z is said to be strictly convex if each of its supporting hyperplanes meets $\mathrm{bd}\,Z$ at exactly one point $z \in \mathrm{bd}\,Z$ and it is called smooth if $\mathrm{bd}\,Z$ is a smooth submanifold of $\mathbb{R}^{n_z}$. Furthermore the boundary of any smooth set Z can be defined implicitly as $\mathrm{bd}\,Z := \{z \in \mathbb{R}^{n_z} | g(z) = 0\}$, for some smooth function $g : \mathbb{R}^{n_z} \to \mathbb{R}$. In particular if $\mathrm{bd}\,Z$ has only regular points, then g is continuously differentiable.

Let $\mathcal{S}^{n_z-1}$ be the unit sphere in $\mathbb{R}^{n_z}$, $Z \in \mathbb{K}_C^n$ and $\mathrm{bd}\,Z$ have only regular points, then the Gauss map $\mathcal{G}_Z : \mathrm{bd}\,Z \to \mathcal{S}^{n_z-1}$ is a continuous function assigning to every boundary point $z \in \mathrm{bd}\,Z$ its unique unit outer normal. If in addition Z is strictly convex, the Gauss map has a continuous inverse, i.e. $\mathrm{bd}\,Z$ is homeomorphic to $\mathcal{S}^{n_z-1}$.

The Gauss map of a smooth set $Z \in \mathbb{K}_C^{n_z}$ is given by

$$\forall \zeta \in \mathrm{bd}\,Z: \quad \mathcal{G}_Z(\zeta) = \left\| \frac{\partial g}{\partial z}(\zeta) \right\|^{-1} \frac{\partial g}{\partial z}(\zeta)\,.$$

Moreover, its differential $\partial \mathcal{G}_Z / \partial z(\zeta)$ defines a linear operator from $T_\zeta Z$ the tangent space of Z at ζ onto itself. If for all $\zeta \in \mathrm{bd}\,Z$

$$w^\mathsf{T} \frac{\partial \mathcal{G}_Z}{\partial z}(\zeta) w > 0\,, \quad \forall w \in T_\zeta Z \setminus \{0\}\,,$$

Z is said to have positive curvature. Notice that a smooth set Z with positive curvature is also strictly convex.

4.2 Review of Existing Convex Set Propagation Methods

4.2.1 Differential Inequalities

The theory of differential inequalities provides sufficient conditions for time-varying interval bounds $Y(t) := \left[y^\mathrm{L}(t), y^\mathrm{U}(t)\right]$ to yield an enclosure of a parametric ODE's reachable set $X(t)$. These conditions can be used for propagating an interval enclosure $Y(t)$ of $X(t)$ along the integration horizon.

Theorem 4.1. *Consider the initial value problem* (4.1)*, and assume that the right-hand side function f is jointly continuous in (t,x,p) and locally Lipschitz-continuous in x uniformly on $[0,T]\times P$, with $P\subset\mathbb{R}^{n_p}$ compact, and the initial value function x_0 is continuous on P. Let the functions $y^{\mathrm{L}},y^{\mathrm{U}}:[0,T]\to\mathbb{R}^{n_x}$ be Lipschitz-continuous, with $y^{\mathrm{L}}(t)\leq y^{\mathrm{U}}(t)$, and satisfy*

$$\text{a.e. } t\in[0,T]\,,\quad \dot{y}_i^{\mathrm{L}}(t)\leq \min_{\xi,\rho}\left\{ f_i(t,\xi,\rho)\,\middle|\, \begin{array}{l}\xi_i=y_i^{\mathrm{L}}(t)\\ \xi\in\left[y^{\mathrm{L}}(t),y^{\mathrm{U}}(t)\right]\\ \rho\in P\end{array}\right\} \tag{4.6}$$

$$\text{and}\quad \dot{y}_i^{\mathrm{U}}(t)\geq \max_{\xi,\rho}\left\{ f_i(t,\xi,\rho)\,\middle|\, \begin{array}{l}\xi_i=y_i^{\mathrm{U}}(t)\\ \xi\in\left[y^{\mathrm{L}}(t),y^{\mathrm{U}}(t)\right]\\ \rho\in P\end{array}\right\}, \tag{4.7}$$

for each $i=1,\ldots,n_x$, with $[y^{\mathrm{L}}(0),y^{\mathrm{U}}(0)]\supseteq\{x_0(p)\mid p\in P\}$. Then, $X(t)\subseteq\left[y^{\mathrm{L}}(t),y^{\mathrm{U}}(t)\right]$ for all $t\in[0,T]$.

This result is well established in the more general setting of a continuous right-hand side function f; see, e.g., [71, 154] for ODE systems without parameters, where the differential inequalities are stated in terms of left-sided Dini derivatives and with strict inequalities. The corresponding result with weak inequalities follows from a uniqueness argument, such as (local) Lipschitz-continuity of the ODE right-hand side [45, 133]. The parameter-dependent case is also addressed in [138, 133].

Revisiting the classical theory of differential inequalities with stronger assumptions has two principal motivations. The first one is related to well-posedness of the initial value problems, as uniqueness of the ODE solutions is required in most practical applications. The second one is concerned with numerical computation of the bounding trajectories, which usually applies interval analysis to estimate the minimum and maximum values in the right-hand sides of (4.6,4.7). From this standpoint, (local) Lipschitz-continuity of the right-hand side function f and of the bounding trajectories y^{L} and y^{U} as well as weak differential inequalities are certainly justified. A proof of these results will be obtained in Sect. 4.4.1 as a corollary of the generalized differential inequalities introduced in Theorem 4.3.

A practical application of Theorem 4.1 is in propagating a continuous-time enclosure of the reachable set by considering equalities in (4.6,4.7) and possibly overestimating the right-hand sides using interval analysis. This approach presents the advantage that the number of ODEs in the auxiliary bounding system is $2n_x$, and therefore the size of the bounding system is proportional to the number of state variables in the original ODE system. This

makes the approach computationally tractable for large-scale systems, at least in principle. Nonetheless, an important limitation with differential inequalities-based bounds is that they often suffer large overestimation, thus providing a poor approximation of the actual reachable set $X(t)$ or even blowing up to infinity in finite time. This instability can even occur in the case of linear ODEs, as illustrated in the following example.

Example 4.1. *Consider the linear differential equations*

$$\begin{aligned} \dot{x}_1(t,p) &= x_2(t,p) \quad \textit{with} \quad x_1(0,p) = p_1 \\ \dot{x}_2(t,p) &= -x_1(t,p) \quad \textit{with} \quad x_2(0,p) = p_2 \end{aligned} \tag{4.8}$$

with $p \in P := [-1,1]^2$. The solutions of this Hamiltonian system remain bounded, although not asymptotically stable. An exact expression for the reachable set is:

$$X(t) = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix} X(0)\,. \tag{4.9}$$

On the other hand, the tightest possible bounding trajectories $y^{\mathrm{L}}, y^{\mathrm{U}}$ satisfying Theorem (4.1) *are the solutions of the auxiliary ODE system*

$$\begin{aligned} \dot{y}_1^{\mathrm{L}}(t) &= y_2^{\mathrm{L}}(t) \quad \textit{with} \quad y_1^{\mathrm{L}}(0) = p_1^{\mathrm{L}} \\ \dot{y}_1^{\mathrm{U}}(t) &= y_2^{\mathrm{U}}(t) \quad \textit{with} \quad y_1^{\mathrm{U}}(0) = p_1^{\mathrm{U}} \\ \dot{y}_2^{\mathrm{L}}(t) &= -y_1^{\mathrm{U}}(t) \quad \textit{with} \quad y_2^{\mathrm{L}}(0) = p_2^{\mathrm{L}} \\ \dot{y}_2^{\mathrm{U}}(t) &= -y_1^{\mathrm{L}}(t) \quad \textit{with} \quad y_2^{\mathrm{U}}(0) = p_2^{\mathrm{U}}\,, \end{aligned} \tag{4.10}$$

which yields an unstable enclosure of the reachable set $X(t)$ of (4.8) *as*

$$Y(t) = \exp(t)\,Y(0)\,. \tag{4.11}$$

This instability is entailed by the inability of interval vectors to represent rotated boxes exactly, the so-called wrapping effect. ◇

Remark 4.1. *In the special case of linear ODE systems, it should be noted that general transformations can be applied in order to avoid the foregoing bound instability. In [121], for instance, the original system is first transformed into complex Jordan normal form and an exponential enclosure technique is used based on complex interval analysis and Piccard iterations. One can also shift the trajectories of the original ODEs as given below in* (4.12)*, and then use differential inequalities to bound the shifted ODEs.*

4.2.2 Ellipsoidal Propagation Approach

In order to improve upon interval enclosures, an alternative approach involves propagating ellipsoidal enclosures of the reachable set $X(t)$. In the presence of uncertain parameters in the ODE right-hand side or initial condition, it is convenient to construct ellipsoidal enclosures after shifting the trajectories so that

$$\forall (t,p) \in [0,T] \times P, \quad x(t,p) - \hat{x}(t) - G(t)(p-\hat{p}) \in \mathcal{E}(Q(t)), \tag{4.12}$$

with $Q(t) \in \mathbb{S}^{n_x}_+$. Here, $\hat{p} \in \mathbb{R}^{n_p}$ can be any reference point, for instance $\hat{p} := \mathrm{mid}\,(P)$; the reference trajectory $\hat{x} : [0,T] \to \mathbb{R}^{n_x}$ is such that $\hat{x}(t) := x(t,\hat{p})$; and $G(t) \in \mathbb{R}^{n_x \times n_p}$ is the solution of the variational differential equation

$$\forall t \in [0,T], \quad \dot{G}(t) \;=\; A(t)G(t) + B(t) \quad \text{with} \quad G(0) \;=\; B_0,$$

where the following shorthand notation is used:

$$A(t) := \frac{\partial f}{\partial x}(t,\hat{x}(t),\hat{p}), \quad B(t) := \frac{\partial f}{\partial p}(t,\hat{x}(t),\hat{p}), \text{ and } \quad B_0 \;:=\; \frac{\partial x_0}{\partial p}(\hat{p}),$$

assuming that the right-hand side function f is continuously-differentiable in (x,p), and the initial value function x_0 continuously-differentiable with respect to p. With this transformation, an enclosure $Y(t)$ of the reachable set $X(t)$ of the original ODEs can then be recovered as:

$$Y(t) = \{\hat{x}(t)\} \oplus G(t)[P-\hat{p}] \oplus \mathcal{E}(Q(t)). \tag{4.13}$$

The following theorem provides sufficient conditions for time-varying ellipsoidal bounds $\mathcal{E}(Q(t))$ to satisfy (4.12). This construction requires a right-hand side nonlinearity bounder $\Omega(t,Q,P,\hat{p}) \in \mathbb{IR}^{n_x}$ such that

$$\begin{aligned} &\forall (\xi,\rho) \in \{\hat{x}(t)\} \oplus G(t)[P-\hat{p}] \oplus \mathcal{E}(Q) \times P, \\ &\quad f(t,\xi,\rho) - f(t,\hat{x}(t),\hat{p}) - A(t)(\xi - \hat{x}(t)) - B(t)(\rho - \hat{p}) \in \Omega(t,Q,P,\hat{p}), \end{aligned} \tag{4.14}$$

as well as an initial value nonlinearity bounder $\Omega_0(P,\hat{p}) \in \mathbb{IR}^{n_x}$ satisfying

$$\forall \rho \in P, \quad x_0(\rho) - \hat{x}(0) - B_0(\rho - \hat{p}) \;\in\; \Omega_0(P,\hat{p}). \tag{4.15}$$

By a slight abuse of notation, the more compact notation $\Omega_t(Q)$ and Ω_0 is used subsequently to denote the nonlinearity bounders as P and $\hat{p}$ are fixed.

Theorem 4.2. *Consider the initial value problem* (4.1)*, and assume that the right-hand side function f is jointly continuous in (t,x,p) and continuously-differentiable in (x,p) for $t \in [0,T]$ and the initial-value function x_0 is continuously-differentiable on P, with $P \subset \mathbb{R}^{n_p}$ compact. Suppose that nonlinearity bounders Ω_t and Ω_0 satisfying* (4.14) *and* (4.15)*, respectively, are available, with Ω_t locally Lipschitz-continuous in Q on $[0,T]$. Let $\kappa : [0,T] \to \mathbb{R}^{n_x}_{++}$ be a continuous function, and let $Q : [0,T] \to \mathbb{S}^{n_x}_{+}$ be a Lipschitz-continuous function satisfying*

$$\begin{aligned} \text{a.e. } t \in [0,T]\,, \quad \dot{Q}(t) \succeq\; & A(t)Q(t) + Q(t)A(t)^{\mathsf{T}} + \sum_{i=1}^{n_x} \kappa_i(t) Q(t) \\ & + \operatorname{diag}(\kappa(t))^{-1} \operatorname{diag}\operatorname{rad}(\Omega_t(Q(t)))^2\,, \end{aligned} \tag{4.16}$$

with $Q(0) \succeq \operatorname{diag}\operatorname{rad}(\Omega_0)^2$. Then, $X(t) \subseteq \{\hat{x}(t)\} \oplus G(t)[P-\hat{p}] \oplus \mathcal{E}(Q(t))$ for all $t \in [0,T]$.

A similar result in the presence of time-varying uncertainties was established in [56]. The state transformation approach used in Theorem 4.2 exploits the fact that the uncertainty is time invariant, thus avoiding over-conservatism. An alternative proof of this results will be obtained in Sect. 4.4.1 as a corollary of the generalized differential inequalities introduced in Theorem 4.3.

Similar to differential inequalities, a practical application of Theorem 4.2 is in propagating a continuous-time enclosure of the reachable set with equality in (4.16). We note that a positive semidefinite solution of (4.16) is guaranteed to exist (at least locally) when the initial value $Q(0)$ is itself positive semidefinite. A clear advantage of this approach over differential inequalities is that it can provide exact bounds for parametric linear ODEs; e.g., the actual reachable set of (4.10) in Example 4.1 can be computed. Nonetheless, the number of ODEs in the auxiliary bounding system now scales as $\mathbf{O}(()n_x^2 + n_x n_p)$, and therefore the size of the bounding system increases quadratically with the number of state variables in the original ODE system. Another aspect that must be taken into account is the differentiability class of the functions f and x_0 in (4.1), as the ellipsoidal method in Theorem 4.2 requires continuous differentiability in both x and p, whereas differential inequalities only require local Lipschitz-continuity in x. Nonetheless, we note that the ellipsoidal approach may still provide valid enclosures in the case that f or x_0 fail to be continuously-differentiable, for instance by choosing $A(t)$ in the Clarke subdifferential $\partial_c f(t,\hat{x}(t),\hat{p})$ [27]—the quadratic convergence property (see Sect. 4.5.2) would be lost however.

Example 4.2. *Consider the following parametric nonlinear ODE system of the Lotka-Volterra type*

$$\dot{x}_1(t,p) = p_1 x_1(t,p)\,[1-x_2(t,p)] \quad \textit{with} \quad x_1(0,p) = 1.2+p_2 \tag{4.17}$$

$$\dot{x}_2(t,p) = p_1 x_2(t,p)\,[x_1(t,p)-1] \quad \textit{with} \quad x_2(0,p) = 1.1+p_2^2\,, \tag{4.18}$$

where $p \in P := ([2.95,3.05],[-0.05,0.05])^{\mathrm{T}}$ *and* $t \in [0,T]$.

In order to propagate an ellipsoidal enclosure of the reachable set of (4.17,4.18), Theorem 4.2 is applied with equality in (4.16)*. Further details about the numerical solution procedure will be given later on in Sect. 4.6. The results at* $t = 1$ *are shown in the left plot of Figure 4.1. The actual reachable set (gray-shaded area) is contained in the reachable set enclosure (thick solid line), which is obtained as the Minkowsky sum of* $\{x(t,\hat{p})\} \oplus G(t)[P-\hat{p}]$ *(dashed line) and* $\mathcal{E}(Q(t))$ *(dotted line). The right plot of Figure 4.1 presents the projections onto* x_1 *of the time-varying reachable set (shaded area) and of the ellipsoidal enclosure (thick solid line). This enclosure is also to be compared to the best possible bounds obtained with differential inequalities (thick dotted line) as given in Theorem 4.1, which turn out to be much more conservative in this example.* ◇

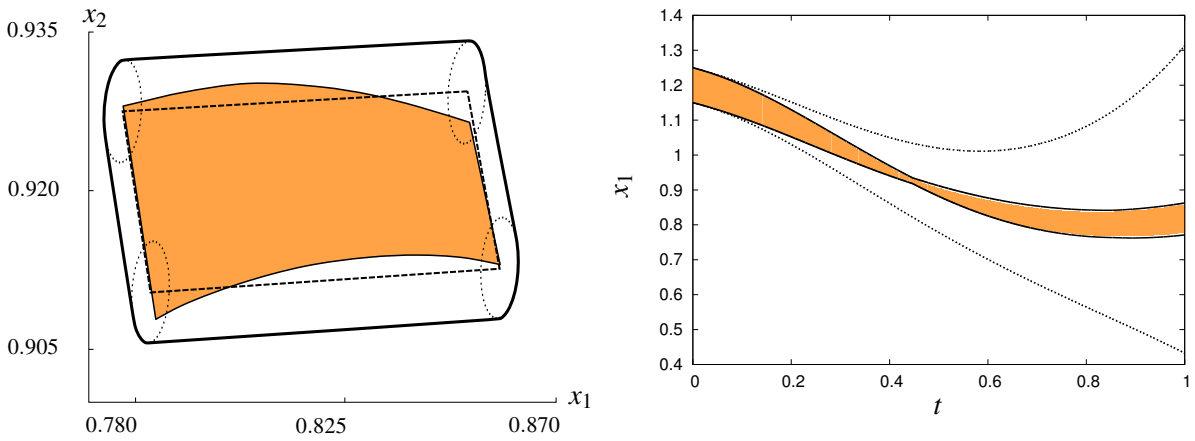

Fig. 4.1 *Left plot:* Ellipsoidal enclosure $\{x(t,\hat{p})\} \oplus G(t)[P-\hat{p}] \oplus \mathcal{E}(Q(t))$ of the reachable set $X(t)$ at $t = 1$. *Right plot:* Projections onto x_1 of the reachable set $X(t)$, the ellipsoidal enclosure $\{x(t,\hat{p})\} \oplus G(t)[P-\hat{p}] \oplus \mathcal{E}(Q(t))$, and the differential inequalities bounds $y^{\mathrm{L}}(t), y^{\mathrm{U}}(t)$ along $[0,T]$. The reachable set and the ellipsoidal enclosure are represented with a shaded area and a thick solid line, respectively.

4.3 Generalized Differential Inequalities

This section presents a unified framework for the continuous-time propagation of pointwise-in-time convex enclosures of the reachable set of nonlinear ODEs. The main result is a set of sufficient conditions, in the form of a generalized differential inequality, for a time-varying support function to describe an enclosure of the reachable set. In particular, this framework provides a direct link with standard differential inequalities and other existing time-continuous enclosure techniques that are amenable to numerical implementation.

The analysis that follows is inspired by, and closely related to, Aubin's viability theory [7]. The properties of the propagation operator Γ defined subsequently have been analyzed exhaustively in the context of viability theory, even in the more general case that p is a time-varying input function.

Definition 4.1. *The set propagation operator* $\Gamma : [0,T] \times [0,T] \times \Pi(\mathbb{R}^{n_x}) \to \Pi(\mathbb{R}^{n_x})$ *associated to the ODE* (4.1) *is defined as*

$$\Gamma(t_1,t_2,Y) := \left\{ x(t_2) \;\middle|\; \begin{array}{l} \exists p \in P : \\ \forall \tau \in [t_1,t_2],\ \dot{x}(\tau) = f(t,x(\tau),p) \\ x(t_1) \in Y \end{array} \right\}, \tag{4.19}$$

for all $t_1,t_2 \in [0,T]$ *with* $t_2 \geq t_1$*, and all* $Y \subseteq \mathbb{R}^{n_x}$*. In particular, for the reachable sets* $X(t_1)$ *and* $X(t_2)$ *at* $t_1 \leq t_2$*, we have*

$$X(t_2) \subseteq \Gamma(t_1,t_2,X(t_1)).$$

Our method of proof follows a stepwise procedure, starting by establishing the result under certain regularity and convexity assumptions, before relaxing these assumptions. In order to avoid disrupting the flow of the paper, a number of technical lemmata used in the proof are reported in Appendix B. The following proposition is also instrumental to prove the result.

Proposition 4.1. *Consider the ODE* (4.1)*, and assume that the right-hand side function* f *is globally Lipschitz-continuous in* x *uniformly on* $[0,T] \times P$*. Let* $Y_\varepsilon : [0,T] \to \mathbb{K}_{\mathrm{C}}^{n_x}$ *be a family of set-valued functions, and suppose that there exist two continuous functions* $\alpha,\beta : \mathbb{R}_+ \to \mathbb{R}_+$*, with* $\alpha(0) = \beta(0) = 0$*, such that*

$$\Gamma(t,t+h,Y_\varepsilon(t)) \subseteq Y_\varepsilon(t+h) \oplus h\,[\alpha(h) + \beta(\varepsilon)]\,\mathscr{B}^{n_x} \tag{4.20}$$

for all sufficiently small $h \geq 0$ with $t+h \in [0,T]$, all $t \in [0,T]$, and all $\varepsilon \geq 0$. Then,

$$\forall \varepsilon \geq 0\,, \quad \Gamma(0,T,Y_\varepsilon(0)) \subseteq Y_\varepsilon(T) \oplus \gamma(\varepsilon)\mathscr{B}^{n_x}\,,$$

for some continuous function $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ with $\gamma(0) = 0$.

Proof. The function f being globally Lipschitz-continuous in x uniformly on $[0,T] \times P$, (unique) solutions $x(t,p)$ of (4.1) are guaranteed to exist on $[0,T]$ for all $p \in P$. Therefore, the propagation operator Γ is well defined. Moreover, by uniform Lipschitzness of f in x on $[0,T] \times P$, it follows that the set-valued map Γ is itself locally Lipschitz-continuous around every $Y \subset \mathbb{R}^{n_x}$ (see [7], Definition 2.1.3); that is, for all $t, t' \in [0,T]$, with $t \leq t'$, there exists $C < \infty$ such that

$$\Gamma(t,t',Y+\Delta) \subseteq \Gamma(t,t',Y) + C\,\mathrm{diam}\,(\Delta)\,\mathscr{B}^{n_x}\,, \tag{4.21}$$

for all $\Delta \subset \mathbb{R}^{n_x}$ with sufficiently small $\mathrm{diam}\,(\Delta)$.

Consider a subdivision of $[0,T]$ into N equidistant intervals $[\theta_{i-1}, \theta_i]$, $i = 1,\ldots,N$, with end-points $\theta_i := ih$ and width $h := \frac{T}{N}$. From (4.20) and for a large enough N—or, equivalently, for a sufficiently small h—we have

$$\forall i \in \{1,\ldots,N\}\,, \quad \Gamma(\theta_{i-1},\theta_i,Y_\varepsilon(\theta_{i-1})) \subseteq Y_\varepsilon(\theta_i) \oplus h\,[\alpha(h)+\beta(\varepsilon)]\,\mathscr{B}^{n_x}\,. \tag{4.22}$$

Consider the sequence of sets X_i, $i = 0,\ldots,N$, defined recursively as

$$X_0 := Y_\varepsilon(0)\,, \quad \text{and} \quad X_i := \Gamma(\theta_{i-1},\theta_i,X_{i-1}) \quad \text{for } i \in \{1,\ldots,N\}\,. \tag{4.23}$$

If readily follows from Definition 4.1 and (4.23) that

$$\forall i \in \{1,\ldots,N\}\,, \quad \Gamma(0,\theta_i,Y_\varepsilon(0)) \subseteq X_i\,. \tag{4.24}$$

Next, by finite induction on $i = 0,\ldots,N-1$, assume that

$$X_i \subseteq Y_\varepsilon(\theta_i) \oplus C_i\,ih\,[\alpha(h)+\beta(\varepsilon)]\,\mathscr{B}^{n_x}\,.$$

for some constant $C_i < \infty$. We have

$$X_{i+1} = \Gamma(\theta_i,\theta_{i+1},X_i) \subseteq \Gamma(\theta_i,\theta_{i+1},Y_\varepsilon(\theta_i) \oplus C_i\,ih\,[\alpha(h)+\beta(\varepsilon)]\,\mathscr{B}^{n_x})\,,$$

and by the Lipschitz-continuity property (4.21), there are constants $C_i', C_{i+1} < \infty$ such that

$$\begin{aligned} X_{i+1} &\subseteq \Gamma(\theta_i, \theta_{i+1}, Y_\varepsilon(\theta_i)) \oplus C_i' \, i\, h\, [\alpha(h) + \beta(\varepsilon)] \mathscr{B}^{n_x} \\ &\overset{(4.20)}{\subseteq} Y_\varepsilon(\theta_{i+1}) \oplus C_{i+1} \, (i+1)\, h\, [\alpha(h) + \beta(\varepsilon)] \mathscr{B}^{n_x}, \end{aligned}$$

for sufficiently small h. Noting that $\Gamma(0, 0, Y_\varepsilon(0)) = X_0$, this induction gives

$$\Gamma(0, T, Y_\varepsilon(0)) \overset{(4.24)}{\subseteq} X_N \subseteq Y_\varepsilon(T) \oplus \gamma(\varepsilon) \mathscr{B}^{n_x} \oplus C_N T \, \alpha(h) \mathscr{B}^{n_x},$$

for sufficiently small h, where we define $\gamma(\varepsilon) := C_N T \, \beta(\varepsilon)$. Here, The sequence of the constants C_i remains bounded because the function f is globally Lipschitz—this is in analogy with Gronwall's lemma. The result of the proposition follows by a compactness argument, noting that $\lim_{h \to 0} C_N T \, \alpha(h) \mathscr{B}^{n_x} = \{0\}$ and $Y_\varepsilon(T) \oplus \gamma(\varepsilon) \mathscr{B}^{n_x}$ is compact by assumption. □

We are now in a position to state the main result of this section. The following theorem exploits the unique and exact representation of convex sets in terms of their support functions in order to obtain sufficient conditions for a time-varying support function of a convex enclosure of the reachable set. As already noted, this analysis is closely related to Aubin's viability theory [7], where the mathematical properties of convex enclosures are analyzed in detail. In this context, a key contribution of this theorem is a mechanism for constructing the support function of a convex enclosure of the reachable set, which can then be translated into practical algorithms as shown subsequently in Sect. 4.4.

Theorem 4.3. *Consider the initial value problem* (4.1) *for $p \in P$ with $P \subset \mathbb{R}^{n_p}$ compact. Assume that the right-hand side function f is jointly continuous in (t, x, p) and locally Lipschitz-continuous in x uniformly on $[0, T] \times P$, with $P \subset \mathbb{R}^{n_p}$ compact, and that the initial value function x_0 is continuous on P. Let $Y : [0, T] \to \mathbb{K}_{\mathrm{C}}^{n_x}$ be a set-valued function, with $V[Y(\cdot)](c)$ Lipschitz-continuous for all $c \in \mathbb{R}^{n_x}$, and suppose that*

$$\text{a.e. } t \in [0, T], \quad \dot{V}[Y(t)](c) \geq \max_{\xi, \rho} \left\{ c^\top f(t, \xi, \rho) \;\middle|\; \begin{array}{l} \xi \in Y(t) \\ c^\top \xi = V[Y(t)](c) \\ \rho \in P \end{array} \right\} \tag{4.25}$$

$$\text{and} \quad V[Y(0)](c) \geq \max_{\rho} \left\{ c^\top x_0(\rho) \;\middle|\; \rho \in P \right\} \tag{4.26}$$

for all $c \in \mathbb{R}^{n_x}$. Then, $X(t) \subseteq Y(t)$ for all $t \in [0, T]$.

Proof. The proof of the theorem proceeds in three steps. In the first step (S1), we establish the result under the following additional assumptions:

A1 The set-valued function Y is such that $Y(t) = \{x \mid g(t,x) \leq 0\}$, where the function g is strictly smooth (infinitely-often differentiable) and convex in x for each $t \in [0,T]$;

A2 The function $V[Y(\cdot)](c)$ is differentiable for all $c \in \mathbb{R}^{n_x}$ on $[0,T]$;

A3 The right-hand side function f is globally Lipschitz-continuous in x uniformly on $[0,T] \times P$.

In a second step (S2), we show that the result also holds without imposing Assumption A1, but with Assumptions A2 and A3 still holding. Finally, the last step (S3) shows that the result remains valid without the need for Assumptions A2 or A3 to hold.

At this point, we also note that the conditions (4.25) and (4.26) are invariant with respect to scaling of the directions $c \in \mathbb{R}^{n_x}$. Therefore, it is sufficient to prove that the generalized differential inequality (4.25,4.26) is satisfied for all directions c with $c^\top c = 1$.

S1 Suppose that Assumptions A1, A2, and A3 are satisfied. Since (4.26) holds for all $c \in \mathbb{R}^{n_x}$, we have $X(0) \subseteq Y(0)$. If we can prove that, for all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$, there exist two continuous functions $\alpha, \beta : \mathbb{R}_+ \to \mathbb{R}_+$, with $\alpha(0) = \beta(0) = 0$, such that

$$\forall t \in [0,T)\,, \quad V[\Gamma(t,t+h,Y(t))](c) \leq V[Y(t+h)](c) + h\,[\alpha(h) + \beta(\varepsilon)] \tag{4.27}$$

for all sufficiently small $h \geq 0$, and all sufficiently small $\varepsilon > 0$, then it will directly follow from Definition 4.1 and Proposition 4.1—after passing to the limit as $\varepsilon \to 0$—that

$$\forall t \in [0,T]\,, \quad X(t) \subseteq \Gamma(0,t,Y(0)) \subseteq Y(t)\,.$$

Let $\chi(\cdot,t,x_t,p)$ denote the solution of the initial value problem

$$\forall \tau \in [t,T], \quad \dot{\chi}(\tau,t,x_t,p) := f(\tau,\chi(\tau,t,x_t,p),p) \quad \text{with} \quad \chi(t,t,x_t,p) = x_t\,.$$

Solutions to this problem are guaranteed to exist for all $p \in P$ and all $x_t \in \mathbb{R}^{n_x}$ by Assumption A3, and we have

$$V[\Gamma(t,t+h,Y(t))](c) \overset{(4.4)}{=} \max_{\xi} \left\{ c^\top \xi \,\middle|\, \xi \in \Gamma(t,t+h,Y(t)) \right\}$$

$$\overset{(4.19)}{=} \max_{\xi,\rho} \left\{ c^\top \chi(t+h,t,\xi,\rho) \,\middle|\, \begin{array}{l} \xi \in Y(t) \\ \rho \in P \end{array} \right\}.$$
$$\leq \max_{\xi,\rho} \left\{ c^\top [\xi + h f(t,\xi,\rho)] \,\middle|\, \begin{array}{l} \xi \in Y(t) \\ \rho \in P \end{array} \right\} + h\,\alpha_1(h)\,,$$

for all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$, and for some continuous function $\alpha_1 : \mathbb{R}_+ \to \mathbb{R}_+$, with $\alpha_1(0) = 0$. Here, the last inequality follows from the fact that $\dot{\chi}(\cdot,t,\xi,\rho)$ exists and is jointly continuous in all the variables.

Now, by Lemma B.1, approximation functions $f_\varepsilon : [0,T] \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$ of f can be constructed that are jointly continuous in (t,x,p), smooth in x on $[0,T] \times P$, and such that

$$\forall x \in \mathbb{R}^{n_x}, \forall t \in [0,T], \forall \rho \in P, \quad \| f_\varepsilon(t,x,\rho) - f(t,x,\rho) \| \leq \beta(\varepsilon)\,, \tag{4.28}$$

for some continuous function $\beta : \mathbb{R}_+ \to \mathbb{R}_+$, with $\beta(0) = 0$. Then, by Assumption A1, we have

$$V[\Gamma(t,t+h,Y(t))](c) \leq \max_{\rho \in P} \max_{\xi} \left\{ c^\top \xi + h c^\top f_\varepsilon(t,\xi,\rho) \,\middle|\, g(t,\xi) \leq 0 \right\}$$
$$+ h\,[\alpha_1(h) + \beta(\varepsilon)]\,, \tag{4.29}$$

for all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$. We now have a closer look at the inner maximization problem

$$\max_{\xi} \left\{ c^\top \xi + h c^\top f_\varepsilon(t,\xi,\rho) \,\middle|\, g(t,\xi) \leq 0 \right\}. \tag{4.30}$$

Since $g(t,\cdot)$ is strictly convex (Assumption A1), the maximizer $\xi_t^*(c) := \arg\max_\xi \left\{ c^\top \xi \,\middle|\, g(t,\xi) \leq 0 \right\}$ is unique and well-defined for all $c \in \mathbb{R}^{n_x} \setminus \{0\}$. Moreover, the following properties hold at $\xi_t^*(c)$ [38]: (i) the constraint $g(t,\xi) \leq 0$ is strongly active, i.e., its KKT multiplier ν_t^* is nonzero; (ii) the linear independence constraint qualification (LICQ) holds since $\frac{\partial g}{\partial \xi}(t,\xi_t^*(c)) = \frac{1}{\nu_t^*} c$; and, (iii) the strong second-order sufficiency conditions (SSOSC) hold. It follows from NLP sensitivity theory [38] that the inner-maximization problem in (4.30) itself has a unique maximizer $\hat{\xi}_t^*(h,\rho,c)$ for all $(c,\rho) \in \mathbb{R}^{n_x} \times P$ and all sufficiently small $h \geq 0$. Moreover, $\hat{\xi}_t^*(\cdot,\rho,c)$ is differentiable and g is strongly active in that neighborhood of $h = 0$.

Therefore, we have

$$0 \;=\; g(t,\hat{\xi}_t^*(h,\rho,c)) \;=\; g(t,\xi_t^*(c)) + h\,\frac{\partial g}{\partial \xi}(t,\xi_t^*(c))\,\frac{\partial \hat{\xi}_t^*}{\partial h}(0,\rho,c) + \mathbf{O}(()\,h^2)\,,$$

so that

$$\frac{\partial g}{\partial \xi}(t,\xi_t^*(c))\,\frac{\partial \hat{\xi}_t^*}{\partial h}(0,\rho,c) = \mathbf{O}(()\,h)\,. \tag{4.31}$$

From the stationarity condition at $\xi_t^*(c)$, we also get

$$c - \nu_t^*\,\frac{\partial g}{\partial \xi}(t,\xi_t^*(c)) \;=\; 0\,,$$

which, when multiplied with $\frac{\partial \hat{\xi}_t^*}{\partial h}(0,\rho,c)$ from the right and using (4.31), gives

$$c^\top \frac{\partial \hat{\xi}_t^*}{\partial h}(0,\rho,c) \;=\; \mathbf{O}(()\,h)\,, \tag{4.32}$$

for all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$. It follows that

$$\begin{aligned}
&\max_{\xi}\,\left\{\, c^\top \xi + h c^\top f_\varepsilon(t,\xi,\rho) \;\middle|\; g(t,\xi) \le 0 \,\right\} \\
&\qquad =\; c^\top \hat{\xi}_t^*(h,\rho,c) + h c^\top f_\varepsilon(t,\hat{\xi}_t^*(h,\rho,c),p) \\
&\qquad =\; c^\top \left[\, \xi_t^*(c) + h\frac{\partial \hat{\xi}_t^*}{\partial h}(0,\rho,c) \,\right] + h c^\top f_\varepsilon(t,\xi_t^*(c),\rho) + \mathbf{O}(()\,h^2) \\
&\qquad \overset{(4.32)}{=}\; c^\top \xi_t^*(c) + h c^\top f_\varepsilon(t,\xi_t^*(c),\rho) + h\,\alpha_2(h)\,,
\end{aligned}$$

for all sufficiently small $h \ge 0$, all ρ in P, all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$, and for some continuous function $\alpha_2 : \mathbb{R}_+ \to \mathbb{R}_+$ with $\alpha_2(0) = 0$. Observing that $c^\top \xi_t^*(c) = V[Y(t)](c)$, we then obtain

$$\begin{aligned}
&\max_{\rho\in P}\,\max_{\xi}\,\left\{\, c^\top \xi + h c^\top f_\varepsilon(t,\xi,\rho) \;\middle|\; g(t,\xi) \le 0 \,\right\} \\
&\;=\; V[Y(t)](c) + h \max_{\rho\in P}\, c^\top f_\varepsilon(t,\xi_t^*(c),\rho) + h\,\alpha_2(h) \\
&\;\le\; V[Y(t)](c) + h \max_{\xi,\rho}\left\{\, c^\top f_\varepsilon(t,\xi,\rho) \;\middle|\; \begin{array}{l} \xi \in Y(t) \\ c^\top \xi = V[Y(t)](c) \\ \rho \in P \end{array} \,\right\} + h\,\alpha_2(h)
\end{aligned}$$

$$\begin{aligned}
&\overset{(4.28)}{\leq} V[Y(t)](c) + h \max_{\xi,\rho} \left\{ c^\top f(t,\xi,\rho) \;\middle|\; \begin{array}{l} \xi \in Y(t) \\ c^\top \xi = V[Y(t)](c) \\ \rho \in P \end{array} \right\} + h\,[\alpha_2(h) + \beta(\varepsilon)] \\
&\overset{(4.25)}{\leq} V[Y(t)](c) + h\dot{V}[Y(t)](c) + h\,[\mathbf{O}(()h) + \alpha_2(\varepsilon)] \\
&\overset{(A2)}{\leq} V[Y(t+h)](c) + h\,[\alpha_2(h) + \alpha_3(h) + \beta(\varepsilon)]\ ,
\end{aligned}$$

for all sufficiently small $h \geq 0$, and all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$. In the last equality, we have used that $V[Y(\cdot)](c)$ is differentiable on $[0,T]$ for all c by Assumption A2. The condition (4.27) thus follows from (4.29).

S2 Suppose now that Assumptions A2 and A3 still hold, but not Assumption A1. Following Lemma B.2 (Appendix B), we construct a family of set-valued functions $Y_\varepsilon : [0,T] \to \mathbb{K}_C^{n_x}$, parameterized by $\varepsilon \geq 0$, such that $Y_\varepsilon(t) := \{x \in \mathbb{R}^{n_x} \mid g_\varepsilon(t,x) \leq 0\} \supseteq Y(t)$, with $g_\varepsilon(t,\cdot)$ strictly convex and smooth for all $\varepsilon > 0$ and all $t \in [0,T]$, and

$$\dot{V}[Y_\varepsilon(t)](c) \geq \dot{V}[Y(t)](c) + L\alpha(\varepsilon)$$

for all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$ and some $L < \frac{1}{T}$. Suppose for a moment that T is sufficiently small in order for the chosen L to remain larger than the uniform Lipschitz constant of the right-hand side function f on the compact sets $\bigcup_{t\in[0,T]} Y_\varepsilon(t)$. This way, we have

$$\begin{aligned}
\dot{V}[Y_\varepsilon(t)](c) &\overset{(4.25)}{\geq} \max_{\xi,\rho} \left\{ c^\top f(t,\xi,\rho) \;\middle|\; \begin{array}{l} \xi \in Y(t) \\ c^\top \xi = V[Y(t)](c) \\ \rho \in P \end{array} \right\} + L\alpha(\varepsilon) \\
&\geq \max_{\xi,\rho} \left\{ c^\top f(t,\xi,\rho) \;\middle|\; \begin{array}{l} \xi \in Y_\varepsilon(t) \\ c^\top \xi = V[Y_\varepsilon(t)](c) \\ \rho \in P \end{array} \right\}
\end{aligned}$$

for all $t \in [0,T]$ and all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$. Then, we can apply the result from part S1 above to show that $X(t) \subseteq Y_\varepsilon(t)$ for all $t \in [0,T]$ and all $\varepsilon > 0$, and the result of the theorem follows by noting that the sets $Y_\varepsilon(t)$ converge to the compact sets $Y(t)$ in the Haussdorf metric as $\varepsilon \to 0$. Finally, the auxiliary assumption that T is sufficiently small—but strictly larger than 0—can be made without loss of generality, since we

can otherwise divide $[0,T]$ into a finite number of sufficiently small subintervals onto which the foregoing procedure can be applied.

S3 In order to complete the proof of the theorem, we finally relax Assumptions A2 and A3. Assume first that the right-hand side function f is locally Lipschitz at each $x \in D$ uniformly on $[0,T] \times P$, where $D \in \mathcal{K}^{n_x}$ is such that $D \supseteq \cup_{t\in[0,T]} Y(t)$—Note that such a set D always exists since the sets $Y(t)$ are bounded by assumption. A modified function $\tilde{f}$ that is globally Lipschitz-continuous in x uniformly on $[0,T] \times P$ can then be constructed, which coincides with f on D. In particular, using $\tilde{f}$ instead of f does not modify the generalized differential inequality (4.25), showing that the assumption that f is globally Lipschitz-continuous in x can be replaced by a local Lipschitzness condition without loss of generality. Finally, because any (locally) Lipschitz-continuous function is differentiable almost everywhere and since perturbing a differential inequality on a set with Lebesgue measure equal to zero does not affect the result, one can drop Assumption A2 as well. □

Remark 4.2. *Theorem 4.3 remains true if the time-invariant parameter $p \in P \subset \mathbb{R}^{n_p}$ is replaced with time-varying measurable function $p : [0,T] \to \mathbb{R}^{n_p}$ such that $p(t) \in P$. This extension can for example be established by dividing the integration horizon $[0,T]$ into small subintervals of width $h := \frac{T}{N}$ for $N \in \mathbb{N}$, where the uncertain parameter is kept constant and Theorem 4.3 can thus be applied. Since the reachable set enclosures $Y(t)$ are compact, one can then consider the limit as $h \to 0$ to prove that the result is valid for general bounded functions, not merely piecewise constant functions.* ◇

4.4 Applications of Generalized Differential Inequalities

This section describes how the generalized differential inequality introduced in Theorem 4.3 can be specialized to yield alternative proofs for the differential inequalities and ellipsoidal propagation results recalled in Sect. 4.2. An approach based on Taylor models with convex remainder terms is also described in order to propagate nonconvex enclosures of the reachable set.

4.4.1 Link with Standard Differential Inequalities and Ellipsoidal Bounding Approach

The case of interval enclosures is addressed first, by specializing the convex enclosure as $Y(t) := \left[y^{\mathrm{L}}(t), y^{\mathrm{U}}(t)\right]$. The following proposition establishes that the result of the standard differential inequalities in Theorem 4.1 is implied by the generalized differential inequalities in Theorem 4.3.

Proposition 4.2. *Any pair of functions $y^{\mathrm{L}}(t), y^{\mathrm{U}}(t) : [0,T] \to \mathbb{R}^{n_x}$ satisfying Theorem 4.1 also satisfies Theorem 4.3 with $Y(t) := \left[y^{\mathrm{L}}(t), y^{\mathrm{U}}(t)\right]$.*

Proof. Noting that

$$\sum_{i=1}^{n_x} \left\{ \begin{array}{ll} c_i y_i^{\mathrm{L}}(t) & \text{if } c_i \leq 0 \\ c_i y_i^{\mathrm{U}}(t) & \text{otherwise} \end{array} \right\} = \frac{1}{2} c^{\mathsf{T}} \left(y^{\mathrm{L}}(t) + y^{\mathrm{U}}(t)\right) + \frac{1}{2} \mathrm{abs}(c)^{\mathsf{T}} \left(y^{\mathrm{U}}(t) - y^{\mathrm{L}}(t)\right) , \quad (4.33)$$

with $\mathrm{abs}(c) := (|c_1|, \ldots, |c_{n_x}|)^{\mathsf{T}}$, the support function of the interval vector $Y(t)$ is given by

$$\forall c \in \mathbb{R}^{n_x}, \quad V\left[Y(t)\right](c) = \frac{1}{2} c^{\mathsf{T}} \left(y^{\mathrm{L}}(t) + y^{\mathrm{U}}(t)\right) + \frac{1}{2} \mathrm{abs}(c)^{\mathsf{T}} \left(y^{\mathrm{U}}(t) - y^{\mathrm{L}}(t)\right) . \quad (4.34)$$

The initial-value condition (4.26) is trivially satisfied with this definition. Moreover, rewriting the (right-hand) derivative of the support function in terms of the (right-hand) derivatives of the bounding trajectories $y^{\mathrm{L}}(t), y^{\mathrm{U}}(t)$ and using the standard differential inequalities result in Theorem 4.1 gives

$$\begin{aligned}
\dot{V}[Y(t)](c) &\overset{(4.34)}{=} \frac{1}{2} (c - \mathrm{abs}(c))^{\mathsf{T}} \dot{y}^{\mathrm{L}}(t) + \frac{1}{2} (c + \mathrm{abs}(c))^{\mathsf{T}} \dot{y}^{\mathrm{U}}(t) \\
&\overset{(4.6),(4.7)}{\geq} \sum_{i=1}^{n_x} \left[\frac{1}{2} \min_{\xi,\rho} \left\{ (c_i - |c_i|) f_i(t,\xi,\rho) \;\middle|\; \begin{array}{l} \xi \in \left[y^{\mathrm{L}}(t), y^{\mathrm{U}}(t)\right] , \\ \xi_i = y_i^{\mathrm{L}}(t) \, , \; \rho \in P \end{array} \right\} \right. \\
&\qquad \left. + \frac{1}{2} \max_{\xi,\rho} \left\{ (c_i + |c_i|) f_i(t,\xi,\rho) \;\middle|\; \begin{array}{l} \xi \in \left[y^{\mathrm{L}}(t), y^{\mathrm{U}}(t)\right] , \\ \xi_i = y_i^{\mathrm{U}}(t) \, , \; \rho \in P \end{array} \right\} \right] \\
&= \sum_{i=1}^{n_x} \max_{\xi,\rho} \left\{ c_i f_i(t,\xi,\rho) \;\middle|\; \begin{array}{l} \xi \in Y(t) \, , \; \rho \in P \, , \\ \xi_i = \left\{ \begin{array}{ll} y_i^{\mathrm{L}}(t) & \text{if } c_i \leq 0 \\ y_i^{\mathrm{U}}(t) & \text{otherwise} \end{array} \right. \end{array} \right\} \\
&\geq \max_{\xi,\rho} \left\{ c^{\mathsf{T}} f(t,\xi,\rho) \;\middle|\; \begin{array}{l} \xi \in Y(t) \, , \; \rho \in P \, , \\ \sum_{i=1}^{n_x} c_i \xi_i = \sum_{i=1}^{n_x} \left\{ \begin{array}{ll} c_i y_i^{\mathrm{L}}(t) & \text{if } c_i \leq 0 \\ c_i y_i^{\mathrm{U}}(t) & \text{otherwise} \end{array} \right. \end{array} \right\}
\end{aligned}$$

$$\overset{(4.33)}{=} \max_{\xi,\rho} \left\{ c^\top f(t,\xi,\rho) \,\middle|\, \begin{array}{l} \xi \in Y(t)\,,\ \rho \in P\,, \\ c^\top \xi = \frac{1}{2} c^\top (y^{\mathrm{L}}(t) + y^{\mathrm{U}}(t)) \\ \qquad\quad + \frac{1}{2} \operatorname{abs}(c)^\top (y^{\mathrm{U}}(t) - y^{\mathrm{L}}(t)) \end{array} \right\}$$

$$\overset{(4.34)}{=} \max_{\xi,\rho} \left\{ c^\top f(t,\xi,\rho) \,\middle|\, \xi \in Y(t)\,,\ c^\top \xi = V[Y(t)](c)\,,\ \rho \in P \right\}\,. \qquad \square$$

The case of ellipsoidal enclosures is addressed by specializing the convex enclosure as $Y(t) := \mathcal{E}(Q(t))$. Like previously with standard differential inequalities, the result of Theorem 4.2 is implied by the generalized differential inequalities in Theorem 4.3.

Proposition 4.3. *Any functions $\kappa : [0,T] \to \mathbb{R}_{++}$ and $Q : [0,T] \to \mathbb{S}^{n_x}_+$ satisfying Theorem 4.2, with associated nonlinearity bounders Ω_0 and Ω_t, also satisfy Theorem 4.3 with $Y(t) := \{\hat{x}(t)\} \oplus G(t)[P - \hat{p}] \oplus \mathcal{E}(Q(t))$.*

Proof. Noting that, for all $c \in \mathbb{R}^{n_x}$, $V[\mathcal{E}(Q(t))](c) = \sqrt{c^\top Q(t) c}$, we have

$$\dot{V}[\mathcal{E}(Q(t))](c) \;=\; \frac{1}{2\sqrt{c^\top Q(t) c}} c^\top \dot{Q}(t) c\,. \tag{4.35}$$

Assume, for a moment, that $Q(t)$ is positive definite and $c \neq 0$, so that $\dot{V}[\mathcal{E}(Q(t))](c)$ exists almost everywhere. Substituting (4.16) in (4.35), and using the fact that $\sum_{i=1}^{n_x} \kappa_i(t) Q(t) + \operatorname{diag}(\kappa(t))^{-1} \operatorname{diag} \operatorname{rad}(\Omega_t(Q(t)))^2$ is minimized by choosing

$$\kappa(t) = \frac{c^\top \operatorname{diag} \operatorname{rad}(\Omega_t(Q(t)))}{\sqrt{c^\top Q(t) c}}\,,$$

we have that

$$\dot{V}[\mathcal{E}(Q(t))](c) \geq \frac{c^\top A(t) Q(t) c}{\sqrt{c^\top Q(t) c}} + \left\| c^\top \operatorname{diag} \operatorname{rad}(\Omega_t(Q(t))) \right\|_1\,, \tag{4.36}$$

for all $\kappa(t) > 0$.

The shifted state trajectories $z(t,p) := x(t,p) - \hat{x}(t) - G(t)[p - \hat{p}]$ satisfy the ODE

$$\begin{aligned} \dot{z}(t,p) \;&= f(t, z(t,p) + \hat{x}(t) + G(t)(p - \hat{p}), p) - f(t, \hat{x}(t), \hat{p}) - [A(t)G(t) + B(t)](p - \hat{p}) \\ &=: \phi(t, z(t,p), p) \qquad (4.37) \\ \text{with} \quad z(0,p) \;&=\; x_0(p) - \hat{x}(0) - G(0)(p - \hat{p}) =: z_0(p)\,. \end{aligned}$$

By construction of the non-linearity bounder Ω_t in (4.14) at a given $t \in [0,T]$, we thus have

$$\phi(t,\xi,\rho) - A(t)\xi \in \Omega_t(Q(t))\,,$$

and

$$\left\| c^\mathsf{T} \operatorname{diag} \operatorname{rad}\left(\Omega_t(Q(t))\right) \right\|_1 \;=\; \max_{\omega \in \Omega_t(Q(t))} c^\mathsf{T}\omega \;\geq\; c^\mathsf{T}\left[\phi(t,\xi,\rho) - A(t)\xi\right]\,,$$

for all $(\xi,\rho) \in \mathcal{E}(Q(t)) \times P$ and all $c \in \mathbb{R}^{n_x}$. In particular, choosing $\xi := \frac{Q(t)c}{\sqrt{c^\mathsf{T}Q(t)c}} \in \mathcal{E}(Q(t))$ gives

$$\left\| c^\mathsf{T} \operatorname{diag} \operatorname{rad}\left(\Omega_t(Q(t))\right) \right\|_1 \;\geq\; \max_{\rho \in P} c^\mathsf{T} \left[\phi\left(t, \frac{Q(t)c}{\sqrt{c^\mathsf{T}Q(t)c}}, \rho\right) - A(t)\frac{Q(t)c}{\sqrt{c^\mathsf{T}Q(t)c}} \right]\,,$$

which after substitution into (4.36) yields

$$\begin{aligned} \forall c \in \mathbb{R}^{n_x}\,, \quad \dot{V}[\mathcal{E}(Q(t))](c) \;&\geq\; \max_{\rho \in P} c^\mathsf{T} \phi\left(t, \frac{Q(t)c}{\sqrt{c^\mathsf{T}Q(t)c}}, \rho\right) \\ &=\; \max_{\rho,\xi} \left\{ c^\mathsf{T}\phi(t,\xi,\rho) \;\middle|\; \begin{array}{l} \xi \in \mathcal{E}(Q(t)) \\ c^\mathsf{T}\xi = V[\mathcal{E}(Q(t))](c) \\ \rho \in P \end{array} \right\}. \end{aligned}$$

Likewise, by construction of the initial value non-linearity bounder Ω_0 in (4.15), we have

$$V[\mathcal{E}(Q(0))](c) = \sqrt{c^\mathsf{T}Q(0)c} \geq \left\| c^\mathsf{T} \operatorname{diag} \operatorname{rad}\left(\Omega_0\right) \right\|_1 \;=\; \max_{\omega \in \Omega_0} c^\mathsf{T}\omega \;\geq\; c^\mathsf{T} z_0(\rho)\,,$$

for all $\rho \in P$ and all $c \in \mathbb{R}^{n_x}$. Therefore, $V[\mathcal{E}(Q(t))](c)$ satisfies the generalized differential inequalities (4.25) and corresponding initial condition.

The case that $Q(t)$ is only positive semi-definite can be treated by approximating $Q(t)$ with a positive definite matrix $Q_\varepsilon \succ 0$ with $\|Q_\varepsilon(t) - Q(t)\| < \varepsilon$ for any sufficiently small $\varepsilon > 0$, then repeating the above construction with Q_ε instead of Q, and finally taking the limit $\varepsilon \to 0$. Notice that this argumentation is analogous to the strongly convex relaxation of the set $Y(t)$ with the set $Y_\varepsilon(t)$ that was used in the proof of Theorem 4.3. □

4.4.2 Propagation of Nonconvex Enclosures using Taylor Models

Convex enclosures of reachable sets can be conveniently computed using the unified framework of generalized differential inequalities, for instance as interval vectors or ellipsoids.

Nonetheless, approximating nonconvex enclosures with convex sets can result in large overestimation, due to both the wrapping effect and the dependency problem. Clearly, the ability to propagate nonconvex enclosures can help mitigate this overestimation.

This section describes a novel way of propagating nonconvex enclosures of the reachable set of the parametric ODE (4.1), whereby high-order polynomial approximations of the parametric ODE solutions are used in combination with the framework of generalized differential inequalities for bounding the approximation error. Specifically, let $\mathcal{P}_x^q : [0,T] \times P \to \mathbb{R}^{n_x}$ denote a qth-order polynomial approximant, and let the remainder function $r_x^q : [0,T] \times P \to \mathbb{R}^{n_x}$ be defined such that

$$\forall (t,p) \in [0,T] \times P\,, \quad r_x^q(t,p) := x(t,p) - \mathcal{P}_x^q(t,p)\,. \tag{4.38}$$

It follows from differentiating (4.38) with respect to time that the remainder function r_x^q satisfies the parametric ODE

$$\dot{r}_x^q(t,p) = f\left(t, \mathcal{P}_x^q(t,p) + r_x^q(t,p), p\right) - \dot{\mathcal{P}}_x^q(t,p) =: r_f^q\left(t, r_x^q(t,p), p\right)\,, \tag{4.39}$$

with initial condition $r_x^q(0,p) = x_0(p) - \mathcal{P}_{x_0}^q(p)$, where $\mathcal{P}_{x_0}^q : P \to \mathbb{R}^{n_x}$ denotes a qth-order polynomial approximation of x_0 on P. The idea is to apply Theorem 4.3 for characterizing pointwise-in-time convex enclosures $\mathcal{R}_x^q(t,P) \supseteq \{r(t,p) | p \in P\}$, and thus obtain a nonconvex enclosure of the reachable set in the form

$$X(t,P) \subseteq \{\mathcal{P}_x^q(t,p) \mid p \in P\} \oplus \mathcal{R}_x^q(t,P)\,.$$

The focus hereafter is on Taylor models [84, 103, 20], although alternative types of polynomial approximation can be used in principle as long as these constructions can be automated for general factorable functions. In the Taylor model approach, the polynomial approximant $\mathcal{P}_x^q(t,\cdot)$ matches the qth-order Taylor expansion of $x(t,\cdot)$ on P at a given reference point $\hat{p} \in P$ [25]:

$$\forall i\{1,\ldots,n_x\}\,,\ \forall t \in [0,T]\,,\ \forall p \in P\,, \quad \mathcal{P}_{x_i}^q(t,p) := \sum_{\substack{\gamma \in \mathbb{N}^{n_p},\\ |\gamma| \leq q}} \frac{\partial^\gamma x_i(t,\hat{p})}{\gamma!}(p-\hat{p})^\gamma\,. \tag{4.40}$$

That is, the time-varying coefficients of $\mathcal{P}_x^q$ are expressed as functions of the state-sensitivities $\partial^\kappa x_i(\cdot,\hat{p})$ of (4.1) up to order q at the reference points $\hat{p}$. In particular, this construction requires that the right-hand side function f and initial-value function x_0 be at

least $(q+1)$-times continuously-differentiable in both x and p. Moreover, it requires that a system of state-sensitivity equations of size $\mathbf{O}(()\, n_x n_p^q)$ be integrated on $[0,T]$.

The application of differential inequalities to propagate interval bounds on the remainder function r_x^q as $\mathcal{R}_x^q(t,P) := \left[r_{x_i}^{\mathrm{L}}(t), r_{x_i}^{\mathrm{U}}(t)\right]$ was first investigated in [25]. The following proposition follows directly on application of Theorem 4.1 to bound the solutions of (4.39).

Proposition 4.4. *Consider the initial value problem* (4.1)*, and assume that the right-hand side function f is jointly continuous in (t,x,p) and $(q+1)$-times continuously-differentiable in (x,p) on $[0,T]$, with $q \geq 1$, and the initial-value function x_0 is $(q+1)$-times continuously-differentiable on P, with $P \subset \mathbb{R}^{n_p}$ compact. Let $\mathcal{P}_x^q : [0,T] \times P \to \mathbb{R}^{n_x}$ be defined as in* (4.40)*, for a given reference point $\hat{p} \in P$. Let the functions $r_x^{q,\mathrm{L}}, r_x^{q,\mathrm{U}} : [0,T] \to \mathbb{R}^{n_x}$ be Lipschitz-continuous, with $r_x^{q,\mathrm{L}} \leq r_x^{q,\mathrm{U}}$, and satisfy*

$$\text{a.e. } t \in [0,T]\,, \quad \dot{r}_{x_i}^{q,\mathrm{L}}(t) \leq \min_{\xi,\rho} \left\{ f_i(t, \mathcal{P}_x^q(t,\rho) + \xi, \rho) - \dot{\mathcal{P}}_x^q(t,\rho) \,\middle|\, \begin{array}{l} \xi_i = r_{x_i}^{q,\mathrm{L}}(t) \\ \xi \in \mathcal{R}_x^q(t,P) \\ \rho \in P \end{array} \right\} \tag{4.41}$$

$$\dot{r}_{x_i}^{q,\mathrm{U}}(t) \geq \max_{\xi,\rho} \left\{ f_i(t, \mathcal{P}_x^q(t,\rho) + \xi, \rho) - \dot{\mathcal{P}}_x^q(t,\rho) \,\middle|\, \begin{array}{l} \xi_i = r_{x_i}^{q,\mathrm{U}}(t) \\ \xi \in \mathcal{R}_x^q(t,P) \\ \rho \in P \end{array} \right\} \tag{4.42}$$

for each $i = 1,\ldots,n_x$, and $\left[r_{x_i}^{\mathrm{L}}(0), r_{x_i}^{\mathrm{U}}(0)\right] \supseteq \{x_0(p) - \mathcal{P}_{x_0}^q \mid p \in P\}$, where $\mathcal{P}_{x_0}^q : \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$ is a truncated Taylor expansion of x_0 on P at $\hat{p}$. Then, $X(t) \subseteq \{\mathcal{P}_x^q(t,p) | p \in P\} \oplus [r_x^{q,\mathrm{L}}(t), r_x^{q,\mathrm{U}}(t)]$, for all $t \in [0,T]$.

In practice, the resulting enclosures are often found to be tighter than upon application of the standard differential inequalities given by Theorem 4.1, especially as the expansion order q increases. This trend is confirmed by the convergence analysis results in Sect. 4.5, where it is established that the overestimation is of order (no less than) $\mathbf{O}(()\operatorname{diam}(P)^{q+1})$ for Taylor models combined with differential inequalities, while it is normally of order $\mathbf{O}(()\operatorname{diam}(P))$ with standard differential inequalities. Nonetheless, it is noteworthy that this trend is only asymptotic as $\operatorname{diam}(P) \to 0$, and so standard differential inequalities can outperform their Taylor model counterparts for large parameter host sets. It is also worth noting that the combination with Taylor models entails a scaling in the size of the bounding system not only with n_x but also with n_p and q as $\mathbf{O}(()\, n_x n_p^q)$. A trade-off can therefore be expected between the enclosure tightness and the increase in computational time for larger expansion orders.

The propagation of ellipsoidal enclosures for the remainder function r_x^q as $\mathcal{R}_x^q(t,P) := \mathcal{E}(Q_x^q(t))$ was recently investigated in [152] as well. One way of applying the ellipsoidal bounding technique of Theorem 4.2 involves rewriting the ODE (4.39) in the form

$$\dot{r}_x^q(t,p) = A_f(t) r_x^q(t,p) + N_f\left(t, r_x^q(t,p), p, \hat{p}\right), \tag{4.43}$$
$$\text{with} \quad A_f(t) := \frac{\partial f}{\partial x}\left(t, \mathcal{P}_x^q(t,\hat{p}), \hat{p}\right) \quad \text{and} \quad N_f(t,r,p) := r_f^q(t,r,p) - A_f(t)\, r\,.$$

Then, assuming that nonlinearity bounders $\Omega_t^q(Q), \Omega_0^q \in \mathbb{IR}^{n_x}$ can be constructed at each $t \in [0,T]$ such that

$$\forall (r,p) \in \mathcal{E}(Q) \times P, \quad N_f(t,r,p) \in \Omega_t^q(Q) \tag{4.44}$$
$$\text{and} \quad \forall p \in P, \quad x_0(p) - \mathcal{P}_{x_0}^q(p) \in \Omega_0^q\,, \tag{4.45}$$

the following proposition follows directly on application of Theorem 4.2 to (4.43).

Proposition 4.5. *Consider the initial value problem* (4.1)*, and assume that the right-hand side function f is jointly continuous in (t,x,p) and $(q+1)$-times continuously-differentiable in (x,p) on $[0,T]$, with $q \geq 1$, and the initial-value function x_0 is $(q+1)$-times continuously-differentiable on P, with $P \subset \mathbb{R}^{n_p}$ compact. Let $\mathcal{P}_x^q : [0,T] \times P \to \mathbb{R}^{n_x}$ be defined as in* (4.40)*, for a given reference point $\hat{p} \in P$. Suppose that nonlinearity bounders $\Omega_t^q(Q)$ and Ω_0^q satisfying* (4.44) *and* (4.45)*, respectively, are available, with Ω_t^q locally Lipschitz-continuous in Q on $[0,T]$. Let $\kappa : [0,T] \to \mathbb{R}_{++}^{n_x}$ be a continuous function, and let $Q_x^q : [0,T] \to \mathbb{S}_+^{n_x}$ be a Lipschitz-continuous function satisfying*

$$\begin{aligned} \text{a.e. } t \in [0,T]\,, \quad \dot{Q}_x^q(t) \succeq A_f(t) Q_x^q(t) + Q_x^q(t) A_f(t)^{\mathsf{T}} + \sum_{i=1}^{n_x} \kappa_i(t) Q_x^q(t) \\ + \operatorname{diag}(\kappa(t))^{-1} \operatorname{diag}\operatorname{rad}\left(\Omega_t^q(Q_x^q(t))\right)^2\,, \end{aligned} \tag{4.46}$$

with $Q_x^q(0) \succeq \operatorname{diag}\operatorname{rad}\left(\Omega_0^q\right)$. Then, $X(t) \subseteq \{\mathcal{P}_x^q(t,p) | p \in P\} \oplus \mathcal{E}(Q_x^q(t))$ for all $t \in [0,T]$.

Like with differential inequalities, the propagation of an ellipsoidal enclosure of the remainder term in a higher-order Taylor model is often found to improve upon the standard ellipsoidal bounds given by Theorem 4.2. In particular, it can be shown that the enclosures obtained by considering a first-order Taylor model with ellipsoidal remainder or a standard ellipsoidal approach are identical when the nonlinearity bounders Ω_0, Ω_t in the latter are derived from first-order Taylor models; that is, $\{x(t,\hat{p})\} \oplus G(t)(p-\hat{p})$ and $\mathcal{P}_x^1(t,p)$ match

and so do $Q(t)$ and $Q_x^1(t)$. Here again, a main limitation of Taylor models with ellipsoidal remainders is the size of the bounding system, which scales as $\mathbf{O}(()\, n_x n_p^q + n_x^2)$.

Example 4.2 (continued)**.** *We revisit the Lotka-Volterra system (4.17,4.18) in the context of nonconvex enclosures based on Taylor models. Second-order Taylor models ($q = 2$) are considered, as expanded at the midpoint $\hat{p} = (3,0)^\mathsf{T}$ and with centered remainder bounds. Both differential inequalities-based bounds and ellipsoidal bounds are considered for the remainder function r_x^q and are computed via, respectively, (4.41,4.42) and* (4.46)*, both with equalities. Further details about the numerical solution procedure will be given later on in Sect. 4.6.*

The results at $t = 1$ are shown in Figure 4.2. The actual reachable set (shaded area) is contained in the reachable set enclosure (thick solid line), for Taylor models with either interval (left plot) or ellipsoidal (right plot) remainder bounds. The former is obtained as the Minkowsky sum of $\{\mathcal{P}_x^2(t,p)|p \in P\}$ (dashed line) and the remainder interval $[r_x^{2,\mathrm{L}}(t), r_x^{2,\mathrm{U}}(t)]$ (dotted lines); the latter, as the Minkowsky sum of $\{\mathcal{P}_x^2(t,p)|p \in P\}$ (dashed line) and the remainder enclosure $\mathcal{E}(Q_x^2(t))$ (dotted lines). In comparing Taylor models with different remainder bounds, it is evident that the wrapping effect and the dependency problem transfers to the enclosures of the remainder function, although the overestimation is greatly reduced in this example by the use of Taylor models. Also note that Taylor models with ellipsoidal remainders outperform their interval remainder counterparts. ◇

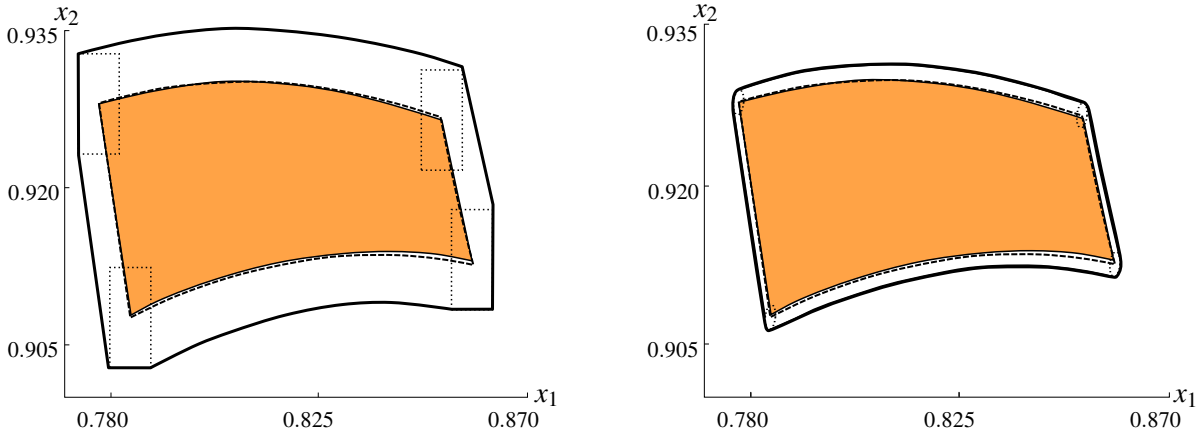

Fig. 4.2 *Left plot:* Second-order Taylor model with interval remainder $\{\mathcal{P}_x^2(t,p)|p \in P\} \oplus [r_x^{2,\mathrm{L}}(t), r_x^{2,\mathrm{U}}(t)]$ of the reachable set $X(t,P)$ at $t = 1$. *Right plot:* Second-order Taylor model with ellipsoidal remainder $\{\mathcal{P}_x^2(t,p)|p \in P\} \oplus \mathcal{E}(Q_x^2(t))$ of the reachable set $X(t,P)$ at $t = 1$. The reachable set and the ellipsoidal enclosure are represented with a shaded area and a thick solid line, respectively.

4.5 Convergence Analysis

We already argued in the introduction section that convergence properties of the computed enclosures to the actual reachable set are important in application areas such as global or robust optimization. This section investigates conditions under which a family of enclosures $Y(\cdot,P)$ that satisfy the generalized differential inequality in Theorem 4.3 exhibits a certain convergence order $k \geq 1$ with respect to the parameter host set P. This generic result is then specialized to interval enclosures, ellipsoidal enclosures, and Taylor models with convex remainders as a means for deriving sharp bounds on their convergence order. The results are illustrated with a numerical example.

4.5.1 General Convergence Theorem

This subsection investigates the convergence properties of enclosure functions that satisfy the generalized differential inequality (4.25,4.26). A formal definition of the convergence order of (the image of) a set-valued function is given below. We formulate this definition in terms of *representable sets*, which can be described exactly by a finite number of parameters. The class of representable sets of interest hereafter are interval vectors and ellipsoids.

Definition 4.2. *Let $\mathscr{S}(\mathbb{K}^n)$ be a class of representable sets in $\mathbb{K}^n$. A set-valued function $\Psi : \mathscr{S}(\mathbb{K}^n) \to \mathbb{K}^m$ is said to have convergence order $k \geq 1$ on a set $Z \in \mathcal{S}(\mathbb{K}^n)$, if there exists a constant $C < \infty$ such that*

$$\operatorname{diam}(\Psi(W)) \leq C \operatorname{diam}(W)^k \ ,$$

for all $W \in \mathscr{S}(\mathbb{K}^n)$ such that $W \subseteq Z$ with sufficiently small $\operatorname{diam}(W)$.

In practice, the maximization problem defining the right-hand side of (4.25) may not be solved exactly and the application of Theorem 4.3 typically relies on the availability of an upper-bounding function $\mathcal{B}_t : \mathbb{R}^{n_x} \times \mathbb{K}^{n_x} \times \mathbb{K}^{n_p} \to \mathbb{R}$ such that

$$\forall (c,Y,Z) \in \mathbb{R}^{n_x} \times \mathbb{K}^{n_x} \times \mathbb{K}^{n_p}\,, \quad \mathcal{B}_t(c,Y,Z) \ \geq \ \max_{\xi,\rho} \left\{ c^{\mathsf{T}} f(t,\xi,\rho) \,\middle|\, \begin{array}{c} \xi \in Y \\ \rho \in Z \\ c^{\mathsf{T}}\xi = V[Y](c) \end{array} \right\} ,$$

as well as another upper-bounding function $\mathcal{B}_0 : \mathbb{R}^{n_x} \times \mathbb{K}^{n_p} \to \mathbb{R}$ such that

$$\forall (c,Z) \in R^{n_x} \times \mathbb{K}^{n_p}\,, \quad \mathcal{B}_0(c,Z) \ \geq \ \max_{\rho} \left\{ c^{\mathsf{T}} x_0(\rho) \mid \rho \in Z \right\} .$$

With this notation, the generalized differential inequality (4.25) can be rewritten in the form

$$\text{a.e. } t \in [0,T]\,, \quad \dot{V}[Y(t,Z)](c) \;=\; \mathcal{B}_t(c,Y(t,Z),Z) \quad \text{with} \quad V[Y(0,Z)](c) \;=\; \mathcal{B}_0(c,Z)\,. \tag{4.47}$$

The following technical assumptions on the upper-bounding functions $\mathcal{B}_t$ and $\mathcal{B}_0$ are made in order to analyze the convergence rate of the enclosure family $Y(\cdot,Z)$ with $Z \subseteq P$ on $[0,T]$.

Assumption 4.1. *Let $\mathcal{S}(\mathbb{K}_{\mathrm{C}}^{n_x})$ a class of representable sets in $\mathbb{K}_{\mathrm{C}}^{n_x}$. There exist integers $m,n \geq 1$ and a set of constants $0 \leq L_0,\ldots,L_n < \infty$ such that the functions $\mathcal{B}_t$ and $\mathcal{B}_0$ satisfy*

$$\begin{aligned}[V[Y](c)+V[Y](-c)]^{n-1}\,[\mathcal{B}_t(c,Y,Z)+\mathcal{B}_t(-c,Y,Z)] \qquad (4.48)\\ \leq \sum_{i=0}^{n} L_i \operatorname{diam}(Y)^{n-i} \operatorname{diam}(Z)^{im}\end{aligned}$$

$$\textit{and} \quad \mathcal{B}_0(c,Z)+\mathcal{B}_0(-c,Z)) \;=\; \mathbf{O}\left(\operatorname{diam}(Z)^m\right) \tag{4.49}$$

for all $c \in \mathbb{R}^{n_x}$ with $\|c\| \leq 1$, all $Y \in \mathcal{S}(\mathbb{K}_{\mathrm{C}}^{n_x})$, and all $Z \subseteq P$.

Theorem 4.4. *Let Assumption 4.1 be satisfied with $m \geq 1$ for a class $\mathcal{S}(\mathbb{K}_{\mathrm{C}}^{n_x})$ of representable sets, and let the right-hand side function $f(t,\cdot,\cdot)$ be locally Lipschitz-continuous. If $V[Y(\cdot,Z)](c) : [0,T] \to \mathbb{R}$ denotes the solution of the initial value problem* (4.47) *for all $c \in \mathbb{R}^{n_x}$ and all $Z \subseteq P$, then $Y(t,\cdot)$ has convergence order m on P, for all $t \in [0,T]$.*

Proof. From conditions (4.48,4.49) of Assumption 4.1 and by property (4.5), there exist an integer $n \geq 1$ and a set of constants $0 \leq L_0,\ldots,L_n < \infty$, $C_0, \geq 1$ and $0 \leq C_1 < \infty$ such that the solutions $V[Y(\cdot,Z)](c)$ to the initial value problem (4.47) also satisfy

$$\begin{aligned}\text{a.e. } t \in [0,T]\,, \quad & \frac{\mathrm{d}}{\mathrm{d}t}\,[V[Y(t,Z)](c)+V[Y(t,Z)](-c)]^n \\ & \leq\; n \sum_{i=0}^{n} \frac{L_i}{C_1} \left[\max_{c' \in \mathbb{R}^n, \|c'\| \leq 1} V[Y(t,Z)](c')+V[Y(t,Z)](-c')\right]^{n-i} \operatorname{diam}(Z)^{im} \\ \text{and} \quad & V[Y(0,Z)](c) \leq C_0 \operatorname{diam}(Z)^m\,,\end{aligned}$$

for all $c \in \mathbb{R}^{n_x}$ with $\|c\| \leq 1$, and all $Z \subseteq P$ with sufficiently small $\operatorname{diam}(Z)$. Then, we have

$$\begin{aligned}\text{a.e. } t \in [0,T]\,, \quad & \frac{\mathrm{d}}{\mathrm{d}t} \left[\max_{c \in \mathbb{R}^n, \|c\| \leq 1} V[Y(t,Z)](c)+V[Y(t,Z)](-c)\right]^n \\ & \leq \max_{c \in \mathbb{R}^n, \|c\| \leq 1} \frac{\mathrm{d}}{\mathrm{d}t}\,[V[Y(t,Z)](c)+V[Y(t,Z)](-c)]^n\end{aligned}$$

$$\leq n \sum_{i=0}^{n} \frac{L_i}{C_1} \left[\max_{c \in \mathbb{R}^n, \|c\| \leq 1} V[Y(t,Z)](c) + V[Y(t,Z)](-c) \right]^{n-i} \operatorname{diam}(Z)^{im} .$$

The result of the theorem follows from the application of Lemma B.3, with the choice of $u(t) := \max_{c \in \mathbb{R}^n, \|c\| \leq 1} V[Y(t,Z)](c) + V[Y(t,Z)](-c)$, and property (4.5). □

4.5.2 Convergence Rate of Standard Differential Inequalities and Ellipsoidal Bounds

Starting with the case of standard differential inequalities as given in Theorem 4.1, we follow the same approach as in the proof of Proposition 4.2 by defining

$$\mathcal{B}_t(c,Y,Z) := c^{\mathsf{T}} \operatorname{mid}(F_t(Y,Z)) + \operatorname{abs}(c)^{\mathsf{T}} \operatorname{rad}(F_t(Y,Z)) ,$$

with the range bounder $F_t(Y,Z) := \left[f_t^{\mathrm{L}}(Y,Z), f_t^{\mathrm{U}}(Y,Z) \right] \in \mathbb{IR}^{n_x}$ such that

$$\left[f_t^{\mathrm{L}}(Y,Z), f_t^{\mathrm{U}}(Y,Z) \right] \supseteq \left[\min_{\xi,\rho} \left\{ f_i(t,\xi,\rho) \middle| \begin{array}{l} \xi_i = y_i^{\mathrm{L}} \\ \xi \in Y \\ \rho \in Z \end{array} \right\}, \max_{\xi,\rho} \left\{ f_i(t,\xi,\rho) \middle| \begin{array}{l} \xi_i = y_i^{\mathrm{U}} \\ \xi \in Y \\ \rho \in Z \end{array} \right\} \right] .$$

for each $i \in \{1,\ldots,n_x\}$. Provided that F_t is such that

$$\forall Y \in \mathbb{IR}^{n_x}, \forall Z \subseteq P, \quad \operatorname{diam}\left(F_t^i(Y,Z)\right) \leq \mathbf{O}(\operatorname{diam}(Y)) + \mathbf{O}(\operatorname{diam}(Z)) , \tag{4.50}$$

it is easy to see that condition (4.48) in Assumption (4.1) can be satisfied with $m = n = 1$ for the class of representable sets $\mathscr{S}(\mathbb{K}_{\mathrm{C}}^{n_x}) := \mathbb{IR}^{n_x}$. In particular, we note that if the right-hand side function $f(t,\cdot,\cdot)$ is locally Lipschitz-continuous on $Y \times Z$, then its range converges with order at least, but in general no better than, $m = 1$; see, e.g., [19, Lemma 1]. Moreover, F_t can be evaluated using natural interval extensions or other convergent forms of interval extensions.

Likewise, provided that a range bounder $X_0(Z) := \left[x_0^{\mathrm{L}}(Z), x_0^{\mathrm{U}}(Z) \right] \in \mathbb{IR}^{n_x}$ is available such that

$$\begin{aligned} X_0(Z) &\supseteq \left[\min_{\rho} \{ x_0(\rho) \,|\, \rho \in Z \}, \max_{\rho} \{ x_0(\rho) \,|\, \rho \in Z \} \right] , \\ \text{and} \quad \forall Z \subseteq P, \quad \operatorname{diam}(X_0(Y,Z)) &\leq \mathbf{O}(\operatorname{diam}(Y)) + \mathbf{O}(\operatorname{diam}(Z)) , \end{aligned} \tag{4.51}$$

condition (4.49) can be satisfied with $m = 1$ as well. This is the case, in general, when the initial value function x_0 is (locally) Lipschitz-continuous and natural interval extensions are used.

With this in mind, the following convergence result follows readily from Theorem 4.4.

Corollary 4.1. *Consider the initial value problem* (4.1)*, and let the right-hand side function f and the initial value function x_0 be locally Lipschitz-continuous in all their arguments. For any subset $Z \subseteq P$ with sufficiently small* $\operatorname{diam}(Z)$*, let the trajectories $y^{\mathrm{L}}(\cdot,Z), y^{\mathrm{U}}(\cdot,Z) : [0,T] \to \mathbb{R}^{n_x}$, with $-\infty < y^{\mathrm{L}}(\cdot,Z) \leq y^{\mathrm{L}}(\cdot,Z) < \infty$, be Lipschitz-continuous and satisfy*

$$\begin{aligned} \text{a.e. } t \in [0,T]\,, \quad \dot{y}_i^{\mathrm{L}}(t,Z) &= f_{t,i}^{\mathrm{L}}\left(\left[y^{\mathrm{L}}(t,Z), y^{\mathrm{U}}(t,Z)\right], Z\right) \\ \dot{y}_i^{\mathrm{U}}(t,Z) &= f_{t,i}^{\mathrm{U}}\left(\left[y^{\mathrm{L}}(t,Z), y^{\mathrm{U}}(t,Z)\right], Z\right), \\ \text{and} \quad y_i^{\mathrm{L}}(0,Z) &= x_0^{\mathrm{L}}(Z), \quad y_i^{\mathrm{U}}(0,Z) = x_0^{\mathrm{U}}(Z)\,. \end{aligned}$$

If conditions (4.50) *and* (4.51) *hold, then the reachable set enclosure $Y(t,\cdot) := \left[y^{\mathrm{L}}(t,\cdot), y^{\mathrm{U}}(t,\cdot)\right]$ has convergence order $m \geq 1$ on P for all $t \in [0,T]$.*

A direct consequence of Corollary 4.1 is that the convergence rate of $Y(t,P)$ towards the actual reachable set $X(t,P)$ as $\operatorname{diam}(P) \to 0$, in the sense of the Hausdorff metric, is itself (at least) linear; that is,

$$d\left(\left[y^{\mathrm{L}}(t,P), y^{\mathrm{U}}(t,P)\right], X(t,P)\right) \leq \mathbf{O}\left(\operatorname{diam}(P)\right)\,.$$

We show in the following example that this bound is sharp.

Example 4.1 (continued)**.** *We revisit the linear ODE system* (4.8) *in the context of convergence analysis. For any interval vector $Z := \rho[-1,1]^2$, with $0 \leq \rho \leq 1$, we have*

$$X(t,Z) = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix} Z \quad \text{and} \quad Y(t,Z) = \exp(t)\,Z\,, \tag{4.52}$$

from which it can be established that $d_{\mathrm{H}}(X(t,Z), Y(t,Z)) \geq (\exp(t) - 1)\operatorname{diam}(Z)$. ◇

Next, we analyze the convergence rate of the ellipsoidal bounding method as formulated in Theorem 4.2. The class of representable sets of interest in this case is $\mathscr{S}(\mathbb{K}_{\mathrm{C}}^{n_x}) := \mathcal{E}(Q)$, with $Q \in \mathbb{S}_+^{n_x}$. Using (4.35), the tight version of the sufficient condition (4.16)—that is, with

equality—can be rewritten in the form of (4.47) with

$$\mathcal{B}_t(c,\mathcal{E}(Q),Z) := \frac{c^\top \left[A(t)Q + QA(t)^\top + \sum_{i=1}^{n_x} \kappa_i(t)Q + (\operatorname{diag}(\kappa(t))^{-1} \operatorname{diag}\operatorname{rad}(\Omega_t(Q))^2\right] c}{2\sqrt{c^\top Qc}}$$

as parameterized by some continuous function $\kappa : [0,T] \to \mathbb{R}^{n_x}_{++}$. It follows that

$$\begin{aligned}
&[V[\mathcal{E}(Q)](c) + V[\mathcal{E}(Q)](-c)]\ [\mathcal{B}_t(c,\mathcal{E}(Q),Z) + \mathcal{B}_t(-c,\mathcal{E}(Q),Z)] \qquad (4.53)\\
&\quad = 2c^\top \left[A(t)Q + QA(t)^\top + \sum_{i=1}^{n_x} \kappa_i(t)Q + (\operatorname{diag}(\kappa(t))^{-1} \operatorname{diag}\operatorname{rad}(\Omega_t(Q))^2\right] c\\
&\quad \leq \|c\|^2 (C_1\|Q\| + C_2\|\operatorname{diag}\operatorname{rad}(\Omega_t(Q))\|)\ ,
\end{aligned}$$

for some constants $0 \leq C_1, C_2 < \infty$. By definition, a nonlinearity bounder Ω_t satisfying (4.14) can always be constructed such that

$$\forall Q \in \mathbb{S}^{n_x}_+,\ \forall Z \subseteq P, \quad \|\operatorname{diag}\operatorname{rad}(\Omega_t(Q))\| \leq \mathbf{O}\left(\operatorname{diam}(\mathcal{E}(Q))^2\right) + \mathbf{O}\left(\operatorname{diam}(Z)^2\right)\ . \qquad (4.54)$$

Combining (4.53) and (4.54) gives

$$\begin{aligned}
&[V[\mathcal{E}(Q)](c) + V[\mathcal{E}(Q)](-c)]\ [\mathcal{B}_t(c,\mathcal{E}(Q),Z) + \mathcal{B}_t(-c,\mathcal{E}(Q),Z)]\\
&\qquad\qquad \leq \mathbf{O}\left(\operatorname{diam}(\mathcal{E}(Q))^2\right) + \mathbf{O}\left(\operatorname{diam}(Z)^4\right),
\end{aligned}$$

for all $c \in \mathbb{R}^{n_x}$ with $\|c\| \leq 1$, all $Q \in \mathbb{S}^{n_x}_+$ with sufficiently small $\operatorname{diam}(\mathcal{E}(Q))$, and all $Z \subseteq P$ with sufficiently small $\operatorname{diam}(Z)$. Therefore, condition (4.48) is satisfied with $m = n = 2$ when (4.54) holds. In practice, Ω_t can be computed as the remainder of a first-order Taylor model of $f(t,\cdot,\cdot)$ at $(\hat{x}(t),\hat{p})$ on $\{\hat{x}(t)\} \oplus G(t)[P - \hat{p}] \oplus \mathcal{E}(Q) \times P$ in the case that $f(t,\cdot,\cdot)$ is twice continuously-differentiable [103, 20].

Likewise, a nonlinearity bounder Ω_0 satisfying (4.15) can be constructed such that

$$\forall Z \subseteq P, \quad \|\operatorname{diag}\operatorname{rad}(\Omega_0)\| \leq \mathbf{O}\left(\operatorname{diam}(Z)^2\right)\ , \qquad (4.55)$$

and condition (4.49) can thus be satisfied with $m = 2$ as well. For instance, Ω_0 can be computed as the remainder of a first-order Taylor model of x_0 at $\hat{p}$ on P provided that x_0 is twice continuously-differentiable.

The following convergence result follows readily from Theorem 4.4 based on the foregoing considerations.

Corollary 4.2. *Consider the initial value problem* (4.1)*, and let the right-hand side function f and the initial value function x_0 be continuously-differentiable in all their arguments. For any subset $Z \subseteq P$ with sufficiently small* $\operatorname{diam}(Z)$*, let the matrix-valued functions $Q(\cdot,Z) : [0,T] \to \mathbb{S}^{n_x}$ be Lipschitz-continuous and such that*

$$\begin{aligned} \text{a.e. } t \in [0,T]\,, \quad \dot{Q}(t,Z) &= A(t)Q(t,Z) + Q(t,Z)A(t)^{\mathsf{T}} + \sum_{i=1}^{n_x} \kappa_i(t)\, Q(t,Z) \\ &\quad + \operatorname{diag}(\kappa(t))^{-1} \operatorname{diag}\operatorname{rad}(\Omega_t(Q(t,Z)))^2 \\ \text{and} \quad Q(0,Z) &= \operatorname{diag}\operatorname{rad}(\Omega_0)^2\,, \end{aligned}$$

where Ω_t and Ω_0 satisfy the conditions (4.54) *and* (4.55)*, and $\kappa : [0,T] \to \mathbb{R}^{n_x}$ is such that $\kappa(t) \in \left[\kappa^{\mathrm{L}}, \kappa^{\mathrm{U}}\right]$ with $0 < \kappa^{\mathrm{L}} \leq \kappa^{\mathrm{U}} < \infty$. Then, the ellipsoidal bound $\mathcal{E}(Q(t,\cdot))$ has convergence order $m \geq 2$ on P for all $t \in [0,T]$.*

A direct consequence of Corollary 4.2 is that the convergence rate of $Y(t,P) := \{\hat{x}(t)\} \oplus G(t)[P-\hat{p}] \oplus \mathcal{E}(Q(t,P))$ towards the actual reachable set $X(t,P)$ when $\operatorname{diam}(P) \to 0$ is itself (at least) quadratic; that is,

$$d\left(\{\hat{x}(t)\} \oplus G(t)[P-\hat{p}] \oplus \mathcal{E}(Q(t,P)), X(t,P)\right) \leq \mathbf{O}\left(\operatorname{diam}(P)^2\right).$$

The sharpness of this bound will be established indirectly in Example 4.3, by checking the convergence rate of first-order Taylor models with ellipsoidal bounds.

4.5.3 Convergence of Taylor Models with Convex Remainder Bound

The following analysis of the convergence properties of the remainder bounds in Taylor models assumes that the right-hand side function $f(t,\cdot,\cdot)$ and the initial value function x_0 are both $(q+1)$-times continuously-differentiable. Here, the bounders $\mathcal{B}_t$ are defined such that

$$\mathcal{B}_t(c,Y,Z) \geq \max_{\xi,\rho} \left\{ c^{\mathsf{T}} r_f^q(t,\xi,\rho) \;\middle|\; \begin{array}{l} \xi \in Y \\ \rho \in Z \\ c^{\mathsf{T}}\xi = V[Y](c) \end{array} \right\},$$

with the remainder function r_f^q given by (4.39). By construction, r_f^q satisfies

$$\begin{aligned}\left\| r_f^q(t,\xi,\rho) \right\| &\leq \left\| f\left(t, \mathcal{P}_x^q(t,\rho), \rho\right) - \dot{\mathcal{P}}_x^q(t,\rho) + \frac{\partial f}{\partial x}(t, \eta\xi, \rho)\,\xi \right\| \quad \text{for some } \eta \in [0,1] \\ &\leq \mathbf{O}\left(\|\rho - \hat{p}\|^{q+1} \right) + \mathbf{O}\left(\|\xi\| \right),\end{aligned}$$

for all $(\xi,\rho) \in \mathcal{E}(Q_x^q(t)) \times P$, where the last inequality follows by the convergence properties of qth-order Taylor models [103, 20]. An enclosure $\mathcal{R}_{f,t}(Y,Z) \supseteq \{r_f^q(t,\xi,\rho) \mid (\xi,\rho) \in Y \times Z\}$ in $\mathbb{K}_{\mathrm{C}}^{n_x}$ satisfying

$$\forall Y \in \mathbb{K}_{\mathrm{C}}^{n_x},\ \forall Z \subseteq P, \quad \operatorname{diam}\left(\mathcal{R}_{f,t}(Y,Z)\right) \leq \mathbf{O}\left(\operatorname{diam}(Y)\right) + \mathbf{O}\left(\operatorname{diam}(Z)^{q+1}\right) \tag{4.56}$$

can thus be obtained, e.g., based on a qth-order Taylor model of $f\left(t, \mathcal{P}_x^q(t,\cdot), \cdot\right)$ on Z as well as on an interval extension of $\frac{\partial f}{\partial x}(t,\cdot,\cdot)$ on $Y \times Z$. It follows that condition (4.48) can be satisfied with $m = q+1$ and $n = 1$ if condition (4.56) holds.

Likewise, the bounder $\mathcal{B}_0$ is defined such that

$$\mathcal{B}_0(c,Z) \geq \max_{\rho} \left\{ c^{\mathsf{T}} \left[x_0(\rho) - \mathcal{P}_{x_0}^q(\rho) \right] \,\middle|\, \rho \in Z \right\},$$

and an enclosure $\mathcal{R}_{x_0}(Z) \supseteq \{x_0(\rho) - \mathcal{P}_{x_0}^q(\rho) \mid \rho \in Z\}$ in $\mathbb{K}_{\mathrm{C}}^{n_x}$ satisfying

$$\forall Z \subseteq P, \quad \operatorname{diam}\left(\mathcal{R}_{x_0}(Z)\right) \leq \mathbf{O}\left(\operatorname{diam}(Z)^{q+1}\right) \tag{4.57}$$

can be obtained, e.g., as the remainder of a qth-order Taylor model of x_0. Therefore, condition (4.49) can be satisfied with $m = q+1$ if (4.57) holds.

At this stage, we note that various parameterization of the remainder enclosure $Y(\cdot,Z)$ can be used, given that conditions (4.48) and (4.49) are satisfied. For instance, both interval remainders and ellipsoidal remainders can be propagated, as described in Sect. 4.4.2. These variants differ in the way the bounders $\mathcal{B}_t$ and $\mathcal{B}_0$ are constructed, similar in essence to the constructions in Sect. 4.5.2.

The following convergence result follows readily from Theorem 4.4 and the above considerations.

Corollary 4.3. *Consider the initial value problem* (4.1)*, and let the right-hand side function f and the initial value function x_0 be $(q+1)$-times continuously-differentiable in all their arguments. For any subset $Z \subseteq P$ with sufficiently small* $\operatorname{diam}(Z)$*, let the set-valued functions $Y(\cdot,Z) : [0,T] \to \mathbb{K}_{\mathrm{C}}^{n_x}$ be such that $V[Y(\cdot,Z)](c)$ is locally Lipschitz-continuous and*

satisfy (4.47) *for all* $c \in \mathbb{R}^{n_x}$. *If conditions* (4.56) *and* (4.57) *hold, then the remainder bound* $\mathcal{R}_x^q(Y(t,\cdot))$ *has convergence order* $m \geq q+1$ *on* P *for all* $t \in [0,T]$.

In turn, the convergence rate of the qth-order Taylor models $\mathcal{P}_x^q(t,P) \oplus \mathcal{R}_x^q(t,P)$ to the actual reachable set $X(t,P)$ as $\operatorname{diam}(P) \to 0$ is of order (at least) $q+1$. For applications in global optimization however, bounding the multivariate polynomial part too presents some challenges. A variety of range-bounding strategies for the Taylor polynomial part have been proposed, some of which enjoy quadratic or higher convergence rate to the actual polynomial range; see, e.g., [77, 103, 80].

In order to keep our considerations general, we assume here that a set-valued function $\mathcal{T}_x^q(\cdot,P) : [0,T] \to \mathbb{R}^{n_x}$ is available such that for all $t \in [0,T]$ we have

$$\mathcal{T}_x^q(t,P) \supseteq \{\mathcal{P}_x^q(t,p) | p \in P\} \;\text{ and }\; d\left(\mathcal{P}_x^q(t,P), \mathcal{T}_x^q(t,P)\right) \leq \mathbf{O}\left(\operatorname{diam}(P)^{\phi(q)}\right),$$

for a known order function $\phi : \mathbb{N} \to \mathbb{N}$. It follows from the triangular inequality (for the Hausdorff metric d) that

$$\forall t \in [0,T], \quad d\left(\mathcal{T}_x^q(t,P) \oplus \mathcal{R}_x^q(t,P), X(t,P)\right) \leq \mathbf{O}\left(\|P\|^{\min\{\phi(q),q+1\}}\right).$$

The convergence properties of Taylor models with interval and ellipsoidal remainder bounds are illustrated in the following example, which assumes exact range bounding of the multivariate polynomial part.

Example 4.3. *The dynamic model of an artificial genetic circuit with three states [158], known as* repressilator*, is considered:*

$$\dot{x}_1 = \frac{p_1}{1+x_3^3} - p_2 x_1 + p_3, \quad \textit{with} \quad x_1(0) = 5.5 \tag{4.58}$$

$$\dot{x}_2 = \frac{p_1}{1+x_1^3} - p_2 x_2 + p_3, \quad \textit{with} \quad x_2(0) = 3.5 \tag{4.59}$$

$$\dot{x}_3 = \frac{p_1}{1+x_2^3} - p_2 x_3 + p_3, \quad \textit{with} \quad x_3(0) = 4.5, \tag{4.60}$$

with $p \in P := ([215,216]\ [0.995,1.005]\ [1.495,1.505])^{\mathsf{T}}$. *Taylor models are propagated continuously in time to enclose the reachable set* $X(t,P)$ *of (4.58-4.60). These estimators are constructed in the same way as earlier in Example 4.2 and further details about the numerical solution procedure will be given in Sect. 4.6 below.*

The left plot of Figure 4.3 shows the projection of $X(t,P)$ *(shaded area) onto* x_1*, as well as the projections of the corresponding Taylor model enclosures* $\mathcal{P}_x^q(t,P) \oplus \mathcal{R}_x^q(t,P)$ *for both*

interval remainder bounds (dashed lines) and ellipsoidal remainder bounds (solid lines). Notice how increasing the Taylor model order q, here from $q = 1$ to 5, delays the time when the enclosure size blows up. Moreover, enclosing the remainder within ellipsoidal bounds instead of interval bounds provides tighter bounds in this case. The right plot of Figure 4.3 illustrates the corresponding convergence rate of $d_{\mathrm{H}}(\mathcal{P}_x^q(t,P) \oplus \mathcal{R}_x^q(t,P), X(t,P))$ at time $t = 10$ as $\mathrm{diam}(P) \to 0$. It is found that all qth-order Taylor enclosure converge with order $q+1$, in agreement with the theory, for both types of remainder enclosures. This also establishes that the convergence bounds derived previously are indeed sharp. ◇

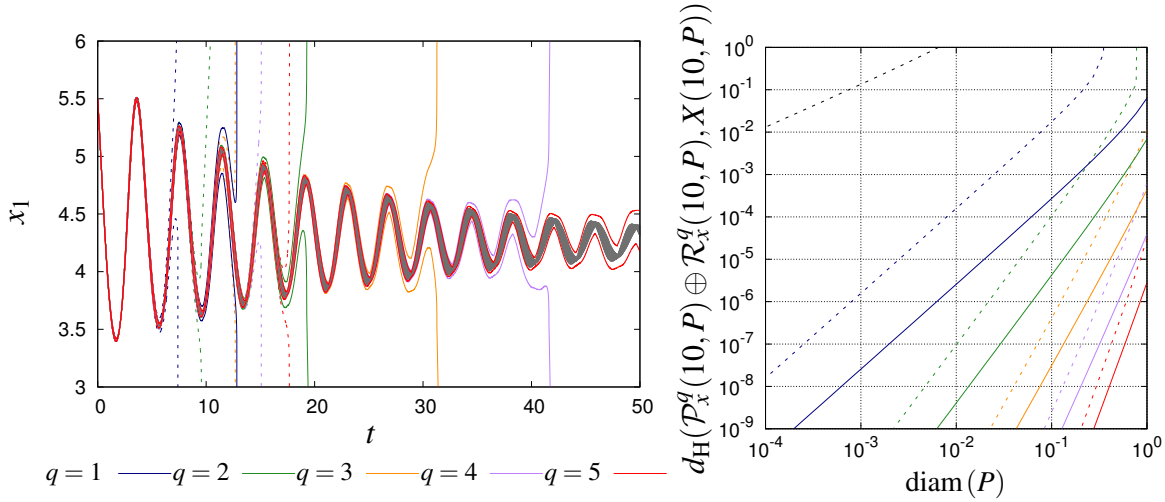

Fig. 4.3 *Left plot:* Projections onto x_1 of the reachable set $X(t)$ and the enclosures $\mathcal{P}_x^q(t,P) \oplus \mathcal{R}_x^q(t,P)$ for both interval and ellipsoidal remainder bounds. *Right plot:* Convergence of the Taylor model enclosure to the actual reachable set at $t = 10$ in the Hausdorff metric $d_{\mathrm{H}}(\mathcal{P}_x^q(t,P) \oplus \mathcal{R}_x^q(t,P), X(t,P))$. The reachable set is represented with a shaded area and the Taylor model enclosures with interval and ellipsoidal remainders with, respectively, dashed lines and solid lines.

4.6 Numerical Implementation and Case Study

This section describes the main implementation details for the bounding approaches presented in Sect. 4.4, before investigating a more challenging case study.

The main difficulty in using the sufficient conditions in Theorems 4.1 and 4.2 and Propositions 4.4 and 4.5 for computing bounds on the parametric solutions of (4.1) is constructing and evaluating the right-hand sides of the auxiliary bounding systems (4.6,4.7), (4.16), (4.41,4.42) and (4.46), respectively. Our implementation assumes that the right-hand side function f in (4.1) and the corresponding initial value function x_0 are both factorable. We

use verified libraries such as PROFIL (`http://www.ti3.tu-harburg.de/`) for interval analysis as well as the library MC++ [93] for Taylor model arithmetic. For numerical ODE integration, we use explicit Runge-Kutta schemes as this avoids computing the Jacobian of the right-hand side of the auxiliary ODEs. More specifically, our implementation relies on the explicit solvers with adaptive step-size control available as part of the GNU Scientific Library (GSL)—function `gsl_odeiv2` in GSL ver. 1.5. The developed continuous-time set-propagation code comes in the form of a C++ class called `ODEBND_GSL`, which is made freely available at: http://www3.imperial.ac.uk/environmentenergyoptimisation/software.

Standard Differential Inequalities Technique Theorem 4.1 is applied with equalities in (4.6,4.7), and the resulting $2n_x$ lower- and upper-bounding ODEs are integrated all together, forward in time, as they are coupled. At a given time $t \in [0,T]$, the right-hand sides of these ODEs are evaluated component-wise. For instance, the right-hand side of the ith lower-bounding ODE (4.6) is evaluated by taking the lower bound from the natural interval extension of $f_i(t,\cdot,\cdot)$ on $L^i \times P$, with $L^i_i := \{\min Y_i(t)\}$ and $L^i_j := Y_j(t)$ if $j \neq i$. Likewise, the right-hand side of the ith upper-bounding ODE (4.7) is evaluated by taking the upper bound from the natural interval extension of $f_i(t,\cdot,\cdot)$ on $U^i \times P$, with $U^i_i := \{\max Y_i(t)\}$ and $U^i_j := Y_j(t)$ if $j \neq i$.

Ellipsoidal Bounding Technique Theorem 4.2 is applied with equality in (4.16), and the ODEs describing $\hat{x}(t)$, $G(t)$ and $Q(t)$ are integrated all together, forward in time. Of the alternatives to construct the right-hand sides of these auxiliary ODEs at a given time $t \in [0,T]$, our implementation considers a q'th-order Taylor model of $f(t,\cdot,\cdot)$ at $(\hat{x}(t),\hat{p})$ on $\Xi_t \times P$, with $q' \geq 1$ and where Ξ_t denotes the interval hull of $\hat{x}(t) \oplus G(t)[P-\hat{p}] \oplus \mathcal{E}(Q(t))$ and the reference point $\hat{p}$ is the midpoint of P. Writing such a Taylor model in the form

$$\forall(\xi,\rho) \in \Xi_t \times P, \quad f(t,\xi,\rho) \in c_{f_t} + L^x_{f_t}(\xi - \hat{x}(t)) + L^p_{f_t}(\rho - \hat{p}) + B_{f_t},$$

with $c_{f_t} \in \mathbb{R}^{n_x}$, $L^x_{f_t} \in \mathbb{R}^{n_x \times n_x}$, $L^p_{f_t} \in \mathbb{R}^{n_x \times n_p}$ and $B_{f_t} \in \mathbb{IR}^{n_x}$, it follows that B_{f_t} can be used as the nonlinearity bounder Ω_t in (4.14) and the right-hand side of the auxiliary bounding ODE system can be evaluated as

$$\begin{aligned}
\dot{\hat{x}}(t) &= c_{f_t} \\
\dot{G}(t) &= L^p_{f_t} \\
\dot{Q}(t) &= (L^x_{f_t})\,Q(t) + Q(t)\,(L^x_{f_t})^{\mathsf{T}} + \sum_{i=1}^{n_x} \kappa_i(t)\,Q(t) + \operatorname{diag}(\kappa(t))^{-1} \operatorname{diag} \operatorname{rad}\left(B_{f_t}\right)^2 .
\end{aligned}$$

The ellipsoidal parameterization $\kappa(t)$ in the latter ODE is so chosen as to minimize $\operatorname{tr}\dot{Q}(t)$. A direct evaluation of $\kappa(t)$ as

$$\forall t \in [0,T]\,, \quad \kappa(t) = \frac{\operatorname{rad}(B_{f_t})}{\sqrt{\operatorname{tr}(Q(t))}}\,,$$

would nonetheless result in $\kappa(t) \to 0$ as $\operatorname{diam}(\mathcal{E}(Q(t))) \to 0$, thereby impairing the quadratic convergence property established in Corollary 4.2. The following modification

$$\forall t \in [0,T]\,,\ \forall i \in \{1,\ldots,N\}\,, \quad \kappa_i(t) = \frac{\sqrt{a_i(t)+\varepsilon}}{\mathrm{TOL}\,\sqrt{a_i(t)+\varepsilon}+\sqrt{b(t)+\varepsilon}} \tag{4.61}$$
$$\text{with} \quad a_i(t) = \frac{1}{s_i(t)}\operatorname{rad}(B_{f_t})\,, \quad b(t) = \operatorname{tr}\Big(\operatorname{diag}(s(t))^{-1}\,Q(t)\Big)\,,$$

and with scaling factors $s_i(t) = 1$ or $s_i(t) = x_i(t)^2$, ensures that $\kappa(t) \geq \frac{1}{1+\mathrm{TOL}} > 0$ for all $t \in [0,T]$—and hence quadratic convergence. Our implementation uses $\varepsilon = 2^{-52}$ and $\mathrm{TOL} = 10^{-6}$ as default values. Initial conditions for the auxiliary ODEs are evaluated as

$$\hat{x}(0) = c_{x_0}\,, \quad G(0) = L_{x_0} \quad \text{and} \quad Q(0) = \operatorname{diag}\operatorname{rad}(B_{x_0})^2\,,$$

based on a q'th-order Taylor model of x_0 at $\hat{p}$ on P given in the form

$$\forall \rho \in P\,, \quad x_0(\rho) \in c_{x_0} + L_{x_0}(\rho - \hat{p}) + B_{x_0}\,,$$

with $c_{x_0} \in \mathbb{R}^{n_x}$, $L_{x_0} \in \mathbb{R}^{n_x \times n_p}$ and $B_{x_0} \in \mathbb{IR}^{n_x}$. Our implementation considers second-order Taylor models of $f(t,\cdot,\cdot)$ and x_0 by default, i.e. $q' = 2$. The use of higher-order Taylor models can provide tighter enclosures by capturing more dependencies in some cases, but this also causes a significant computational overhead.

Taylor Model with Differential Inequalities Bounding Technique A qth-order Taylor model with interval remainder, $(\mathcal{P}_x^q(t,\cdot),[r_x^{q,\mathrm{L}}(t),r_x^{q,\mathrm{U}}(t)])$, is computed based on Proposition 4.4 with equalities in (4.41,4.42). The $2n_x$ remainder bounds are propagated together with the coefficients $\partial^\kappa x(\cdot,\hat{p})$ of the Taylor polynomial (4.40), forward in time. At a given time $t \in [0,T]$ and for each component $i = 1,\ldots,n_x$, our implementation computes a qth-order Taylor model on P at $\hat{p}$ of the composite function

$$f_i(t,(\mathcal{P}_x^q(t,\cdot),L_x^i),\cdot) =: (\mathcal{P}_{f_i}^q, L_{f_i})\,, \quad \text{with} \quad L_{x_j}^i := \begin{cases} [r_{x_j}^{q,\mathrm{L}}(t), r_{x_j}^{q,\mathrm{U}}(t)] & \text{if } i \neq j, \\ \{r_{x_i}^{q,\mathrm{L}}(t)\} & \text{otherwise.} \end{cases}$$

By construction, the coefficients of the multivariate polynomial $\mathcal{P}_{f_i}^q$ match the time derivatives $\partial^\kappa \dot{x}(t,\hat{p})$ of the polynomial coefficients in (4.40), and the lower bound $(\min L_{f_i})$ provides the right-hand side of (4.41). Likewise, the right-hand side of (4.42) is constructed based on a qth-order Taylor model on P at $\hat{p}$ of the composite function

$$f_i(t,(\mathcal{P}_x^q(t,\cdot),U_x^i),\cdot) =: (\mathcal{P}_{f_i}^q, U_{f_i}), \quad \text{with} \quad U_{x_j}^i := \begin{cases} [r_{x_j}^{q,\mathrm{L}}(t), r_{x_j}^{q,\mathrm{U}}(t)] & \text{if } i \neq j, \\ \{r_{x_i}^{q,\mathrm{U}}(t)\} & \text{otherwise.} \end{cases}$$

Taylor Model with Ellipsoidal Remainder Bounding Technique A qth-order Taylor model with ellipsoidal remainder, $(\mathcal{P}_x^q(t,\cdot), Q_x^q(t))$, is computed based on Proposition 4.5 with equality in (4.46). Here, the shape matrix Q_x^q of the ellipsoidal remainder is propagated with the coefficients $\partial^\kappa x(\cdot,\hat{p})$ of the Taylor polynomial (4.40), forward in time. Of the alternatives to construct the right-hand sides of these auxiliary ODEs at a given time $t \in [0,T]$, one approach—referred to as the *full approach* subsequently—involves computing a qth-order Taylor model (with interval remainder) of the composite function $f(t,\mathcal{P}_x^q(t,p)+r,p)$ at $(0,\hat{p})$ for $(r,p) \in R_t \times P$, where R_t denotes the interval hull of $\mathcal{E}(Q_x^q(t))$. Writing such a Taylor model in the form

$$\forall (r,\rho) \in R_t \times P, \quad f(t,\mathcal{P}_x^q(t,\rho)+r,\rho) \in \sum_{\substack{\gamma \in \mathbb{N}^{n_p} \\ |\gamma| \leq q}} \alpha_{f_t,\gamma}(\rho-\hat{p})^\gamma + L_{f_t}^r r + B_{f_t},$$

with $\alpha_{f_t,\gamma} \in \mathbb{R}^{n_x}$, $L_{f_t}^r \in \mathbb{R}^{n_x \times n_x}$ and $B_{f_t} \in \mathbb{IR}^{n_x}$, we have that $L_{f_t}^r$ and B_{f_t} can be used, respectively, as the matrix $A_f(t)$ in (4.43) and as the nonlinearity bounder Ω_t^q in (4.44); that is, the right-hand side of the auxiliary bounding ODE system can be evaluated as

$$\begin{aligned} &\partial^\gamma \dot{x}(t,\hat{p}) = \alpha_{f_t,\gamma}, \quad \text{for all } \gamma \in \mathbb{N}^{n_p}, \text{ with } |\gamma| \leq q \\ &\dot{Q}_x^q(t) = (L_{f_t}^r)\, Q_x^q(t) + Q_x^q(t)\, (L_{f_t}^r)^\mathsf{T} + \sum_{i=1}^{n_x} \kappa_i(t)\, Q_x^q(t) + \operatorname{diag}(\kappa(t))^{-1} \operatorname{diag} \operatorname{rad}\left(B_{f_t}\right)^2. \end{aligned}$$

Moreover, the ellipsoidal parameterization $\kappa(t)$ in (4.46) is chosen in a similar way as (4.61), here for minimizing $\operatorname{tr} \dot{Q}_x^q(t)$, possibly after scaling.

The full approach involves Taylor expanding the function $f(t,\mathcal{P}_x^q(t,p)+r,p)$ jointly in (r,p) up to order q, which can prove computationally demanding when either the dynamic system size n_x or the expansion order q is large. An alternative approach—referred to as the *mean-value approach* hereafter—proceeds in two steps:

i. Compute a qth-order Taylor model (with interval remainder) of the composite function $f(t, \mathcal{P}_x^q(t,\cdot),\cdot)$ at $\hat{p}$ on P:

$$\forall \rho \in P, \quad f(t, \mathcal{P}_x^q(t,\rho),\rho) \in \sum_{\substack{\gamma \in \mathbb{N}^{n_p} \\ |\gamma| \leq q}} \alpha_{f_t,\gamma}(\rho - \hat{p})^\gamma + B_{f_t}^0,$$

with $\alpha_{f_t,\gamma} \in \mathbb{R}$ and $B_{f_t}^0 \in \mathbb{IR}^{n_x}$;

ii. Compute a q'th-order Taylor model (with interval remainder) of the Jacobian matrix $\frac{\partial f}{\partial x}(t, \mathcal{P}_{x_t}^{q'}(p) + r, p)$ at $(0,\hat{p})$ for $(r,p) \in R_{x_t}^{q'} \times P$, with $0 \leq q' \leq q$ and where $(\mathcal{P}_{x_t}^{q'}, R_{x_t}^{q'})$ is a q'th-order Taylor model of $x(t,\cdot)$ on P at $\hat{p}$:

$$\forall (r,\rho) \in R_{x_t}^{q'} \times P, \quad \frac{\partial f}{\partial x}(t, \mathcal{P}_{x_t}^{q'}(\rho) + r, \rho) \in C_{\partial_x f_t} + B_{\partial_x f_t},$$

with $C_{\partial_x f_t} \in \mathbb{R}^{n_x \times n_x}$ and $B_{\partial_x f_t} \in \mathbb{IR}^{n_x \times n_x}$.

By construction, $C_{\partial_x f_t}$ can be used as the matrix $A_f(t)$ in (4.43), whereas $B_{f_t}^0 + (B_{\partial_x f_t} - C_{\partial_x f_t}) R_{x_t}^{q'}$ can be used as the nonlinearity bounder Ω_t^q in (4.44). This way, the right-hand side of the auxiliary bounding ODE system can be evaluated as

$$\begin{aligned}
\partial^\gamma \dot{x}(t,\hat{p}) &= \alpha_{f_t,\gamma}, \quad \text{for all } \gamma \in \mathbb{N}^{n_p}, \text{ with } |\gamma| \leq q \\
\dot{\mathcal{Q}}_x^q(t) &= (C_{\partial_x f_t})\, \mathcal{Q}_x^q(t) + \mathcal{Q}_x^q(t)\, (C_{\partial_x f_t})^\top + \sum_{i=1}^{n_x} \kappa_i(t)\, \mathcal{Q}_x^q(t) \\
&\quad + \operatorname{diag}(\kappa(t))^{-1} \operatorname{diag} \operatorname{rad} \left(B_{f_t}^0 + (B_{\partial_x f_t} - C_{\partial_x f_t}) R_{x_t}^{q'} \right)^2.
\end{aligned}$$

Our implementation of the simplified approach uses the automatic differentiation package FADBAD++ (http://www.fadbad.com/fadbad.html) in order to compute the Jacobian matrix $\frac{\partial f}{\partial x}$ and first-order Taylor models are used to bound the entries of this matrix by default ($q' = 1$). The use of second- or higher-order Taylor models can provide tighter enclosures by capturing more dependencies, but this also causes a significant computational overhead.

4.6.1 Case Study: Anaerobic Digestion

In this section, the anaerobic digestion model presented in Chapter 3 (cf. 3.6.2) is considered

$$\dot{X}_1 = (\mu_1(S_1) - \alpha D) X_1 \tag{4.62}$$

$$\dot{X}_2 = (\mu_2(S_2) - \alpha D) X_2 \tag{4.63}$$

$$\dot{S}_1 = D(S_1^{\text{in}} - S_1) - k_1\mu_1(S_1)X_1 \tag{4.64}$$

$$\dot{S}_2 = D(S_2^{\text{in}} - S_2) + k_2\mu_1(S_1)X_1 - k_3\mu_2(S_2)X_2 \tag{4.65}$$

$$\dot{Z} = D(Z^{\text{in}} - Z) \tag{4.66}$$

$$\dot{C} = D(C^{\text{in}} - C) - q_{\text{CO}_2} + k_4\mu_1(S_1)X_1 + k_5\mu_2(S_2)X_2\,, \tag{4.67}$$

with

$$q_{\text{CO}_2} := k_{\text{L}}a(C + S_2 - Z - K_{\text{H}}P_{\text{CO}_2}) \tag{4.68}$$

$$P_{\text{CO}_2} := \frac{\phi_{\text{CO}_2} - \sqrt{\phi_{\text{CO}_2}^2 - 4K_{\text{H}}P_{\text{t}}(C + S_2 - Z)}}{2K_{\text{H}}} \tag{4.69}$$

$$\phi_{\text{CO}_2} := C + S_2 - Z + K_{\text{H}}P_{\text{t}} + \frac{k_6}{k_{\text{L}}a}\mu_2(S_2)X_2 \tag{4.70}$$

$$\mu_1(S_1) := \bar{\mu}_1\frac{S_1}{S_1 + K_{S_1}} \tag{4.71}$$

$$\mu_2(S_2) := \bar{\mu}_2\frac{S_2}{S_2 + K_{S_2} + S_2^2/K_{I_2}}\,. \tag{4.72}$$

First, we consider initial conditions as $X_1(0) \in 0.5 \times [0.94, 1.06]$ g(COD) L^{-1}, $X_2(0) \in [0.94, 1.06]$ g(COD) L^{-1}, $S_1(0) = 1$ g(COD) L^{-1}, $S_2(0) = 5$ mmol L^{-1}, $Z(0) = 50$ mmol L^{-1}, and $C(0) \in 40 \times [0.94, 1.06]$ mmol L^{-1}; that is, the model contains three uncertain quantities. Projections onto the variables X_1 and S_2 of the actual reachable set and of the enclosures obtained with standard differential inequalities and the ellipsoidal bounding technique as well as their Taylor model counterparts for various expansion orders are shown in Fig. 4.4. Standard differential inequalities are found to produce rather weak bounds here, blowing up after about $t \approx 0.09$ day. Combining Taylor models with differential inequalities delays this blow up time significantly, up to about $t \approx 4.01$ day with 4th-order Taylor models. In comparison, bounds computed with the ellipsoidal technique blow up around $t \approx 1.61$ day, thus outperforming 2nd-order Taylor models with differential inequalities remainder bounds. When used in combination with ellipsoidal remainder bounds, Taylor models delay the blow up time significantly as well. It is even observed that 4th- or higher-order Taylor models with ellipsoidal remainder bounds can stabilize the reachable set enclosure for this level of uncertainty; that is, the bounds converge to the actual steady-state values as $t \to \infty$.

The results reported in Table 4.1 are for reduced uncertainty in the initial conditions as $X_1(0) \in 0.5 \times [0.98, 1.02]$ g(COD) L^{-1}, $X_2(0) \in [0.98, 1.02]$ mmol L^{-1}, $S_1(0) = 1$ mmol L^{-1}, $S_2(0) = 5$ mmol L^{-1}, $Z(0) = 50$ mmol L^{-1}, and $C(0) \in 40 \times [0.98, 1.02]$ mmol L^{-1}. Here

Table 4.1 Comparison of computational performance for various continuous-time enclosure methods. The numerical integration algorithm used is the explicit embedded Runge-Kutta-Fehlberg (4, 5) method in GSL, with relative tolerance RTOL $= 10^{-6}$ and absolute tolerance ATOL $= 10^{-8}$. The reported CPU times are for an Intel™ CORE™i7 vPro™ (4×2.1 GHz) computer with 8Gb RAM, running Ubuntu 12.04.3 LTS (GNU/Linux 3.5.0-45-generic x86_64) and gcc version 4.7.3 (with -O2 compilation option).

Differential inequalities (DI) and ellipsoidal bounding (EB) techniques				
Technique q	DI	EB ($q' = 1$)	EB ($q' = 2$)	
Blow up time [day]	0.39	1.99	13.55	
RHS evaluations for $t \in [0,4]$	N/A	N/A	812	
CPU time in [s] for $t \in [0,4]$	N/A	N/A	0.54	
Taylor model (TM) w/ differential inequalities remainder bounds				
TM Order q	2	3	4	5
Blow up time [day]	3.07	5.31	7.58	9.89
RHS evaluations for $t \in [0,4]$	N/A	1,077	1,191	1,272
CPU time in [s] for $t \in [0,4]$	N/A	0.39	0.72	1.33
Taylor model (TM) w/ ellipsoidal remainder bounds – Mean-value approach ($q' = 0$)				
TM Order q	2	3	4	5
Blow up time [day]	4.23	11.12	∞	∞
RHS evaluations for $t \in [0,4]$	972	1,077	1,191	1,272
CPU time in [s] for $t \in [0,4]$	0.14	0.23	0.41	0.71
Taylor model (TM) w/ ellipsoidal remainder bounds – Mean-value approach ($q' = 1$)				
TM Order q	2	3	4	5
Blow up time [day]	∞	∞	∞	∞
RHS evaluations for $t \in [0,4]$	954	1,077	1,191	1,272
CPU time in [s] for $t \in [0,4]$	0.98	1.20	1.47	1.85
Taylor model (TM) w/ ellipsoidal remainder bounds – Full approach				
TM Order q	2	3	4	5
Blow up time [day]	39.84	∞	∞	∞
RHS evaluations for $t \in [0,4]$	954	1,077	1,191	1,272
CPU time in [s] for $t \in [0,4]$	0.35	1.49	6.35	26.37

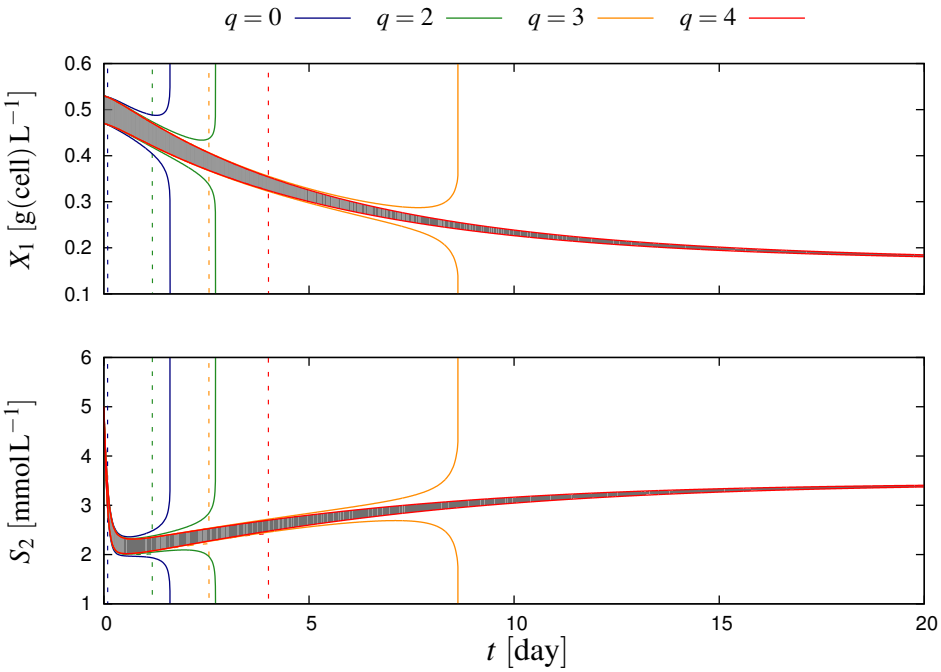

Fig. 4.4 Projections onto X_1 (top plot) and C (bottom plot) of the reachable set $X(t)$ and the enclosures $Y(t)$ computed with various approaches. The reachable set is represented with a shaded area. The enclosures labelled with $q = 0$ correspond to standard differential inequalities (dashed lines) and the ellipsoidal bounding approach (solid lines). The enclosures labelled $q = 2$, 3 and 4 are Taylor model enclosures with interval remainders (dashed lines) and ellipsoidal remainders (with full approach; solid lines).

again, standard differential inequalities produce the weakest bounds, with a blow up time of $t \approx 0.40$ day. The use of ellipsoidal calculus improves the bounds significantly, by delaying the blow up time to $t \approx 1.99$ day and $t \approx 13.52$ day when the nonlinearity bounder Ω_t is constructed by using, respectively, 1st-order and 2nd-order Taylor models. The use of Taylor models as a means of propagating nonconvex enclosures provides tighter bounds as the Taylor model order is increased, which illustrates well the advantage of bounding techniques enjoying higher-order Hausdorff convergence. Ellipsoidal remainder bounds are found to outperform differential inequalities remainder bounds in this case as well. In particular, all three variants of Taylor models with ellipsoidal remainders appear to stabilize the reachable set enclosures for small Taylor expansion orders—as small as $q = 2$ in the mean-value approach with $q' = 1$.

Comparisons of computational performance are made for a fixed integration horizon of $t \in [0,4]$. As far as Taylor model-based techniques are concerned, the number of right-hand side evaluations—and therefore the number of integration steps—increases slowly with the Taylor expansion order between $q = 2,\ldots,5$, which is due to a larger size of the auxiliary

bounding system while the tolerances remain unchanged. For a particular Taylor expansion order q, the use of various bounding schemes for the Taylor remainder does not affect the number of right-hand side evaluations though; this suggests that the step-size is mainly determined by the propagation of the Taylor polynomial here. Regarding CPU times, a significant increase is observed in the full approach of ellipsoidal remainder bounds with the Taylor expansion order between $q = 2,\ldots,5$ compared to differential inequalities remainder bounds. This computational burden is greatly reduced by the use of the mean-value approach with either $q' = 0$ or 1. In particular, the ellipsoidal remainder bound variant with $q' = 0$ appears to be both faster and to provide tighter bounds than the differential inequalities remainder bound technique as well as the standard ellipsoidal bounding technique.

4.7 Robust Forward Invariant Tubes for Control-Affine Systems

In this section we study the construction of enclosures for for nonlinear control systems. In particular, the focus is on control-affine systems of the form:

$$\dot{x}(t) = f(x(t), w(t)) + G(x(t))u(t) \tag{4.73}$$

where $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \to \mathbb{R}^{n_x}$, $G : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x \times n_u}$ and $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_w} \to \mathbb{R}^{n_x}$ are potentially nonlinear functions, for which regularity assumption will be stated later on as necessary. The state trajectory is denoted by $x \in \mathbb{L}^{n_x}$, $u \in \mathbb{U} := \{u \in \mathbb{L}^{n_u} \mid \forall t \in \mathbb{R},\ u(t) \in U \subseteq \mathbb{R}^{n_u}\}$ denotes the control, and $w \in \mathbb{W} := \{w \in \mathbb{L}^{n_w} \mid \forall t \in \mathbb{R},\ w(t) \in W \subseteq \mathbb{R}^{n_w}\}$ denotes the exogenous disturbance.

Before proceeding with the construction of tubes for such control systems, we introduce the following technical assumptions:

Assumption 4.2. *The sets $U \subseteq \mathbb{R}^{n_u}$ and $W \subseteq \mathbb{R}^{n_w}$ are compact and convex, i.e. $U \in \mathbb{K}_C^{n_u}$ and $W \in \mathbb{K}_C^{n_w}$.*

Assumption 4.3. *The function f is jointly continuous in x, w and locally Lipschitz-continuous in x. Moreover, the function G is continuously differentiable.*

For a given control $u \in \mathbb{U}$ and a given set of initial states $X_1 \in \mathbb{K}_C^{n_x}$ at t_1, we denote the reachable set of (4.73) at $t_2 > t_1$ as:

$$X(t_2) := \left\{ \xi \in \mathbb{R}^{n_x} \;\middle|\; \begin{array}{l} \exists x \in \mathbb{L}^{n_x}, \exists w \in \mathbb{W} : \\ \text{a.e. } t \in [t_1, t_2], \\ \quad \dot{x}(t) = f(x(t), w(t)) + G(x(t))u(t) \\ x(t_1) \in X_1\,, \;\; x(t_2) = \xi \end{array} \right\}.$$

An enclosure of the reachable set for the class of systems considered in this section is given by a straightforward application of Theorem 4.3

Corollary 4.4. *Consider the uncertain dynamic system* (4.73) *with initial condition* $x(t_1) \in X_1$*, with* $X_1 \in \mathbb{K}_C^{n_x}$*, and a given control* $u \in \mathbb{U}$*, and let Assumptions 4.2 and 4.3 hold. Let* $Y : [t_1, t_2] \to \mathbb{K}_C^{n_x}$ *be a set-valued function such that*

1. *the function* $V[Y(\cdot)](c)$ *is for all* $c \in \mathbb{R}^{n_x}$ *Lipschitz-continuous on* $[t_1, t_2]$ *and*

2. *the set-valued function* Y *satisfies for all* $c \in \mathbb{R}^{n_x}$ *the differential inequality*

$$\text{a.e. } t \in [t_1, t_2], \quad \dot{V}[Y(t)](c) \geq \max_{\xi, \omega} \left\{ c^\top f(\xi, \omega) + c^\top G(\xi) u(t) \;\middle|\; \begin{array}{l} c^\top \xi = V[Y(t)](c) \\ \xi \in Y(t) \\ \omega \in W \end{array} \right\}$$

$$\text{and} \qquad V[Y(t_1)](c) \geq \max_{\xi} \left\{ c^\top \xi \,\middle|\, \xi \in X_1 \right\}.$$

Then, Y *is an enclosure of the reachable set of* (4.73)*, i.e.* $Y(t) \supseteq X(t)$ *for all* $t \in [t_1, t_2]$*.*

Proof. Follows directly from Theorem 4.3 and Remark 4.2. □

The idea now is that the control policy u, is not given, but chosen in the set $\mathbb{U}$ of admissible controllers in order to reduce the cross-section of the tube, while accounting for every possible realization of the exogenous disturbance $w \in \mathbb{W}$. Furthermore, we are interested in computing a feedback control function, such that if the state of the system is within the tube, we can make sure to keep it inside regardless of the uncertainty. This is formalized in the following definition

Definition 4.3. *The set-valued function* $Y : [t_1, t_2] \to \Pi(\mathbb{R}^{n_x})$ *is called a RFIT for* (4.73) *on* $[t_1, t_2]$*, if there exists a feedback control law* $\mu : [t_1, t_2] \times \mathbb{R}^{n_x} \to U$ *such that any solution of*

the controlled system

$$\forall t \in [t_1, t_2], \quad \dot{x}(t) = f(x(t), w(t)) + G(x(t))\mu(t, x(t)),$$

with $x(t) \in Y(t)$, satisfies $x(t') \in Y(t')$ for all $t, t' \in [t_1, t_2]$ with $t' \geq t$ and all $w \in \mathbb{W}$.

The following theorem, provides sufficient conditions for a set-valued function to be an RFIT of a given control-affine nonlinear system. These sufficient conditions come in the form of a min-max differential inequality (DI), which describe the convex cross-sections of a RFIT in terms of their support functions.

Theorem 4.5. *Consider the uncertain dynamic system* (4.73)*, and let Assumptions 4.2 and 4.3 hold. Let $Y : [t_1, t_2] \to \mathbb{K}_C^{n_x}$ be a set-valued function such that*

1. *the function $V[Y(\cdot)](c)$ is for all $c \in \mathbb{R}^{n_x}$ Lipschitz-continuous on $[t_1, t_2]$ and*
2. *the set-valued function Y satisfies for all $c \in \mathbb{R}^{n_x}$ the differential inequality*

$$\text{a.e. } t \in [t_1, t_2],$$
$$\dot{V}[Y(t)](c) \geq \min_{\nu \in U} \max_{\xi, \omega} \left\{ c^\top f(\xi, \omega) + c^\top G(\xi)\nu \;\middle|\; \begin{array}{l} c^\top \xi = V[Y(t)](c) \\ \xi \in Y(t) \\ \omega \in W \end{array} \right\} \tag{4.74}$$

Then, Y is an RFIT for all $t \in [t_1, t_2]$.

Proof. The proof proceeds in two steps. In the first step (S1), we establish the results under the following auxiliary assumptions:

A4 The set of admissible controls U is smooth with positive curvature;

A5 The pointwise-in-time cross-sections $Y(t)$ of the tube Y are smooth with positive curvatures at each $t \in [t_1, t_2]$.

In the second step (S2), we argue that the result still holds by removing these extra assumptions.

S1 We start by noting that since the inequality (4.74) is invariant under scaling of the directions $c \in \mathbb{R}^{n_x}$, it is sufficient to consider those directions c with $c^\top c = 1$, namely $c \in \mathcal{S}^{n_x - 1}$.

It follows from Assumption A5 that the Gauss map $\mathcal{G}_{Y(t)} : \mathrm{bd}\, Y(t) \to \mathcal{S}^{n_x-1}$ is a diffeomorphism [125]. For all $c \in \mathcal{S}^{n_x-1}$ and all $t \in [t_1, t_2]$, the inverse Gauss map values $\mathcal{G}^{-1}_{Y(t)}(c)$ correspond to the elements of the singletons

$$\Psi_t(c) := \left\{ \xi \in \mathbb{R}^{n_x} \middle| \begin{array}{c} \xi \in Y(t) \\ c^\top \xi = V[Y(t)](c) \end{array} \right\}.$$

In particular, we have

$$\begin{aligned} \min_{\nu \in U} V[\Gamma_g(\nu, c, Y(t))](c) &= \max_{\substack{\xi \in \Psi_t(c), \\ \omega \in W}} c^\top f(\xi, \omega) + \min_{\nu \in U} \max_{\xi \in \Psi_t(c)} c^\top G(\xi)\nu \\ &= \max_{\omega \in W} c^\top f\left(\mathcal{G}^{-1}_{Y(t)}(c), \omega\right) + \min_{\nu \in U} c^\top G\left(\mathcal{G}^{-1}_{Y(t)}(c)\right)\nu . \end{aligned} \tag{4.75}$$

Moreover, by Lipschitz continuity of $V[Y(\cdot)](c)$ on $[t_1, t_2]$, the functions $\mathcal{G}^{-1}_{Y(\cdot)}(c) : [t_1, t_2] \to \mathrm{bd}\, Y(t)$ are continuous for each $c \in \mathcal{S}^{n_x-1}$.

Next, we focus on the minimization subproblem in the right-hand side of (4.75). By continuity of $\mathcal{G}^{-1}_{Y(t)}$ and G (Assumption 4.3) and by compactness of U (Assumption 4.2), the sets

$$\arg\min_{\nu \in U} c^\top G\left(\mathcal{G}^{-1}_{Y(t)}(c)\right)\nu$$

are singletons for all $c \in \mathcal{S}^{n_x-1}$ and all $t \in [t_1, t_2]$, and we can define the function μ_t^* as

$$\mu_t^*(c) := \arg\min_{\nu \in U} c^\top G\left(\mathcal{G}^{-1}_{Y(t)}(c)\right)\nu .$$

Since $\mu_t^*(c)$ is always attained at the boundary of U, it follows by Assumption A4, by continuous differentiability of $\mathcal{G}^{-1}_{Y(t)}$ and G (Assumption 4.3), and from sensitivity theory [38] that $\mu_t^*(\cdot)$ is continuously differentiable on $\mathcal{S}^{n_x-1}$, for each $t \in [t_1, t_2]$. Moreover, the function $\mu^*_{(\cdot)}(c) : [t_1, t_2] \to U$ is continuous for each $c \in \mathcal{S}^{n_x-1}$.

The result follows from the application of Corollary 4.4 to the auxiliary ODE

$$\dot{x}(t) = f(x(t), w(t)) + G(x(t))\mu(t, x(t)) ,$$

with $\mu(t,x) := \mu_t^*(\mathcal{G}_{Y(t)}(x))$. In particular, $\mu(t,x)$ provides a feedback control law for the RFIT Y under the auxiliary Assumptions A4 and A5.

S2 In the case that certain tube cross-sections $Y(t)$ or the control constraint set U fail to be smooth with positive curvature on $[t_1,t_2]$, we can always construct a family of set-valued functions $Y_\varepsilon : [t_1,t_2] \to \mathbb{K}_C^{n_x}$ as well as a family of compact sets $U_\varepsilon \subseteq U$ with smooth boundary and postive curvature such that the following statements hold for all $\varepsilon > 0$:

1. $Y_\varepsilon(t) \supseteq Y(t)$ for all $t \in [t_1,t_2]$, $\quad U_\varepsilon \supseteq U$.

2. There exists a continuous function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ with $\alpha(0) = 0$ such that

$$\begin{aligned} &d_{\mathrm{H}}(Y_\varepsilon(t), Y(t)) \leq \alpha(\varepsilon), \quad d_{\mathrm{H}}(U_\varepsilon, U) \leq \alpha(\varepsilon), \\ &\text{and} \quad \dot{V}[Y_\varepsilon(t)](c) \geq \dot{V}[Y(t)](c) + L\alpha(\varepsilon) \end{aligned}$$

   for all $t \in [t_1,t_2]$ with $L := \frac{1}{2(t_2-t_1)}$. In particular, L can be made arbitrarily large by choosing $t_2 - t_1$ significantly small.

Such outer approximations have been used in the Proof of Theorem 4.3 (See Lemma B.1 and B.2). This way, the result follows from the application of the procedure in S1 above and taking the limit as $\varepsilon \to 0$ by invoking a continuity argument. In detail, we can (without loss of generality) choose $t_2 - t_1$ sufficiently small so that

$$\begin{aligned} \dot{V}[Y_\varepsilon(t)](c) &\geq \dot{V}[Y(t)](c) + L\alpha(\varepsilon) \\ &\geq \min_{\nu \in U} V[\Gamma_g(\nu, c, Y(t))](c) + L\alpha(\varepsilon) \\ &\geq \min_{\nu \in U_\varepsilon} V[\Gamma_g(\nu, c, Y_\varepsilon(t))](c) \end{aligned} \tag{4.76}$$

for sufficiently small $\varepsilon > 0$. Thus, it follows from step S1 that Y_ε is a RFIT for all sufficiently small $\varepsilon > 0$. The final technical difficulty involves analyzing the limit behavior of the sequence

$$\begin{aligned} \mu_\varepsilon(t,\xi) &:= \mu^*_{t,\varepsilon}(\mathcal{G}_{Y_\varepsilon(t)}(\xi)) \\ \text{with } \mu^*_{t,\varepsilon}(c) &:= \arg\min_{\nu \in U_\varepsilon} c^\top G\left(\mathcal{G}^{-1}_{Y_\varepsilon(t)}(c)\right)\nu \,, \end{aligned}$$

which may fail to converge as $\varepsilon \to 0$. Since $\mu_\varepsilon(t,\xi)$ takes values in U_ε only and the sets U_ε converge in the Hausdorff sense to a compact set U, the sequence $\mu_\varepsilon(t,\xi)$ is bounded uniformly with respect to $\varepsilon > 0$. Consequently, we can use the Bolzano-Weierstrass theorem to establish the existence of a sequence $\varepsilon_1, \varepsilon_2, \ldots \in \mathbb{R}_+$ with $\lim_{i\to\infty} \varepsilon_i \to 0$ such that the limit

$$\mu(t,\xi) = \lim_{i \to \infty} \mu_{\varepsilon_i}(t,\xi)$$

exists. By construction, $\mu(t,\xi)$ is a control law that generates the limit tube Y, therefore Y is an RFIT. □

The following corollary is a direct side-product of the proof of Theorem 4.5.

Corollary 4.5. *Let the set-valued function $Y : [t_1,t_2] \to \mathbb{K}_C^{n_x}$ satisfy the conditions of Theorem 4.5. Under the additional regularity conditions that the boundaries of the tube cross-sections $Y(t)$ for all $t \in [t_1,t_2]$ and the set of admissible controls U are smooth with positive curvature, an explicit feedback control law generating the RFIT is given by*

$$\mu(t,x(t)) = \mu_t^* \left(\mathcal{G}_{Y(t)}(x(t))\right), \tag{4.77a}$$

where $\mathcal{G}_{Y(t)}$ denotes the Gauss map of $Y(t)$ and

$$\mu_t^*(c) := \underset{v\in U}{\arg\min}\, c^\top G \left(\mathcal{G}_{Y(t)}^{-1}(c)\right) v\,, \tag{4.77b}$$

with $\mathcal{G}_{Y(t)}^{-1}$ being the inverse Gauss map of $Y(t)$.

Proof. Follows directly from Step S1 in the proof of Theorem 4.5. □

Remark 4.3. *The feedback control law given by Eqs.* (4.77) *is not necessarily unique.*

Although heavily inspired by set-theoretic methods for the synthesis of model predictive controllers (see [115] for an introduction), Theorem 4.5 also provides a constructive approach for nonlinear feedback control laws by exploiting properties at the boundaries of RFITs. To the best of our knowledge, this approach has not been exploited as of yet in the robust MPC literature. This construction is the focus of the following sections.

4.8 Ellipsoidal Robust Forward Invariant Tubes

This section derives practical and computationally tractable conditions for checking whether a particular set-valued function Y is a RFIT for system (4.73). The focus is on tubes with ellipsoidal cross-sections given by

$$Y(t) = \mathcal{E}(q_x(t), Q_x(t))\,, \tag{4.78}$$

where $q_x(t) \in \mathbb{R}^{n_x}$ and $Q_x(t) \in \mathbb{S}_+^{n_x}$ denote the center and shape matrix of the tube, pointwise in time. Moreover, we make the following assumptions:

Assumption 4.4. *The uncertainty and admissible control sets are such that* $W \subseteq \mathcal{E}(q_w, Q_w)$ *and* $U \subseteq \mathcal{E}(q_u, Q_u)$*, with* $(q_w, Q_w) \in \mathbb{R}^{n_w} \times \mathbb{S}_+^{n_w}$ *and* $(q_u, Q_u) \in \mathbb{R}^{n_u} \times \mathbb{S}_+^{n_u}$.

Assumption 4.5. *The functions* f *and* G *are twice continuously differentiable in all of their arguments.*

The following construction of ellipsoidal tubes based on Theorem 4.5 uses the same ideas as the construction of ellipsoidal bounds for uncertain ODEs based on Theorem 4.3; see, e.g., [68, 56, 148]. It involves decomposing the state x, control u, and disturbance w into their nominal and perturbed components as

$$\begin{aligned} x(t) &= q_x(t) + \delta_x(t) \\ u(t) &= q_u + \delta_u(t) \\ w(t) &= q_w + \delta_w(t)\,. \end{aligned}$$

First, we choose q_x so that it satisfies the ODE

$$\dot{q}_x(t) = f(q_x(t), q_w) + G(q_x(t))\, u_x(t)\,,$$

for some reference input $u_x(t) \in \mathcal{E}(q_u, Q_u)$. Here, $u_x \in \mathbb{U}$ is the control input for the nominal system, i.e. when no disturbance is present or in our case, when the disturbance takes its nominal value q_w. It follows that the perturbed state component δ_x satisfies the ODE

$$\begin{aligned} \dot{\delta}_x(t) = \;& A(q_x(t))\delta_x(t) + B(q_x(t))\delta_w(t) + G(q_x(t) + \delta_x(t))\delta_u(t) \\ & + n(t, \delta_x(t), \delta_w(t), \delta_u(t))\,, \end{aligned} \tag{4.79}$$

where $n : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ is defined in such a way that (4.79) is equivalent to (4.73); and

$$A(q_x(t)) := \frac{\partial f}{\partial x}(q_x(t), q_w) + \frac{\partial G}{\partial x}(q_x(t))q_u\,, \quad \text{and} \quad B(q_x(t)) := \frac{\partial f}{\partial w}(q_x(t), q_w)\,.$$

At this point, we introduce the following technical assumptions regarding the control constraint set and the nonlinearities in the functions G and n.

Assumption 4.6. *For any nominal control* $u_x \in \mathbb{U}$*, there exists a matrix-valued function* $R_u : [t_1, t_2] \to \mathbb{S}_+^{n_u}$ *such that*

$$\forall t \in [t_1, t_2] : \quad \mathcal{E}(q_u, R_u(t)) \subseteq \mathcal{E}(u_x(t), Q_u)\,,$$

Assumption 4.7. *There exists a nonlinearity bounder $\Omega_n : \mathbb{R}^{n_x} \times \mathbb{S}_+^{n_x} \to \mathbb{S}_+^{n_x}$ for the function n such that*

$$n(t, \xi, \omega, \nu) \in \mathcal{E}(\Omega_n(q_x(t), Q_x(t))),$$

for all $t \in [t_1, t_2]$, all $\xi \in \mathcal{E}(Q_x(t))$, all $\omega \in \mathcal{E}(Q_w)$, and all $\nu \in \mathcal{E}(Q_u)$.

Assumption 4.8. *There exists a nonlinearity bounder $\Omega_G : \mathbb{R}^{n_x} \times \mathbb{S}_+^{n_x} \times \mathbb{S}_+^{n_u} \times \mathbb{R}^{n_x \times n_u} \to \mathbb{S}_+^{n_x}$ such that*

$$\begin{aligned}\Omega_G(q_x(t), Q_x(t), R_u(t), S_0) \succeq \; & Q_x^{\frac{1}{2}}(t) S_0 R_u^{\frac{1}{2}}(t) G(\xi)^\mathsf{T} + G(\xi) R_u^{\frac{1}{2}}(t) S_0^\mathsf{T} Q_x^{\frac{1}{2}}(t) \\ & - Q_x^{\frac{1}{2}}(t) S_0 R_u^{\frac{1}{2}}(t) G(q_x(t))^\mathsf{T} - G(q_x(t)) R_u^{\frac{1}{2}}(t) S_0^\mathsf{T} Q_x^{\frac{1}{2}}(t),\end{aligned}$$

for all $t \in [t_1, t_2]$, all $\xi \in \mathcal{E}(q_x(t), Q_x(t))$ all $S_0 \in \mathbb{R}^{n_x \times n_u}$ with $S_0 S_0^\mathsf{T} \preceq I$ and all R_u satisfying Assumption 4.6.

A set of sufficient conditions for a tube with ellipsoidal cross-section to be a RFIT for system (4.73) are stated in the following theorem. For notational convenience, we introduce the set-valued function $\Phi_g : \mathbb{R}^{n_x} \times \mathbb{S}_+^{n_x} \times \mathbb{R}^{n_x \times n_u} \times \mathbb{S}_+^{n_u} \times \mathbb{R}_{++} \times \mathbb{R}_{++} \to \mathbb{S}_+^{n_x}$ associated with the right-hand-side function of (4.73) given by

$$\begin{aligned}\Phi(q_x(t), Q_x(t), S_0, R_u(t), \lambda_0, \kappa_0) := \; & A(q_x(t)) Q_x(t) + Q_x(t) A(q_x(t))^\mathsf{T} \\ & + Q_x^{\frac{1}{2}}(t) S_0 R_u^{\frac{1}{2}}(t) G(q_x(t))^\mathsf{T} + G(q_x(t)) R_u^{\frac{1}{2}}(t) S_0^\mathsf{T} Q_x^{\frac{1}{2}}(t) \\ & + \left(\frac{1}{\lambda_0} + \frac{1}{\kappa_0} \right) Q_x(t) + \Omega_G(q_x(t), Q_x(t), S_0) \\ & + \lambda_0 B(q_x(t)) Q_w B(q_x(t))^\mathsf{T} + \kappa_0 \Omega_n(q_x(t), Q_x(t)).\end{aligned}$$

Theorem 4.6. *Consider the uncertain dynamic system* (4.73)*, let $u_x \in \mathbb{U}$ be a reference control and let Assumptions 4.4-4.8 hold. If the functions $Q_x : [t_1, t_2] \to \mathbb{S}_+^{n_x}$ and $q_x : [t_1, t_2] \to \mathbb{R}^{n_x}$ satisfy*

$$\dot{q}_x(t) = f(q_x(t), q_w) + G(q_x(t)) u_x(t), \tag{4.80}$$

$$\dot{Q}_x(t) \succeq \Phi(q_x(t), Q_x(t), S(t), R_u(t), \lambda(t), \kappa(t)), \tag{4.81}$$

for some functions $\lambda, \kappa : [t_1, t_2] \to \mathbb{R}_{++}$ and $S : [t_1, t_2] \to \mathbb{R}^{n_x \times n_u}$ with $S(t) S(t)^\mathsf{T} \preceq I$, then $Y(t) := \mathcal{E}(q_x(t), Q_x(t))$ is a RFIT for (4.73) *on $[t_1, t_2]$.*

Proof. In analogy to the proof of Theorem 4.5, the following proof proceeds in two steps. In the first step (S1), we establish the results under the following auxiliary assumption:

A6 The shape matrices Q_u and $Q_x(t)$, $t_1 \leq t \leq t_2$, are positive definite.

In the second step (S2), we argue that the result still holds by removing this extra assumption.

S1 The idea in this part of the proof is to show that the conditions (4.80)–(4.81) imply the min-max differential inequality (4.74) for $Y(t) := \mathcal{E}(q_x(t), Q_x(t))$. Using the state decomposition into nominal part (4.80) and perturbed part (4.79) as well as Assumptions 4.6 and 4.7, we want to show that

$$\begin{aligned}
&\dot{V}[\mathcal{E}(Q_x(t))](c) \\
&\quad \geq \min_{v \in \mathcal{E}(R_u(t))} \max_{\xi, \omega_1, \omega_2} \left\{ \begin{array}{l|l} c^\top A(q_x(t))\xi & c^\top \xi = V[\mathcal{E}(Q_x(t))](c) \\ + c^\top B(q_x(t))\omega_1 & \xi \in \mathcal{E}(Q_x(t)) \\ + c^\top G(q_x(t)+\xi)v & \omega_1 \in \mathcal{E}(Q_w) \\ + c^\top \omega_2 & \omega_2 \in \mathcal{E}(\Omega_n(q_x(t), Q_x(t))) \end{array} \right\}.
\end{aligned} \tag{4.82}$$

for a.e. $t \in [t_1, t_2]$ and all $c \in \mathbb{R}^{n_x}$ such that $c^\top c = 1$ and $\mathcal{E}(R_u(t)) \subseteq \mathcal{E}(u_x(t) - q_u, Q_u)$. Notice that the ellipsoid $\mathcal{E}(R_u(t))$ needs to be introduced since the central path defined by Eq. (4.80) is evaluated along $u_x(t) \in \mathcal{E}(q_u, Q_u)$ and not along the center q_u.

Now, consider a family of ellipsoids parameterised by the matrix valued function $Q_x : [t_1, t_2] \to \mathbb{S}^{n_x}_{++}$. For each $t \in [t_1, t_2]$ the Gauss map $\mathcal{G}_{\mathcal{E}_x(t)} : \mathrm{bd}\,\mathcal{E}(Q_x(t)) \to \mathcal{S}^{n_x - 1}$ is a diffeomorphism under Assumption A6, given by:

$$\mathcal{G}_{\mathcal{E}_x(t)}(\xi) := \frac{Q_x^{-1}(t)\xi}{\left\| Q_x^{-1}(t)\xi \right\|_2}, \qquad \mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c) = \frac{Q_x(t)c}{\sqrt{c^\top Q_x(t) c}}.$$

In particular, the right-hand side of Condition (4.82) is given by

$$\begin{aligned}
&c^\top A(q_x(t))\mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c) + \min_{v \in \mathcal{E}(R_u(t))} c^\top G\left(q_x(t) + \mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c)\right) v \\
&\quad + \max_{\omega_1 \in \mathcal{E}(Q_w)} c^\top B(q_x(t))\omega_1 + \max_{\omega_2 \in \mathcal{E}(\Omega_n(q_x(t), Q_x(t)))} c^\top \omega_2 .
\end{aligned}$$

Since, for any matrices $D \in \mathbb{R}^{n_x \times n_\zeta}$ and $Q_\zeta \in \mathbb{S}^{n_\zeta}_+$

$$\max_\zeta / \min_\zeta \left\{ c^\top D\zeta \,\middle|\, \zeta \in \mathcal{E}(Q_\zeta) \right\} = \pm\sqrt{c^\top D Q_\zeta D^\top c},$$

we can write Condition (4.82) as

$$\begin{aligned}\dot{V}[\mathcal{E}(Q_x(t))](c) \geq \; & c^\mathsf{T} A(q_x(t))\mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c) \\ & - \sqrt{c^\mathsf{T} G\left(q_x(t)+\mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c)\right) R_u(t) G\left(q_x(t)+\mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c)\right)^\mathsf{T} c} \\ & + \sqrt{c^\mathsf{T} B(q_x(t)) Q_w B(q_x(t))^\mathsf{T} c} + \sqrt{c^\mathsf{T}\Omega_n(q_x(t),Q_x(t))c}\,.\end{aligned}$$

Using the support function of the ellipsoids $\mathcal{E}(Q_x(t))$

$$V[\mathcal{E}(Q_x(t))](c) = \sqrt{c^\mathsf{T} Q_x(t) c}\,,$$

we have that Condition (4.82) is equivalent to

$$\begin{aligned}\frac{1}{2}c^\mathsf{T}\dot{Q}_x(t)c \geq \; & c^\mathsf{T} A(t) Q_x(t) c \\ & - \left\| c^\mathsf{T} G\left(q_x(t)+\mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c)\right) R_u^{\frac{1}{2}}(t)\right\|_2 \left\| Q_x^{\frac{1}{2}}(t)c\right\|_2 \\ & + \left\| Q_x^{\frac{1}{2}}(t)c\right\|_2 \left\| Q_w^{\frac{1}{2}} B(q_x(t))^\mathsf{T} c\right\|_2 \\ & + \left\| Q_x^{\frac{1}{2}}(t)c\right\|_2 \left\| \Omega_n^{\frac{1}{2}}(q_x(t),Q_x(t))c\right\|_2,\end{aligned} \tag{4.83}$$

for a.e. $t \in [t_1,t_2]$ and all $c \in \mathbb{R}^{n_x}$ with $c^\mathsf{T} c = 1$.

At this point, we use the following identities,

$$\begin{aligned}\|C_1 y\|_2 \|C_2 y\|_2 &= \max_S \; y^\mathsf{T} C_1^\mathsf{T} S C_2 y \;\text{ s.t. }\; SS^\mathsf{T} \preceq I \\ &= \inf_{\lambda>0} \frac{1}{2\lambda} y^\mathsf{T} C_1^\mathsf{T} C_1 y + \frac{\lambda}{2} y^\mathsf{T} C_2^\mathsf{T} C_2 y,\end{aligned}$$

in order to establish that (4.83) holds whenever there exist real-valued functions $\lambda, \kappa : [t_1,t_2] \to \mathbb{R}_{++}$ and a matrix-valued function $S : [t_1,t_2] \to \mathbb{R}^{n_x \times n_u}$ with $SS^\mathsf{T} \preceq I$ such that

$$\begin{aligned}\frac{1}{2}c^\mathsf{T}\dot{Q}_x(t)c \geq \; & c^\mathsf{T} A(t) Q_x(t) c \\ & - c^\mathsf{T} Q_x^{\frac{1}{2}}(t) S(t) R_u^{\frac{1}{2}}(t) G(q_x(t)+\mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c))^\mathsf{T} \\ & + \left(\frac{1}{2\lambda(t)} + \frac{1}{2\kappa(t)}\right) c^\mathsf{T} Q_x(t) c \\ & + \frac{\kappa(t)}{2} c^\mathsf{T} B(q_x(t)) Q_w B(q_x(t))^\mathsf{T} c + \frac{\lambda(t)}{2} c^\mathsf{T}\Omega_n(q_x(t),Q_x(t))c,\end{aligned} \tag{4.84}$$

for a.e. $t \in [t_1,t_2]$ and all $c \in \mathbb{R}^{n_x}$ with $c^\mathsf{T} c = 1$. In particular, condition (4.81) along with Assumption 4.8 ensure that

$$\begin{aligned}\dot{Q}_x(t) \succeq\ & A(q_x(t))Q_x(t) + Q_x(t)A(q_x(t))^\mathsf{T} \\ & + Q_x^{\frac{1}{2}}(t)S(t)R_u^{\frac{1}{2}}(t)G(\xi)^\mathsf{T} + G(\xi)R_u^{\frac{1}{2}}(t)S(t)^\mathsf{T}Q_x^{\frac{1}{2}}(t) \\ & + \left(\frac{1}{\lambda(t)} + \frac{1}{\kappa(t)}\right) Q_x(t) + \Omega_G(q_x(t),Q_x(t),S(t)) \\ & + \lambda(t)B(q_x(t))Q_w B(q_x(t))^\mathsf{T} + \kappa(t)\,\Omega_n(q_x(t),Q_x(t))\,,\end{aligned}$$

for a.e. $t \in [t_1,t_2]$ and for all $\xi \in \mathcal{E}(q_x(t),Q_x(t))$, which also implies condition (4.84) since $\mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c) \in \mathcal{E}(Q_x(t))$.

The result that $\mathcal{E}(q_x(t),Q_x(t))$ describes a RFIT on $[t_1,t_2]$ follows from Theorem 4.5. Moreover, a feedback control law for this tube is given by $\mu(t,x(t)) = \mu_t^*(\mathcal{G}_{\mathcal{E}_x(t)}(x(t) - q_x(t)))$ with

$$\begin{aligned}\mu_t^*(c) &:= \operatorname*{arg\,min}_{v \in \mathcal{E}(u_x(t),R_u(t))} c^\mathsf{T} G\left(q_x(t) + \mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c)\right) v \\ &= u_x(t) - \frac{R_u(t)G\left(q_x(t) + \mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c)\right)^\mathsf{T} c}{\left\| R_u^{\frac{1}{2}}(t)G\left(q_x(t) + \mathcal{G}^{-1}_{\mathcal{E}_x(t)}(c)\right)^\mathsf{T} c \right\|_2},\end{aligned} \tag{4.85}$$

for all $c \neq 0$, i.e. for $x(t) \neq q_x(t)$. On the other hand, $\mu(t,x(t)) = u_x(t)$ provides a natural feedback law in the case that $x(t) = q_x(t)$ by construction of (4.80).

S2 In order to show that the result also holds for general positive semidefinite matrices, we can add a small regularization term εI to the matrices Q_w, Q_u and $Q_x(t)$, and then take limits as $\varepsilon \to 0$ by invoking exactly the same continuity argument as in Step S2 of the proof of Theorem 4.5. The feedback control law is then given by $\mu(t,x(t)) = \mu_t^*(\mathcal{G}_{\mathcal{E}_x(t)}(x(t) - q_x(t)))$ with μ_t^* given by Eq. (4.85) and the Gauss map $\mathcal{G}_{\mathcal{E}_x(t)}$ given by

$$\mathcal{G}_{\mathcal{E}_x(t)}(\xi) = \frac{Q_x^\dagger(t)\xi}{\left\| Q_x^\dagger(t)\xi \right\|_2},$$

which follows from the fact that

$$\lim_{\varepsilon \to 0} (Q_x(t) + \varepsilon I)^{-1}\xi = Q_x^\dagger(t)\xi$$

for all $\xi \in \operatorname{bd}\mathcal{E}(Q_x(t)) \subseteq \operatorname{span}(Q_x(t))$. □

The following corollary is again a direct side-product of the proof of Theorems 4.5 (Step S2) and 4.6.

Corollary 4.6. *Let the set-valued function $Y : [t_1, t_2] \to \mathbb{K}_C^{n_x}$ with $Y(t) := \mathcal{E}(q_x(t), Q_x(t))$ satisfy the conditions of Theorem 4.6. Then, an explicit feedback law associated with this RFIT is given by*

$$\mu(t, x(t)) = \begin{cases} \mu_t^*(\mathcal{G}_{Y(t)}(x(t))) & \text{if } x(t) \neq q_x(t) \\ u_x(t) & \text{otherwise} \end{cases} \tag{4.86}$$

where $\mathcal{G}_{Y(t)}$ and $\mathcal{G}_{Y(t)}^{-1}$ denote the Gauss and inverse Gauss maps of $\mathcal{E}(q_x(t), Q_x(t))$ given respectively by

$$\mathcal{G}_{Y(t)}(\xi) = \frac{Q_x^\dagger(t)(\xi - q_x(t))}{\left\| Q_x^\dagger(t)(\xi - q_x(t)) \right\|_2}, \quad \text{and} \quad \mathcal{G}_{Y(t)}^{-1}(c) = q_x(t) + \frac{Q_x(t)c}{\sqrt{c^\top Q_x(t) c}},$$

and the control-minimizer map μ_t^* given by

$$\mu_t^*(c) = u_x(t) - \frac{R_u(t) G \left(\mathcal{G}_{Y(t)}^{-1}(c) \right)^\top c}{\left\| R_u^{\frac{1}{2}}(t) G \left(\mathcal{G}_{Y(t)}^{-1}(c) \right)^\top c \right\|_2}.$$

Proof. Follows directly from the proof of Theorem 4.6 Step S2 □

4.8.1 Practical Considerations for the Construction of Ellipsoidal RFITs

The application of Theorem 4.6 relies on the availability of an inner approximation of the control set shifted by nominal controller u_x through Assumption 4.6. The construction of such approximation can be done using Lemma B.4 in Appendix B. It also requires nonlinear bounders through Assumptions 4.7 and 4.8. Such bounders may be constructed either symbolically, as shown in [56], or numerically, e.g. using tools from interval analysis [148]. A difficulty with the latter, however, is that operations performed using interval arithmetic are Lipschitz continuous, yet typically nonsmooth, which would impair the use of gradient-based methods for solving the optimization problems. Instead of applying interval analysis directly, Lemma B.5 in Appendix B presents a way of constructing smooth nonlinearity bounders for twice-continuously-differentiable functions using the Frobenius norm of certain matrices.

4.9 Robust Tube-Based MPC Based on Min-Max Differential Inequalities

The main motivation for the development of methods to characterize RFITs of control systems, is to consider so-called tube-based approaches for robust model predictive control [88]. The tube MPC approach MPC approach proposed here, involves solving parametric optimization problems of the form

$$\begin{aligned} &\inf_{Y\in\mathcal{Y}} \int_t^{t+T} \ell(Y(\tau))\,\mathrm{d}\tau \\ &\text{s.t.} \quad \forall \tau \in [t,t+T], \quad Y(\tau) \subseteq F_x \\ &\qquad Y(t) = \{\hat{x}_t\}\,, \end{aligned} \tag{4.87}$$

where $\mathcal{Y}$ denotes the set of all RFITs for (4.73) on $[t,t+T]$; $\ell : \Pi(\mathbb{R}^{n_x}) \to \mathbb{R}$ is the objective of the MPC controller; the feasibility set F_x is a subset of $\mathbb{R}^{n_x}$; and $\hat{x}_t$ is the state measurement at time t, assumed to be noisefree.

Problem 4.87 may seem odd at a first sight, since we are optimizing over tubes while standard MPC formulations optimize over control functions. In this formulation, optimizing over the tube Y is equivalent to optimize over a feedback control policy μ, since by Definition 4.3 every Y is generated by at least one μ. In other words, Problem 4.87 can be written as:

$$\begin{aligned} &\inf_{x,\mu,Y} \int_t^{t+T} \ell(Y(\tau))\,\mathrm{d}\tau \\ &\text{s.t.} \quad Y \text{ is an enclosure of the system:} \\ &\qquad \dot{x}(\tau) = f(x(\tau),w(\tau)) + G(x(\tau))\mu(\tau,x(\tau)) \\ &\qquad x(t) = \hat{x}_t \quad Y(t) = \{\hat{x}_t\} \\ &\qquad \text{for } w \in \mathbb{W} \text{ and all } \tau \in [t,t+T] \\ &\qquad \forall \tau \in [t,t+T], \quad Y(\tau) \subseteq F_x\,. \end{aligned}$$

The following analysis aims to develop a tractable computational approach to addressing the tube-based robust MPC problem 4.87. For simplicity our analysis does not consider computational delays. Specifically, given any feedback control policy $\mu(t,x)$ keeping the response x in an optimal RFIT Y^* —e.g. as found from the repeated solution of (4.87) in a receding horizon manner— we assume that the control $u(t) = \mu(t,\hat{x}_t)$ is fed back into the system instantaneously.

Recall that Theorem 4.5 provides sufficient conditions for a convex set-valued function to be an RFIT for (4.73). Therefore, any solution to the following optimization problem turns out to also be a feasible solution to the tube-based MPC problem (4.87), in the case of RFITs with convex cross-sections:

$$
\begin{aligned}
\inf_{Y} \quad & \int_t^{t+T} \ell(Y(\tau))\mathrm{d}\tau \\
\text{s.t.} \quad & \text{a.e. } \tau \in [t, t+T],\ \forall c \in \mathbb{R}^{n_x}, \\
& \dot{V}[Y(\tau)](c) \geq \min_{\nu \in U} \max_{\xi, \omega} \left\{ c^\top f(\xi, \omega) + c^\top G(\xi)\nu \;\middle|\; \begin{array}{l} c^\top \xi = V[Y(t)](c) \\ \xi \in Y(t) \\ \omega \in W \end{array} \right\} \\
& Y(t) = \{\hat{x}_t\} \\
& \forall \tau \in [t, t+T], \quad Y(\tau) \subseteq F_x \,.
\end{aligned}
\tag{4.88}
$$

Notice that (4.88) is not a standard optimal control problem, as it involves semi-infinite differential inequality constraints. However, discretization of this Problem leads to a band-structured optimization problem whose complexity scales linearly with respect to the length of the time horizon.

With the results from Theorem 4.6, Problem (4.88) can be further specialized to the case of tubes with ellipsoidal cross-sections as:

$$
\begin{aligned}
\inf_{\substack{q_x, Q_x, \\ u_x, R_u, S, \\ \lambda, \kappa, \gamma}} \quad & \int_t^{t+T} \ell(\mathcal{E}(q_x(\tau), Q_x(\tau)))\mathrm{d}\tau \\
\text{s.t.} \quad & \text{a.e. } \tau \in [t, t+T], \\
& \dot{q}_x(\tau) = f(q_x(\tau), q_w) + G(q_x(\tau))u_x(\tau) \\
& \dot{Q}_x(\tau) = \Phi(q_x(\tau), Q_x(\tau), S(\tau), R_u(\tau), \lambda(\tau), \kappa(\tau)) \\
& R_u(\tau) = (1-\gamma(\tau))Q_u + \left(1 - \frac{1}{\gamma(\tau)}\right)(u_x(\tau) - q_u)(u_x(\tau) - q_u)^\top \\
& q_x(t) = \hat{x}_t, \quad Q_x(t) = 0_{n_x \times n_x} \\
& \forall \tau \in [t, t+T], \\
& Q_x(\tau) \succeq 0, \quad R_u(\tau) \succeq 0, \quad S(\tau)S(\tau)^\top \preceq I, \\
& \kappa(\tau) > 0, \quad \lambda(\tau) > 0, \quad 0 < \gamma(\tau) < 1, \\
& \mathcal{E}(q_x(\tau), Q_x(\tau)) \subseteq F_x, \quad u_x(\tau) \in \mathcal{E}(q_u, Q_u) \,.
\end{aligned}
\tag{4.89}
$$

Observe that 4.89 now, yields a standard optimal control problem with linear matrix inequality (LMI) constraints. A solution to this problem provides a RFIT in the form $Y(\tau) = \mathcal{E}(q_x(\tau), Q_x(\tau))$, from which an explicit feedback control law can be derived (See Corollary 4.6).

Practical implementation of this tube-based MPC scheme calls for the specification of the performance criterion ℓ and the feasibility set F_x. In the case of tracking control, we may use the so-called generalized rotational inertia of the set $Y(t)$ with respect to a given reference x_{ref} [52], defined by:

$$\ell(Y(t)) := \frac{\int_{Y(t)} (x - x_{\text{ref}})^\top D (x - x_{\text{ref}}) \mathrm{d}x}{\int_{Y(t)} 1 \mathrm{d}x}\,, \tag{4.90}$$

where $D \in \mathbb{S}^{n_x}_{++}$ is any weighting matrix. In the ellipsoidal case, $Y(t) := \mathcal{E}(q_x(t), Q_x(t))$, we have (see Appendix C for a derivation)

$$\ell(\mathcal{E}(q_x(t), Q_x(t))) = (q_x(t) - x_{\text{ref}})^\top D (q_x(t) - x_{\text{ref}}) + \frac{1}{n_x + 2} \operatorname{Tr}(D Q_x(t))\,. \tag{4.91}$$

Regarding the feasible set, we may consider linear state constraints of the form

$$F_x := \left\{ x \in \mathbb{R}^{n_x} \middle| h_i^\top x \leq \eta_i,\ i = 1, \ldots, n_h \right\},$$

with $h_i \in \mathbb{R}^{n_x}$ and $\eta_i \in \mathbb{R}$. In the ellipsoidal case, the feasibility constraint $\mathcal{E}(q_x(\tau), Q_x(\tau)) \subseteq F_x$ can be rewritten as [69]:

$$\forall \tau \in [t, t+T], \quad h_i^\top q_x(\tau) + \sqrt{h_i^\top Q_x(\tau) h_i^\top} \leq \eta_i\,.$$

One of the main issues in robust MPC is ensuring recursive feasibility, namely the ability to find, for every possible initial state, a feasible state at every time along the closed-loop trajectory. This requirement can be addressed by adding the following constraints to the optimization problem (4.89):

$$Y(t+T) \subseteq Y_{\text{ref}}\,, \tag{4.92}$$

where $Y_{\text{ref}} \subseteq F_x$ is a robust forward invariant set, i.e. a time-invariant RFIT. If Y_{ref} satisfies Definition 4.3 on any time interval, then the sets $\{\mu(t+T, x(t+T)) | x(t+T) \in Y(t+T)\} \in U$ will remain non-empty by construction, and the MPC procedure discussed previously is indeed recursively feasible.

In the case of the ellipsoidal tube-based MPC, the set $\mathcal{E}(x_{\text{ref}}, Q_{\text{ref}})$ is a robust forward invariant set for System (4.73) if

$$\Phi_g(x_{\text{ref}}, Q_{\text{ref}}, S_{\text{ref}}, Q_u, \lambda_{\text{ref}}, \kappa_{\text{ref}}) \preceq 0\,,$$

for some scalar $\lambda_{\text{ref}}, \kappa_{\text{ref}} \in \mathbb{R}_{++}$ and some matrix $S_{\text{ref}} \in \mathbb{R}^{n_x \times n_u}$ with $S_{\text{ref}} S_{\text{ref}}^{\mathsf{T}} \preceq I$. For instance, one such matrix Q_{ref} can be found by solving the following optimization problem:

$$\begin{aligned} \inf_{\substack{Q_{\text{ref}}, \lambda_{\text{ref}}, \\ \kappa_{\text{ref}}, S_{\text{ref}}}} \quad & \operatorname{Tr}(Q_{\text{ref}}) \\ \text{s.t.} \quad & \Phi(x_{\text{ref}}, Q_{\text{ref}}, S_{\text{ref}}, Q_u, \lambda_{\text{ref}}, \kappa_{\text{ref}}) \preceq 0 \\ & Q_{\text{ref}} \in \mathbb{S}_+^{n_x},\ \lambda_{\text{ref}},\ \kappa_{\text{ref}} > 0,\ S_{\text{ref}} S_{\text{ref}}^{\mathsf{T}} \preceq I\,. \end{aligned} \tag{4.93}$$

In the next section we present the ellipsoidal approach for tube-based MPC on a numerical case-study.

4.9.1 Numerical Case Study

We consider a benchmark spring-mass-damper system from [124],

$$\underbrace{\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix}}_{\dot{x}(t)} = \underbrace{\begin{pmatrix} x_2(t) + w_1(t) \\ -\frac{k(x)}{M} x_1(t) - \frac{h_d}{M} x_2(t) + \frac{1}{M} w_2(t) \end{pmatrix}}_{f(x(t), w(t))} + \underbrace{\begin{pmatrix} 0 \\ \frac{1}{M} \end{pmatrix}}_{G(x(t))} u(t)$$

where x_1 and x_2 denote the displacement of the cart with respect to the equilibrium position and its velocity respectively, M is the mass of the cart, $k(x) := k_0 e^{-x_1}$ is the stiffness of the spring, and h_d the damping factor. The values of the parameters are $M = 1$ kg, $k_0 = 0.33$ N/m and $h_d = 1.1$ Ns/m.

Bounds for the disturbance and the control sets are given by the centered ellipsoids $\mathcal{E}(Q_w) \in \mathbb{K}_C^2$, $\mathcal{E}(Q_u) \in \mathbb{K}_C$ with $Q_w = \operatorname{diag}(1e^{-2}\ \text{m}^2/\text{s}^2,\ 0.25\ ,\text{N}^2/\text{m}^2)$ and $Q_u = 36\ \text{N}^2$. The length of the prediction horizon has been set to $T = 10$ s and the tube based MPC controller is started at the initial state $x_{\text{start}} = (0.7, 0.7)^{\mathsf{T}}$. The setup is based on Problem 4.89 minimizing the term

$$\int_0^T \left(\|q_x(t)\|_2^2 + \frac{1}{4} \operatorname{Tr}(Q_x(t)) + u_x(t)^2 \right) \mathrm{d}t\ ,$$

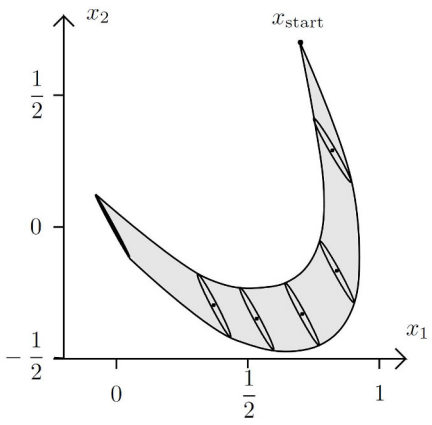

Fig. 4.5 Projection of the optimal ellipsoidal RFIT for $\hat{x}_t = x_{\text{start}}$ (grey area) onto the (x_1, x_2) space without state constraints. Selected ellipsoidal cross section are shown for $t \in \{1/4, 3/4, 5/4, 7/4, 9/4, 10\}$.

which corresponds to the generalized rotational inertia except for the term $u_x(t)^2$, which can be interpreted as a control regularization term. The nonlinearity bounder is constructed by using the technique from Appendix B. We first run the controller without state constraints, i.e. $F_x := \mathbb{R}^2$.

Problem (4.89) is solved numerically using the optimal control sofware ACADO [51][1]. Additionally, we have used a piecewise constant control discretization on 40 equidistant intervals. The numerical solution of the first optimal control problem is visualized in Figure 4.5. Notice that the center of the RFIT at time $t = 10$ is close to the set point $x^* = 0$ as expected. On the other hand the uncertainty cannot be fully compensated, as our adverse player can choose the functions w_1 and w_2 while our control function is scalar valued—thus, in our example the diameter of the predicted RFIT is not necessarily small for large t. A closed-loop simulation of the controller without disturbances indicates that the controller is stable and brings the system to its set point—there are no surprises. However, one thing that is more interesting to analyze is how the predicted RFIT changes if we add a state constraint. Figure 4.6 shows the solution of Problem 4.89 with the state constraint $\mathbb{F}_x = \{x \mid x_1 \leq 0.85\}$. Notice first that the center of the RFIT at $t = 10$ is close to the reference point $x_{\text{ref}} = 0$. Quite remarkably, as the shape of the ellipsoidal tube is an optimization that only needs to satisfy a conservative approximation of the min-max differential inequality, the tube MPC controller is able to shift and turn the ellipsoidal cross sections of the RFIT in such a way

[1] Since ACADO Toolkit does not support LMI constraints, the LMI constraints in Problem 4.89 have in our implementation been replaced by equivalent standard nonlinear state constraints using Schur complement techniques.

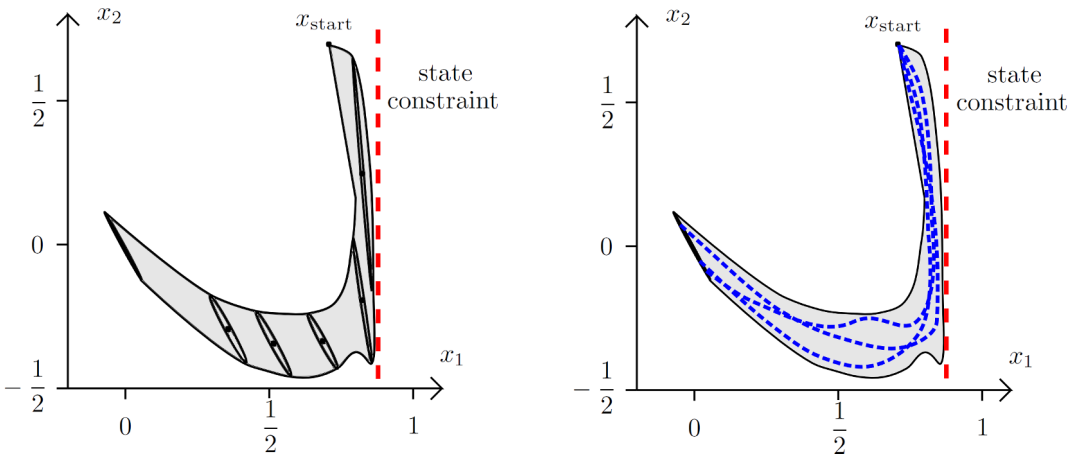

Fig. 4.6 Projection of the optimal ellipsoidal RFIT for $\hat{x}_t = x_{\text{start}}$ (grey area) onto the (x_1, x_2) space. The red line shows the state constraint $\mathbb{F}_x = \{x \mid x_1 \leq 0.85\}$. Left: Selected ellipsoidal cross section are shown for $t \in \{1/4, 3/4, 5/4, 7/4, 9/4, 10\}$. Right: Trajectories for selected uncertainty scenarios are shown in blue.

that constraint violations can be avoided for all possible uncertainty scenarios –which can be confirmed for some uncertainty scenarios in the right panel. On the other hand the uncertainty cannot be fully compensated, as our adverse player can choose the functions w_1 and w_2 while our control function is scalar valued—thus, in our example the diameter of the predicted RFIT is not necessarily small for large t.

4.10 Conclusions

This chapter has presented a unified framework for continuous-time propagation of enclosures for the reachable set of parametric nonlinear ODEs. The main contribution is the formulation of a generalized differential inequality, which provides sufficient conditions for the construction of convex enclosures. This framework has been shown to encompass the classical theory of differential inequalities as well as the ellipsoidal bounding technique. Besides being directly applicable for the construction of other types of convex enclosures, it can also be used for constructing nonconvex enclosures via its combination with Taylor models. Another principal contribution has been using this generalized differential inequality to analyze the convergence properties of various kinds of enclosures, namely interval, ellipsoidal, and Taylor model bounds. Sharp bounds on the convergence order of these methods have been obtained and illustrated with numerical examples. Finally, implementation details for the various bounding techniques have been outlined, and we recall that the

developed C++ code can be obtained freely at: http://omega-icl.bitbucket.org/cronos. The numerical case study of a six-state dynamic model of anaerobic digestion has shown the superior stability properties of nonconvex set-propagation techniques based on Taylor models with ellipsoidal remainder bounds compared to either Taylor models with differential inequalities remainder bounds or convex set-propagation techniques. On the whole, this approach appears to be well suited for use in global and robust dynamic optimization. It is worth mentioning that although not performed in this chapter, the stability analysis of discrete-time set propagation methods can be extended to the continuous-time counterpart.

The generalized differential inequality was also shown to be applicable to control-affine nonlinear systems for the characterization of robust forward invariant tubes in the presence of uncertainties. For this purpose a formulation based on a min-max differential inequality has been introduced, which provides sufficient conditions for a time-varying convex set-valued function to be a RFIT for control-affine nonlinear systems. This min-max DI has also been used to propose a tube-based MPC procedure. Unlike other robust MPC approaches the procedure based on the min-max DI does not use any particular parameterization for the feedback control law, while benefiting from having linear complexity with respect to the time horizon. Another benefit of the proposed approach is that a semi-explicit representation of a feedback control law is obtained as a side-product of the RFIT computation whenever the tube and control sets have smooth boundaries with positive curvature. In particular, this approach has been used to derive a practical implementation tubes tubes with ellipsoidal cross-sections. This was tested on a spring-mass-damper system using the optimal control software ACADO. Future work will deal with the derivation of tractable implementation for other convex sets such as interval boxes and polytopes. Another interesting direction is the use of the min-max DI in order to derive RFIT around a given nominal trajectory and use the semi-explicit feedback law in order to transform a nominal optimal open-loop controller into a robust closed-loop robust controller.

Chapter 5

Constraint Projection

This chapter is concerned with the construction of enclosures for sets defined implicitly by systems of constraints. In particular we address two different problems in set-membership (or guaranteed) estimation and analysis of mathematical models. The first problem is guaranteed parameter estimation, which deals with the problem of computing the set of parameters (in a given domain of interest) whose model output is consistent with a given set of measurements within a prespecified error. Clearly this is a set defined implicitly by constraints that depend on the solution of parametric ODEs. The second problem—guaranteed asymptotic analysis— is concerned with the enclosure of the set of fixed-points (or equilibrium manifold) of an ODE system, the characterization of its stability and the determination of its bifurcation points. In this case, the set is defined implicitly by systems of nonlinear algebraic equations and inequalities. These problems are addressed within the constraint projection framework, which is concerned with the exhaustion of elements of a set satisfying a system of constraints. We provide a branch-and-prune algorithm, which converges in a finite number of iterations and whose output is the enclosure of the implicit set to a prescribed tolerance. The algorithm requires the construction of convergent enclosures for the constraint system, in particular we focus on methods with higher-order convergence properties (i.e. polynomial models). The performance of the algorithm is enhanced by the use of optimization-based domain reduction based on polyhedral relaxations for polynomial models. A domain-reduction strategy in reduced space based on Newton-like methods for polynomial models is also introduced for constraint systems which contain a system of underdetermined nonlinear algebraic equations. This method is capable of mitigating the clustering effect inherent to constraint projection problems. The performance of the algorithm is illustrated with three case studies: Guaranteed parameter estimation and guaranteed

asymptotic analysis for an anaerobic digestion model and guaranteed asymptotic analysis for a nutrient-resource-consumer model.

The remainder of this chapter is organized as follows. Section 5.1 introduces the problems of guaranteed parameter estimation and guaranteed asymptotic analysis in the context of constraint projection. Section 5.2 introduces the branch-and-prune algorithm for constraint projection together with a detailed explanation of the bounding strategy and domain reduction and CPU-time reduction strategies. Section 5.3 presents a brief analysis on the clustering effect for constraint projection problems and a novel domain reduction strategy in reduced space. Section 5.4 illustrates the effects of the enhancements of the branch-and-prune algorithm applied to the guaranteed parameter estimation of an anaerobic digestion model. Section 5.5 presents the constraint projection approach applied to the guaranteed asymptotic analysis for the anaerobic digestion model, while Section 5.6 illustrates the guaranteed asymptotic analysis of a nutrient-resource-consumer model. Finally, Section 5.7 concludes the chapter.

5.1 Problem Definition

The focus of this chapter is on the characterization of implicitly defined sets, such as

$$Z_{g,Z_0,\Gamma} := \{ z \in Z_0 \mid g(z) \in \Gamma \} . \tag{5.1}$$

The set $Z_{g,Z_0,\Gamma}$ defines the preimage of $\Gamma \subset \mathbb{R}^{n_g}$ under a function $g : \mathbb{R}^{n_z} \to \mathbb{R}^{n_g}$ on a domain of interest $Z_0 \in \mathbb{K}^{n_z}$. As a small abuse of notation, we will often write Z_g to denote $Z_{g,Z_0,\Gamma}$, when the domain Z_0 and the set Γ are clear from the context.

An exact or analytical characterization for sets of the form (5.1) is not possible in general. Our focus here is on constraint projection methods, which allow computing enclosures of $\mathscr{Z}_{g,Z_0,\Gamma}$ using subpavings i.e. a collection of non overlapping boxes Z arranged in partitions $\mathbb{Z}_{\text{in}}$ and $\mathbb{Z}_{\text{bnd}}$ satisfying

$$\bigcup_{Z \in \mathbb{Z}_{\text{in}}} Z \subseteq Z_{g,Y_0,\Gamma} \subseteq \bigcup_{Z \in \mathbb{Z}_{\text{in}} \cup \mathbb{Z}_{\text{bnd}}} Z , \tag{5.2}$$

and $\mathbb{Z}_{\text{bnd}}$ being sufficiently small.

Characterization of implicitly defined sets occurs often in the analysis of processes and systems, e.g. constraint satisfaction problems [103, 57]. In this chapter we are concerned with the application of constraint projection to the analysis of parametric dynamic systems.

In particular we focus on two specific problems: guaranteed parameter estimation and guaranteed asymptotic analysis of parametric dynamic systems.

5.1.1 Guaranteed Parameter Estimation

Consider the following dynamic system

$$\forall t \in [0,T] : \dot{x}(t,p) = f(x(t,p),p) \quad \text{with} \quad x(0,p) = x_0(p), \tag{5.3a}$$

$$y(t,p) = h(x(t,p),p), \tag{5.3b}$$

where $x : [0,t_N] \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$ denotes the vector of process states, $p \in R^{n_p}$ stands for the vector of (unknown) process parameters, and $y : [0,t_N] \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_y}$ denotes the n_y-dimensional vector of model outputs (predictions). Associated to this model, we also have the pointwise in time reachable sets introduced in Chapters 3 and 4 for a given set of parameters $P \in \mathbb{K}^{n_p}$

$$X(t,P) := \{x(t,p) \mid p \in P\} \quad \text{and} \quad Y(t,P) := \{y(t,p) \mid p \in P\} .$$

For a given set of output measurements $y_{\mathrm{m}}(t_i)$ at N time points $t_1,\ldots,t_i,\ldots,t_N$, *classical* parameter estimation seeks for *one* particular instance p_{e} of the parameter values for which the (possibly weighted) normed difference between these measurements and the corresponding model outputs y is minimized. This optimization problem, for instance in the least-square sense, is given by:

$$\begin{gathered} p_{\mathrm{e}} \in \arg\min_{p \in P_0} \sum_{i=1}^{N} \|y_{\mathrm{m}}(t_i) - y(t_i,p)\|_2^2 \\ \text{s.t. Equations (5.3)} \end{gathered} \tag{5.4}$$

where $P_0 := [p_0^L, p_0^U]$ denotes the a priori set of admissible values for the parameters.

In contrast, *guaranteed* (bounded-error) parameter estimation accounts for the fact that the actual process outputs, y_{p}, are only known within some bounded measurement error $e \in E := [e^L, e^U]$, so that

$$y_{\mathrm{p}}(t_i) \in y_{\mathrm{m}}(t_i) + [e^L, e^U] =: Y_{\mathrm{p}}(t_i) . \tag{5.5}$$

Then, the main objective is to estimate the set P_e of *all* possible parameter values p such that $y(t_i;p) \in Y_p(t_i)$ for every $i = 1,\ldots,N$; that is,

$$P_{h,P_0,Y_p} := \left\{ p \in P_0 \;\middle|\; \begin{array}{ll} \exists x \text{ such that:} & \\ \dot{x}(t,p) = f(x(t,p),p) & \text{with} \quad x(0,p) = x_0(p), \\ h(x(t_i,p),p) \in Y_p(t_i), & i = 1,\ldots,N \end{array} \right\}. \tag{5.6}$$

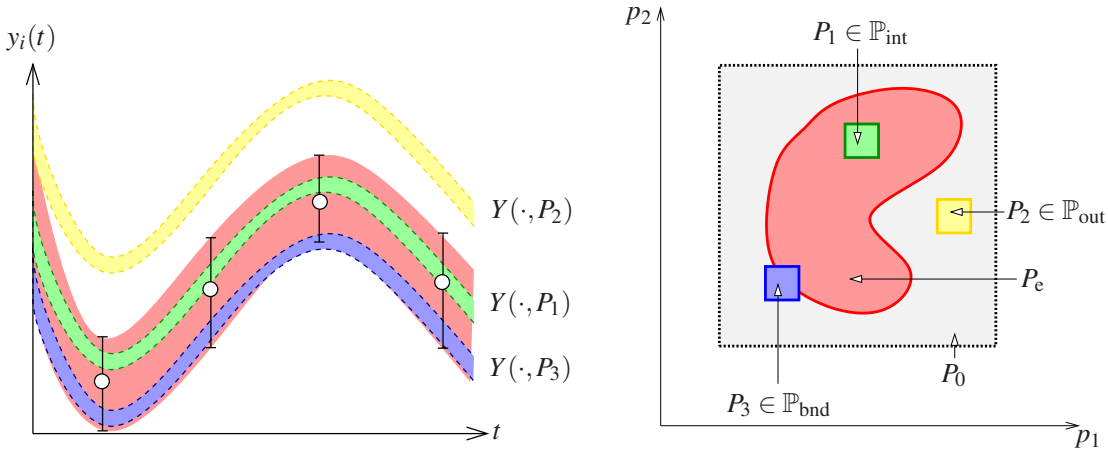

Fig. 5.1 Illustration of guaranteed parameter estimation concepts in the space of output trajectories (left plot) and in the parameter space (right plot).

Depicted in red on the left plot in Fig. 5.1 is the set of all output trajectories satisfying $y(t_i,p) \in Y_p(t_i)$ with $i = 1,\ldots,N$, and on the right plot the corresponding set P_{h,P_0,Y_p} projected onto the (p_1,p_2) space.

5.1.2 Guaranteed Asymptotic Analysis of Parametric Dynamic Systems

Asymptotic analysis refers to the characterization of the equilibrium —and equilibrium-related phenomena— of a dynamic system. Consider again the dynamic system given by Equation (5.3a), the equilibrium manifold of this dynamic system is given by

$$Y_{f,Y_0,\{0\}} = \{ y \in Y_0 \mid f(y) \in \{0\} \}\ , \tag{5.7}$$

where $y = (x,p)^\mathsf{T}$ denotes the vector of stacked states and parameters. This problem requires finding all the solutions of a system of nonlinear algebraic equations in the domain

Y_0. In the sequel we will only refer to under-determined system of equations, since the fully determined (all parameters are known) can be treated in an analogous manner.

Once the equilibrium manifold is known, we can also characterize it via the local stability of its elements. Recall that a dynamic system is stable if and only if the determinant of its n_x-by-n_x Hurwitz matrix,

$$H(x,p) := \begin{pmatrix} c_1(x,p) & c_3(x,p) & c_5(x,p) & \cdots & 0 \\ c_0(x,p) & c_2(x,p) & c_4(x,p) & \cdots & 0 \\ 0 & c_1(x,p) & c_3(x,p) & \cdots & 0 \\ 0 & c_0(x,p) & c_2(x,p) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & c_{n_x}(x,p) \end{pmatrix}, \tag{5.8}$$

and all its leading principal minors are positive, where $c_i(x,p)$ are the coefficients of the characteristic polynomial of the Jacobian matrix $J_f := \frac{\partial f}{\partial x}$,

$$\mathrm{Det}(J_f(x,p) - \lambda I) =: \lambda^{n_x} + \sum_{i=1}^{n_x} c_i(x,p)\lambda^{n_x - i}.$$

In practice, symbolic expressions for the coefficients c_i can be obtained using the Faddeev-Leverrier algorithm [46].

For under-determined systems, the equilibrium manifold is parameterized by the value of p and variations in the parameters may change the location, stability and number of the equilibrium points. We call a point $(x,p) \in Y_f$ a local bifurcation if in a small neighborhood of that point changes in the number of equilibria or its local stability occur. Bifurcation analysis is tool widely used in control and process design and it is typically done via numerical continuation methods.

In this chapter, we will focus on two different types of local bifurcations, namely steady-state and Hopf bifurcations. Steady-state bifurcations are associated with changes in the number of equilibria, i.e. creation or collapse of branches in the equilibrium manifold. Steady-state bifurcations occur when the determinant of the Jacobian matrix is singular.

On the other side, Hopf bifurcations are associated with changes in the stability in the equilibrium manifold. Hopf bifurcations typically occur in the presence of limit cycles and their occurrence induces or destroys oscillatory behavior in dynamic systems. Hopf bifurcation occurs when the characteristic polynomial of the Jacobian matrix has a conjugate pair of purely imaginary eigenvalues while the other roots have negative real part. The above

condition can be expressed algebraically as

$$c_{n_x} > 0,\ \Delta_{n_x-1} = 0,\ \Delta_{n_x-2} > 0,\ \ldots, \Delta_1 > 0, \tag{5.9}$$

with Δ_i the ith principal minor of the Hurwitz matrix H (Eq. (5.8)) [60].

5.2 Constraint Projection Algorithm

Algorithm 2 shows the B&P procedure which is at the core of the methods presented in this chapter. It is worth noticing that the most basic version of a branch and prune algorithm requires a convergent procedure capable of computing an enclosure of $\{g(z)|z \in Z\}$ —the image of a given subpartition $Z \in Z_0$ under the function g (Step 3), the exclusion tests given in Step 4 and a termination criteria $\varepsilon_{\text{box}} \geq 0$. With these three ingredients, Jaulin and Walter [59] has shown that the algorithm terminates in a finite number of iterations. The number of iterations for such a simple B&P algorithm is bounded by $(\text{diam}(Z_0)/\varepsilon_{\text{box}} + 1)^{n_z}$ [59].

In the remainder of this section we present the details of the different steps on the algorithm.

5.2.1 Bounding Strategies

In Step 3 of Algorithm 2 a procedure is called to construct a set-valued function $\bar{g} : \mathbb{K}^{n_z} \to \mathbb{K}^{n_g}$ bounding the image of Z under g, i.e. satisfying $\bar{g}(Z) \supseteq \{g(z)|z \in Z\}$.

The main requirement of this step is for the set-valued function $\bar{g}$ to be Hausdorff convergent, namely if $\text{diam}(Z) \to 0$ then $d_{\text{H}}(\{g(z) \mid z \in Z\}, \bar{g}(Z)) \to 0$. This in turn is needed to ensure the convergence of the algorithm in a finite number of iterations [59]. From a practical perspective, we are interested in set-valued functions with higher order convergence properties, since they provide tighter enclosures and hence it may reduce the number of iterations needed .

If the function is factorable e.g., in the case of computing the solution manifold of a system of implicit equations, the methods based on set-valued arithmetics presented in Chapter 2 can be used. For guaranteed parameter estimation, the measurement function h is factorable, but depends on the solution of an ODE, which is not factorable. In particular Step 3 involves computing enclosures of the reachable set $X(t_i, P)$ with $1 \in \{1, \ldots, N\}$ for a given P in the boundary partition of $P_{h,P_0,Y_{\text{p}}}$. In this case, the methods developed in Chapters 3 and 4 can be used.

Input: Termination tolerances $\varepsilon_{\text{box}} \geq 0$ and $\varepsilon_{\text{bnd}} \geq 0$, initial domain Z_0 and image constraint set Γ.

Initialization: Set partitions $\mathbb{Z}_{\text{bnd}} = \{Z_0\}$, $\mathbb{Z}_{\text{in}} = \emptyset$ and counter $k = 0$

Main Loop:

1. Select a box $Z \in \mathbb{Z}_{\text{bnd}}$ and remove it from $\mathbb{Z}_{\text{bnd}}$
2. Apply domain reduction to Z
3. Compute $\overline{g}(Z) \supseteq \{g(z) \mid z \in Z\}$
4. Exclusion Tests:
   (a) **If** $\overline{g}(Z) \subseteq \Gamma$, insert Z into $\mathbb{Z}_{\text{in}}$
   (b) **Else if** $\overline{g}(Z) \cap \Gamma = \emptyset$, fathom Z
   (c) **Else** bisect Z and insert subsets back into $\mathbb{Z}_{\text{bnd}}$
5. Termination Tests:
   (a) **If** $\text{width}(Z) \leq \varepsilon_{\text{box}}$ for all $Z \in \mathbb{Z}_{\text{bnd}}$, **stop**
   (b) **If** $V_{\text{bnd}} := \sum_{Z \in \mathbb{Z}_{\text{bnd}}} \text{volume}(Z) \leq \varepsilon_{\text{bnd}}$, **stop**
6. Increment counter k+=1; **Return** to step 1

Output: Partitions $\mathbb{Z}_{\text{in}}$ and $\mathbb{Z}_{\text{bnd}}$; Iteration count k

Algorithm 2 Branch and Prune

A comparison between different continuous-time enclosure strategies in the context of guaranteed parameter estimation is presented next for a simple case study.

Example 5.1. *Consider the following dynamic model involving two state variables* $x = (x_1, x_2)^{\mathrm{T}}$ *and three uncertain parameters* $p = (p_1, p_2, p_3)^{\mathrm{T}} \in [0.01, 1]^3$ *[61]:*

$$\dot{x}_1(t) = -(p_1 + p_3)x_1(t) + p_2 x_2(t) \qquad \textit{with} \quad x_1(0) = 1, \tag{5.10a}$$

$$\dot{x}_2(t) = p_1 x_1(t) - p_2 x_2(t) \qquad \textit{with} \quad x_2(0) = 0. \tag{5.10b}$$

This system has a single output variable y, which corresponds to the state variable x_2*,* $y(t,p) := x_2(t,p)$*, with* $N = 15$ *measurements corresponding to the time instants* $t_i = 1, \ldots, 15$*. Synthetic experimental data are generated by simulating the model* (5.10) *with parameter values* $p^* = (0.6, 0.15, 0.35)^{\mathrm{T}}$*, and then rounding the output* $y(t_i)$ *up or down to the nearest value by retaining two significant digits only; then, measurement error ranges of* $\pm 5 \times 10^{-3}$ *are added around these values.*

The constraint projection algorithm (Sect. 2) is implemented in a C++ program that uses the library MC++ for computations involving Taylor models. Moreover, the code calls the ODE integration methods in the GNU Scientific Library (GSL) to bound the parametric ODEs based on the techniques presented in Chapter 4. All the numerical results presented subsequently use the explicit embedded Runge-Kutta-Fehlberg (4,5) method, with both relative and absolute tolerances set to 10^{-7}*, and are obtained on a workstation with Intel Core i7-3770 processors at 3.40 GHz and running 64-bit Linux.*

The performance of guaranteed parameter estimation is investigated for continuous-time ODE bounding techniques propagating Taylor models of orders $q = 1,\ldots,4$ *with interval or ellipsoidal remainders and compared to standard differential inequalities. In order to allow for fair comparisons, the termination criterion is defined in terms of the level of accuracy* ε_{bnd} *of the solution set (Test 5b) in the range* $10^{-3} \rightarrow 5\times10^{-6}$*—the termination criterion in terms of the minimum box size* ε_{box} *(Test 5a) is set to zero, on the other hand. The results are shown in Fig. 5.2 in terms of the number of iterations (left plot) and CPU time (right plot).*

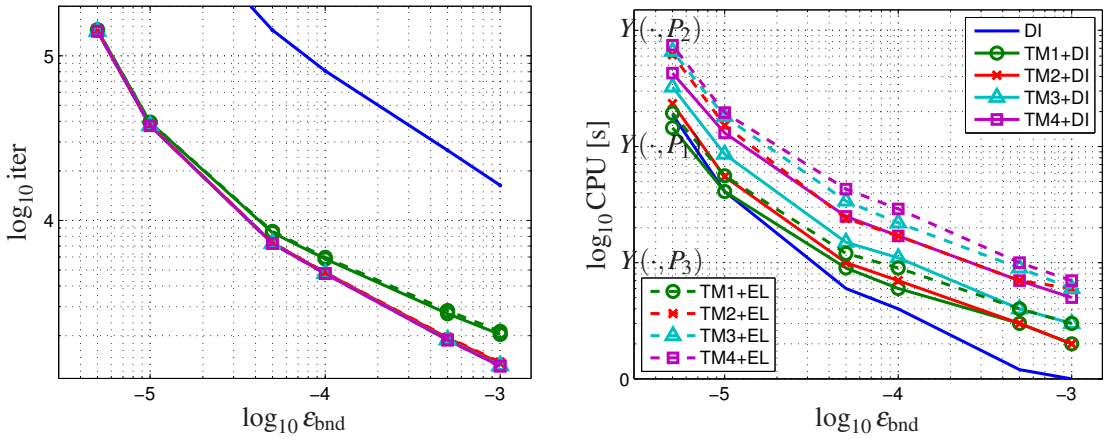

Fig. 5.2 Performance of guaranteed parameter estimation using various ODE bounding techniques in the set-inversion algorithm. Left: Number of iterations vs. convergence threshold. Right: CPU time vs. convergence threshold.

It is evident that classical differential inequalities (DI) require by far the largest number of iterations, at any accuracy level. For accuracies of $\varepsilon_{\text{bnd}} = 10^{-5}$ *and* $\varepsilon_{\text{bnd}} = 5\times10^{-6}$*, respectively 932, 454 and 3,612,968 iterations are needed. Memory storage of such a high number of parameter boxes during the course of the algorithm can become a serious issue with an increasing number of uncertain parameters, calling for less conservative bounding techniques. Despite the large number of iterations however, this approach allows for the fastest computations for accuracies down to* $\varepsilon_{\text{bnd}} \approx 10^{-5}$ *due to its simplicity. At higher*

accuracy levels, bounding techniques based on Taylor models are seen to exhibit faster convergence as the extra computational burden of these higher-order bounding techniques is overpowered by a dramatic reduction in overall number of iterations (more than an order of magnitude). The shortest run-time is obtained with first-order Taylor model with interval remainder bounds (labelled TM1+DI *in Fig. 5.2) for* $\varepsilon_{bnd} < 10^{-5}$ *here.*

In terms of overall number of iterations, the performance of the set-inversion algorithm between first-order Taylor models (both variants TM1+DI *and* TM1+EL*), on the one hand, and between all Taylor models of second-, third- and fourth-order (both variants* TM2+DI, TM3+DI, TM4+DI *and* TM2+EL, TM3+EL, TM4+EL*), on the other hand, is about the same. The lower performance of first-order Taylor models compared to higher-order Taylor models can be attributed to the fact that first-order Taylor models compute convex enclosures and are thus limited for the approximation of (potentially) nonconvex reachable sets. In terms of the overall run-time though, first-order Taylor models with interval remainders are found to outperform the other bounding techniques based on higher-order Taylor models in this case study. Bearing in mind the trade-off between a smaller number of iterations and a larger processing time needed for a single iteration, it is expected that higher-order bounding techniques will become advantageous for dynamic models of higher complexity or with more uncertain parameters nonetheless.*

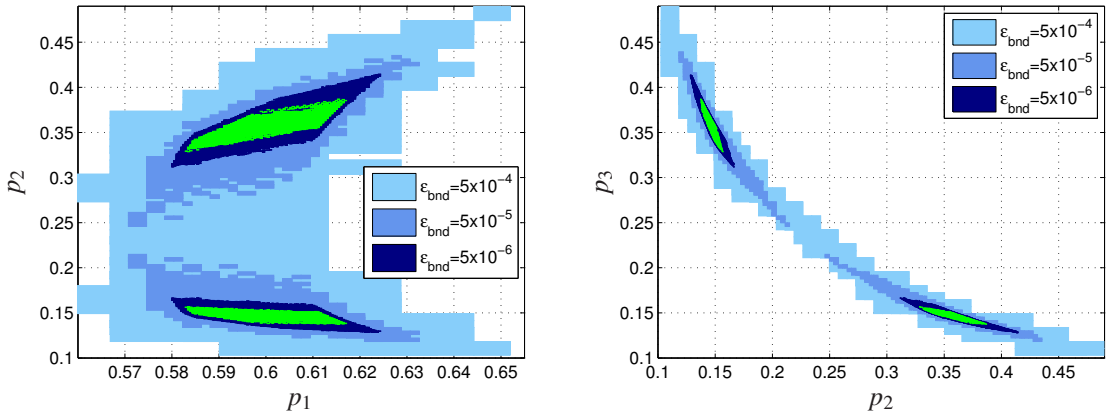

Fig. 5.3 Outer approximations of the sets of guaranteed parameter estimates for different levels of accuracy ε_{bnd}. Inner approximation of the set of guaranteed parameter estimates for $\varepsilon_{\text{bnd}} = 5 \times 10^{-6}$ plotted in green. Left: Projections onto (p_1, p_2) space. Right: Projections onto (p_2, p_3) space.

Finally, the left and right plots in Fig. 5.3 show projections of the approximate solution sets—$\mathbb{P}_{int}$ *and* $\mathbb{P}_{bnd}$ *are shown using the same color scheme as in Fig. 5.1 above—onto the* (p_1, p_2) *and* (p_2, p_3) *subspaces, respectively, for different levels of accuracy* ε_{bnd}*. Observe first that the 'true' parameter values* p^* *used to generate the pseudo-experimental data are*

part of the solution set, for each reported level ε_{bnd}. Moreover, the solution set for this problem turns out to be disconnected, thus suggesting a possible structural identifiability problem. Interestingly, this non-connectedness of the solution set can only be detected when the accuracy level ε_{bnd} is already lower than 5×10^{-5}. This clearly supports the need for developing strategies that can accelerate the convergence of the set-inverse algorithm, such as box-reduction and other CPU-time-reduction approaches.

5.2.2 Optimization-Based Domain Reduction and CPU Reduction Strategies

Step 2 of the B&P algorithm is an improvement over the basic algorithm. The main idea behind domain reduction is to use constraints on the variables in order to exclude parts of a domain Y where these restrictions are violated. This idea was used in Lin and Stadtherr [80], Kletting et al. [63] to contract the solution space by deriving contracted bounds for the variables using the linear part of Taylor models.

The following optimization-based procedure is also introduce in order to reduce the domain at a particular iteration. This procedure makes use of the dependencies captured by the polynomial model.

Consider a box $Z := [z^{\mathrm{L}}, z^{\mathrm{U}}] \in \mathbb{IR}^{n_z}$ and a q-th order polynomial model $\mathcal{T}^q_{g,Z} = (\mathcal{P}^q_{g,Z}, \mathcal{R}^q_{g,Z})$ of g on Z. The interval Z can be tightened by solving optimization problems of the form:

$$z_j^{\mathrm{L/U}} = \min_z / \max_z z_j \quad \text{s.t.} \quad \mathcal{P}^q_{g,Z}(z) \in Z \ominus \mathcal{R}^q_{g,Z} \,. \tag{5.11}$$

With this procedure, the size of the interval vector can be reduced by solving $2n_z$ optimization problems. As stated, Equation (5.11) is a nonconvex optimization problem for $q \geq 2$. In order to reduce the complexity of these problems, affine relaxations of the model have to be used. One way of computing such relaxations is by considering only the linear part of the model, bounding the nonlinear terms and adding them to the remainder term. Another method consists in computing polyhedral relaxations for the polynomial (cf. Chapter 2). These relaxations allow us to exploit the robustness, efficiency and speed of state-of-the-art LP solvers such as GUROBI or CPLEX.

We also note that further improvements could be obtained by tightening the relaxations, for instance by exploiting intermediate substructures in the factored optimization problem [159, 90]. Moreover, in case the reduction of a box Z is larger than a given threshold, for instance $\geq 20\%$ in volume, it can be repeated multiple times. It is important to bear in mind that repeating the reduction several times requires recomputing the enclosures $\overline{g}(Z)$ of

the function g on the reduced box Z though. This defines a clear trade-off between the extra computational burden and the reduction in the size of the partition $\mathbb{Z}_{\text{bnd}}$, which is of course problem-dependent.

When combined with domain-reduction techniques, polynomial models can improve the convergence speed of the set-inversion algorithm significantly. But because polynomial models can also cause a large computational overhead, this benefit is mostly noticeable at an early stage of the constraint projection procedure, when many boxes can be fathomed or greatly reduced. This calls for further CPU-time-reduction strategies in order to make constraint projection more competitive for high-order polynomial models.

In the basic constraint projection algorithm, the enclosures $\overline{g}(Z)$ are recomputed at every iteration because of the overestimation inherent to the bounding techniques. In this context, a simple CPU-time-reduction strategy involves reusing the enclosures computed at a parent node (i.e., for a larger parameter box Z) as soon as the overestimation of the enclosure has become smaller than a given threshold $\varepsilon_{\text{cvg}} > 0$. For polynomial models, such overestimation is directly measured by the remainder term $\mathcal{R}^q_{g,Z}$, and a possible re-usability condition thus reads

$$\operatorname{diam}(\mathcal{R}^q_{g,Z}) \leq \varepsilon_{\text{cvg}} \,. \tag{5.12}$$

As soon as this condition is met, the corresponding polynomial model $(\mathcal{P}^q_{g,Z}, \mathcal{R}^q_{g,Z})$ can indeed be stored and used later on in any child node $Z' \subseteq Z$, effectively by-passing the bounding step 3. Variants of this approach can of course be used that consider relative convergence criteria and scaling for instance.

In addition to reusing polynomial models at children nodes, a further CPU-time-reduction strategy involves reducing the order of the polynomial models, which can lead to significant savings in connection to the relaxation and solution of the optimization-based domain-reduction problems (5.11). Consider without loss of generality the polynomial part expressed as

$$\mathcal{P}^q_{g,Z}(z) = \sum_{\substack{\gamma \in \mathbb{N}^{n_z}, \\ |\gamma| \leq q}} c_\gamma z^\gamma \,,$$

a simple order-reduction procedure is as follows:

In particular, bounding the ρ-th terms can be done using any of the bounding functions mentioned in Chapter 2 (cf. Sec. 2.2.1). Regarding the convergence threshold ε_{cvg} finally, notice that a larger threshold will lead to reusing polynomial models from parent nodes earlier as well as reducing their order faster, but too large a threshold can prevent convergence

Input: Convergence threshold $\varepsilon_{\text{cvg}} > 0$; domain box Z; qth-order polynomial model $(\mathcal{P}^q_{g,Z}, \mathcal{R}^q_{g,Z})$ of g on Z satisfying (5.12)

Initialization: Set reduced order $\rho = q$

Main Loop:

1. Compute bounds for ρ-th terms:

$$\mathcal{B}^\rho \supseteq \left\{ \sum_{\substack{\gamma \in \mathbb{N}^{n_z}, \\ |\gamma| = \rho}} c_\gamma z^\gamma \;\middle|\; z \in Z \right\}$$

2. **If** $\operatorname{diam}(\mathcal{B}^\rho) > \varepsilon_{\text{cvg}}$, **stop**
3. Reduce order ρ-=1; **Return** to step 1

Output: Reduced Taylor model order ρ

Algorithm 3 Order-reduction for polynomial models

of the set-inversion algorithm if the stopping criterion is based solely on the total volume threshold ε_{bnd} (Step 5b).

Example 5.1 (continued)**.** *We continue the case study of the dynamic system* (5.10) *in order to investigate the effect of optimization-based domain reduction and CPU-time reduction. Guaranteed parameter estimation is applied with and without the use of domain reduction as an extra step in the set-inversion algorithm (reduction threshold of 20% and maximum of 10 reduction loops at each iteration). Taylor models of orders $q = 1, \ldots, 4$ are considered for enclosing the output reachable set $\overline{Y}(\cdot, P)$, and the termination criteria remain the same as defined previously.*

The number of iterations and the CPU time required by the set-inversion algorithm to terminate with different Taylor model orders and with or without the use of domain reduction are reported on the left and right plots of Fig. 5.4, respectively, as a function of the termination tolerance ε_{bnd}. It is evident that the number of iterations decreases significantly when domain reduction is used—here by at least one order of magnitude for all considered tolerance levels ε_{bnd}. Moreover, the higher the order of the Taylor model, the smaller the number of iterations required by the algorithm to converge for a given accuracy level. In terms of overall CPU time, the use of domain reduction is found to be mostly beneficial at an

early stage of the set-inversion procedure, where many boxes can be significantly reduced or even eliminated using optimization-based domain reduction.

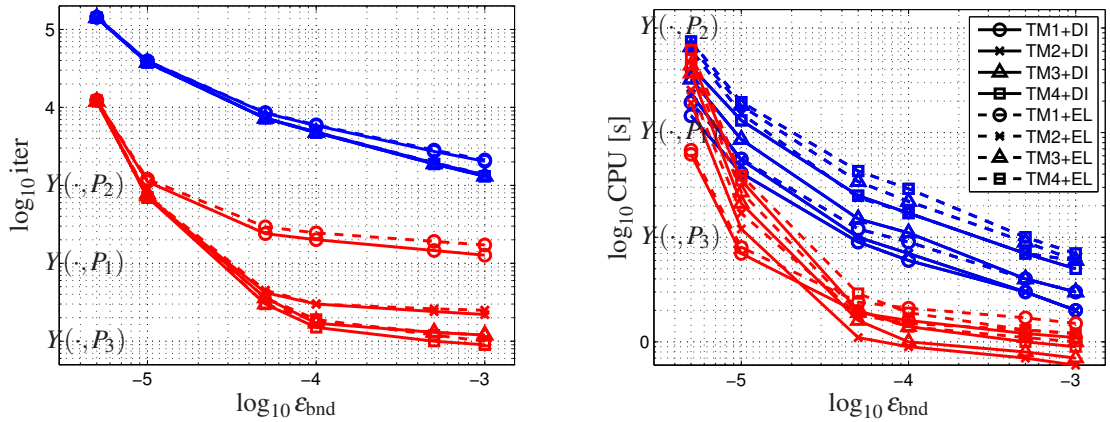

Fig. 5.4 Performance of guaranteed parameter estimation with (red lines) and without (blue lines) the use of domain reduction and with ODE bounding techniques based on Taylor models of orders $q = 1,\ldots,4$. Left: Number of iterations vs. convergence threshold. Right: CPU time vs. convergence threshold.

A certain trade-off is observed in terms of CPU time on the right plot of Fig. 5.4, whereby higher-order Taylor models can cause a significant computational overhead. On the whole, first- or second-order Taylor models with interval remainder bounds are found to enable the fastest computations in this case study. Although higher-order Taylor models reduce the overestimation, lower-order Taylor models eventually become computationally advantageous as the parameter boxes shrink. Another trade-off is observed in terms of the overhead caused by the application of domain reduction (construction and solution of relaxed LP problems). These trends show a clear need for CPU-time-reduction strategies in connection to Taylor model-based ODE bounding. Nonetheless, when used in combination with domain reduction, Taylor model-based ODE bounders now greatly outperforms classical differential inequalities (see Fig. 5.2).

The plots in Fig. 5.5 show the projections of the solution set outer-approximation onto the (p_1,p_2) and (p_2,p_3) subspaces, for increasing accuracy levels of $\varepsilon_{bnd} = 5\times10^{-4}$, 5×10^{-5}, and 5×10^{-6}, using optimization-based domain reduction and second-order Taylor models with interval remainder terms for ODE bounding. In comparing outer-approximations of the guaranteed parameter set P_e for various accuracy levels, it is found that setting $\varepsilon_{bnd} = 5\times10^{-5}$ already provides a tight approximation of P_e, with only 34 boxes and a run-time of about 2 sec.

As expected, a much tighter approximation is obtained by setting $\varepsilon_{bnd} = 5 \times 10^{-6}$*, yet this is at the price of a much finer box partition comprising 11,250 boxes here and a corresponding run-time of over* 60 sec. *For the sake of comparison we also note that, when no domain reduction is used, the partition comprises over 2,200 boxes with* $\varepsilon_{bnd} = 5 \times 10^{-5}$ *and over 70,000 boxes with* $\varepsilon_{bnd} = 5 \times 10^{-6}$*. These results also suggest that the efficiency of the set-inversion algorithm in computing highly-accurate set approximations could be improved significantly if affine cuts were enabled in addition to simple bounds contraction during the domain-reduction procedure. Such cuts would provide the extra flexibility needed to closely approximate the actual parameter set and will be the topic of future research.*

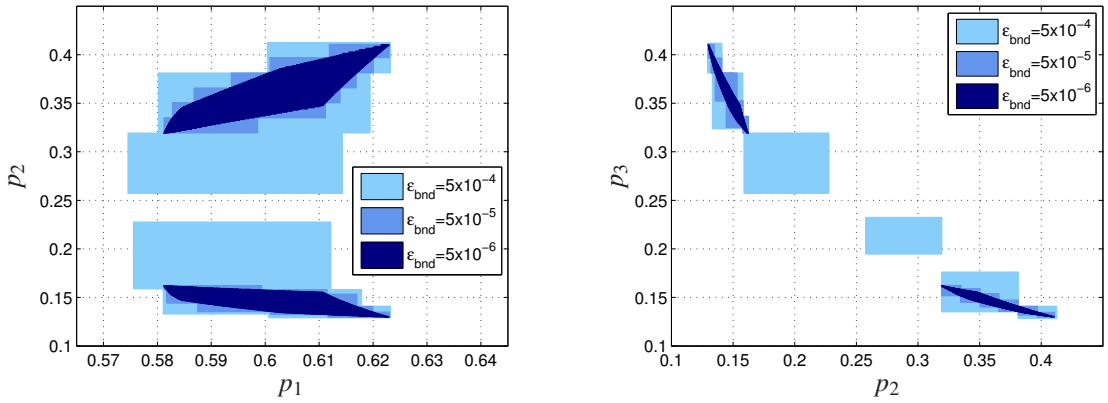

Fig. 5.5 Outer approximations of the sets of guaranteed parameter estimates for different levels of accuracy ε_{bnd} using domain reduction and second-order Taylor models. Left: Projections onto (p_1, p_2) space. Right: Projections onto (p_2, p_3) space.

Finally, we investigate the effect of CPU-time reduction, by considering both strategies of reusing and reducing the order of Taylor models computed at parent nodes. A convergence threshold of $\varepsilon_{\text{cvg}} = 10^{-4}$ *(determined heuristically) is used here.*

Computational time requirements for the set-inversion algorithm to converge are shown in Fig. 5.6 for various termination tolerances ε_{bnd}*. Not reported on this plot are the CPU times for first-order Taylor models since the corresponding improvement is marginal—convergence of first-order Taylor models within* $\varepsilon_{\text{cvg}} = 10^{-4}$ *is only achieved for very small parameter boxes in this case. For higher-order Taylor models, it is evident that the CPU-time-reduction strategies are effective. The best performance is achieved when second-order Taylor models with interval remainder bounds are used, but third- and fourth-order Taylor models lead to comparable run-times nonetheless.*

With all the proposed improvements used together, guaranteed parameter estimation of the dynamic system (5.10) *can be solved to within* $\varepsilon_{bnd} = 5 \times 10^{-6}$ *in less than 60 seconds. This is a three-fold reduction compared to the classical method of differential inequalities.*

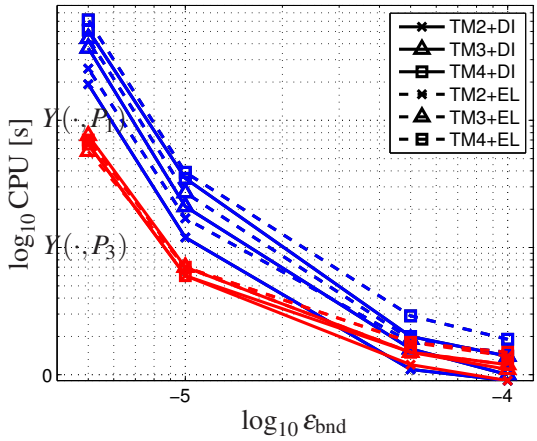

Fig. 5.6 Performance of guaranteed parameter estimation with (red lines) and without (blue lines) CPU-time-reduction strategies, in combination with domain-reduction strategy and ODE bounding techniques based on Taylor models of orders $q = 2, \ldots, 4$: CPU time vs. convergence threshold.

5.3 Domain Reduction in Reduced Space

One of the biggest problems arising when solving a constraint projection problem is the accumulation of boxes in the boundary of the set in order to meet the pre-specified approximation tolerance. In complete methods for global optimization, e.g. those based on branch and bound algorithms, this problem is known as the clustering effect. It is also known that the convergence order of the enclosure functions plays a crucial role in mitigating it [33, 102, 155].

The clustering of boxes in constraint projection is inherent due to the nature of the problem. In global optimization boxes can be fathomed since we are only looking for one value (the global optimum). In constraint projection problems boxes are fathomed only if they do not belong to the set, hence they are bound to accumulate in the boundary of the set.

Example 5.2. *Consider the problem of enclosing an arc of the unit sphere in* $\mathbb{R}^2$*, given by*

$$g(z) := z_1^2 + z_2^2 - 1 = 0 \quad z \in Z_0 = [\delta, 1-\delta]^2 \,, \tag{5.13}$$

with $\delta > 0$*. It is clear that this equation is underdetermined with one degree of freedom, hence the solution manifold is a line. It should also be clear that in this case we expect a big number of boxes in order to meet the prespecified tolerance. In fact using Algorithm 2 with Chebyshev models of order* $q = 2$ *and termination tolerance* $\varepsilon_{box} = 10^{-4}$*, the computed enclosure consists of* 15377 *boxes. Since the system is underdetermined and the domain*

excludes the poles of the unit sphere, we can always parameterize the arc as a function of one of the variables via the implicit function theorem, i.e. we can write

$$\chi(p) = \sqrt{1-p^2} \quad p \in [\delta, 1-\delta] \tag{5.14}$$

and then enclose Eq. (5.14) *using a polynomial model. Figure 5.7 shows a comparison between the enclosure of the solution set of Equation* (5.13) *and the solution set of Equation* (5.14) *using a Chebyshev model with* $q = 2$.

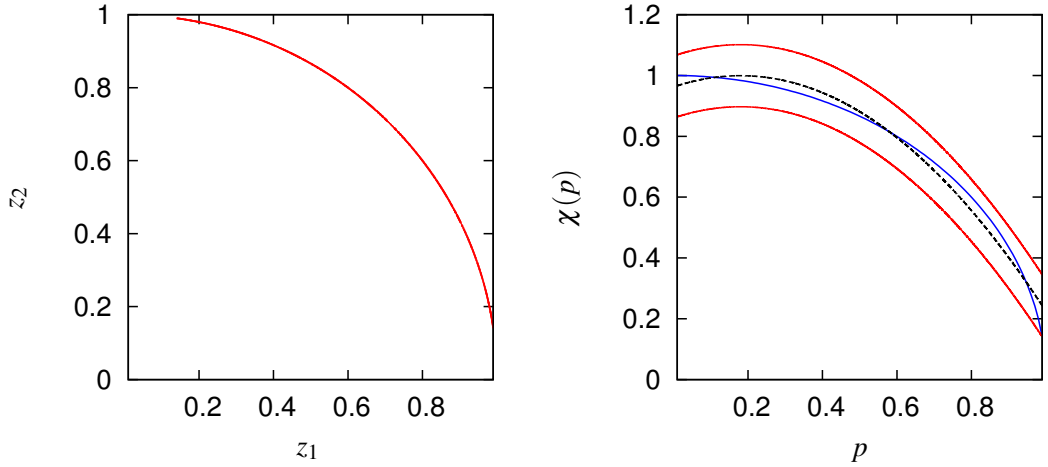

Fig. 5.7 Left: Enclosure of the solution set of $z_1^2 + z_2^2 - 1 = 0$ using Algorithm 2 (15377 boxes). Right: Enclosure of the solution set of $\chi(p) = \sqrt{1-p^2}$ using a Chebyshev Model with $q = 2$. The exact solution set is shown in blue, the polynomial $\mathcal{P}^2_{\chi,[\delta,1-\delta]}(p)$ is shown using the dashed black line and the enclosure $\mathcal{P}^2_{\chi,[\delta,1-\delta]}(p) \oplus \mathcal{R}^2_{\chi,[\delta,1-\delta]}$ is shown in red.

The strategy proposed in this section is also geometric in nature. Similarly to the approach presented in Example 5.2, it relies on the use of the implicit function theorem to exploit the structure of the boundary of the manifold in order to represent it as the union of polynomial models in a reduced space.

Assume that the set $Z_{g,Z_0,\Gamma}$ in (5.1) is defined by a set of equality and inequality constraints, e.g.

$$Z_{g,Z_0,\Gamma} = \left\{ z \in Z_0 \;\middle|\; g(z) = \begin{pmatrix} f(z) \\ h(z) \end{pmatrix} \in \Gamma \right\}$$

with $\Gamma := \{0_{n_f}\} \times \mathbb{R}^{n_g - n_f}$ and functions $f : \mathbb{R}^{n_z} \to \mathbb{R}^{n_f}$ and $h : \mathbb{R}^{n_z} \to \mathbb{R}^{n_g - n_f}_-$. Assume also that the system defined by f is underdetermined, i.e. $n_z \geq n_f$, then we can rewrite the system as $f(x,p) = 0$, where $x \in X$ and $p \in P$ for any box $Z = X \times P \in \mathbb{Z}_{\text{bnd}}$ denote the dependent

and independent variables respectively. According to the implicit function Theorem, if the Jacobian matrix is everywhere invertible on Z, there exists a k-smooth diffeomorphism $\chi : P \to X$ such that $f(\chi(p), p) = 0$.

The observation made in the preceding paragraph is the key to our domain reduction in reduced space, given a boundary box $Z = X \times P$, we can compute an enclosure for the Jacobian matrix of f on Z and if the matrix is everywhere invertible we can compute a q-th order Polynomial model $T^q_{\chi,P} = (\mathcal{P}^q_{\chi,P}, \mathcal{R}^q_{\chi,P})$ of χ on P. The construction of this polynomial model can be done by applying Newton-like operators such as the Gauss-Seidel operator for polynomial models [112].

The Gauss-Seidel operator on polynomial models proceeds by iteratively computing a sequence of Polynomial models $\{T^{q,k}_{\chi,P} = (\mathcal{P}^{q,k}_{\chi,P}, \mathcal{R}^{q,k}_{\chi,P})\}_{k\in\mathbb{N}}$ such that

$$\{\mathcal{P}^{q,k}_{\chi,P}(p) \mid p \in P\} \oplus \mathcal{R}^{q,k}_{\chi,P} \subseteq \{\chi(p) \mid p \in P\}$$

Given a polynomial model $\mathcal{T}^{q,k}_{\chi,P} = (\mathcal{P}^{q,k}_{\chi,P}, \mathcal{R}^{q,k}_{\chi,P})$ a new model $\mathcal{T}^{q,k+1}_{\chi,P} = (\mathcal{P}^{q,k+1}_{\chi,P}, \mathcal{R}^{q,k+1}_{\chi,P})$ can be computed using the iteration

$$\mathcal{T}^{q,k}_{b,P}(p) := W^k \mathcal{T}^q_{f,P}(\mathcal{P}^{q,k}_{\chi,P}(p), p) \tag{5.15a}$$

$$\mathcal{T}^{q,k}_{G,P}(p) := W^k \mathcal{T}^q_{\frac{\partial f}{\partial \chi},P}(\mathcal{P}^{q,k}_{\chi,P}(p) + [0,1]\mathcal{R}^{q,k}_{\chi,P}, p) \tag{5.15b}$$

for every $i \in \{1, \ldots, n_x\}$:

$$\mathcal{T}^{q,k+1}_{\Gamma_i,P}(p) := \mathcal{P}^{q,k}_{\chi_i,P}(p) - \frac{1}{\mathcal{T}^{q,k}_{G_{i,i},P}(p)} \left[\mathcal{T}^{q,k}_{b_i,P}(p) + \sum_{j=1}^{i-1} \Delta_j(p) + \sum_{j=i+1}^{n_x} \Lambda_j(p) \right] \tag{5.15c}$$

with

$$\Delta_j(p) = \left[\mathcal{T}^{q,k}_{G_{i,j},P} \left(\mathcal{T}^{q,k+1}_{\chi_j,P} - \mathcal{P}^{q,k}_{\chi_j} \right) \right](p) \quad \text{and} \quad \Lambda_j(p) = \mathcal{T}^{q,k}_{G_{i,j},P}(p) \mathcal{R}^k_{\chi_j,P} \, .$$

The new polynomial enclosure for χ on P is given by:

$$\mathcal{T}^{k+1}_{\chi_i,P} := \left(\mathcal{P}^{q,k+1}_{\Gamma_i,P}(p), \mathcal{R}^{q,k+1}_{\Gamma_i,P} \cup \left[\mathcal{P}^{q,k}_{\chi_i} - \mathcal{P}^{q,k+1}_{\Gamma_i,P}(P) + \mathcal{R}^{q,k}_{\chi_i,P} \right] \right) \, , \tag{5.15d}$$

For any preconditioning matrix $W^k \in \mathbb{R}^{n_x \times n_x}$. Such matrix can be computed by taking the midpoint of an interval enclosure of $\mathcal{T}^q_{\frac{\partial f}{\partial \chi},P}(\mathcal{P}^{q,k}_{\chi,P}(p) + [0,1]\mathcal{R}^{q,k}_{\chi,P}, p)$.

Example 5.2 (continued)**.** *Consider again enclosing the manifold given by Eq.* (5.13) *using the same tolerance and Chebyshev model order but now also including the domain reduction in reduced space. In this case the Chebyshev model in reduced space can be of a different order and in this case consider* $q = 3$ *for the reduced space. For these settings the arc is enclosed by the union of* 26 *Chebyshev models. This is depicted in Figure 5.8.*

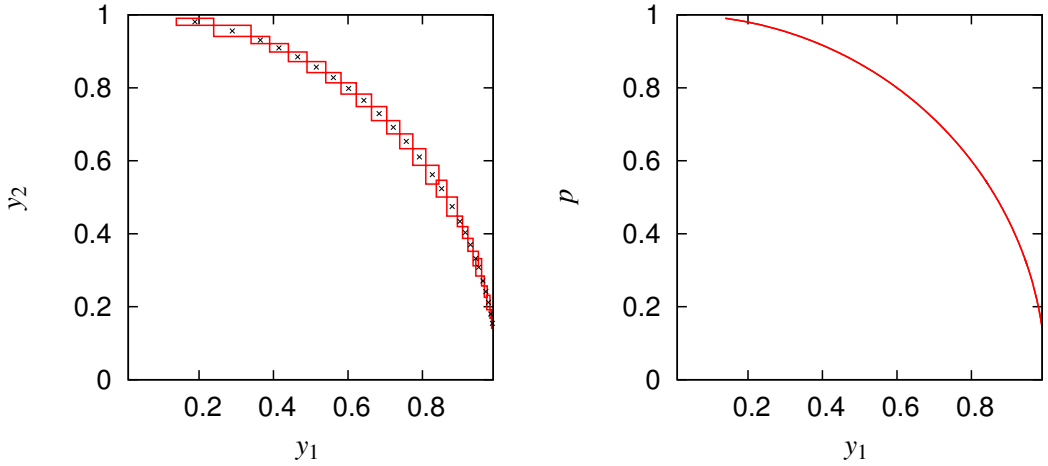

Fig. 5.8 Left: Enclosure of the solution set of $y_1^2 + y_2^2 - 1 = 0$ using Algorithm 2 (15377 boxes). Right: Enclosure of the solution set of $\chi(y_2) = \sqrt{1 - y_2^2}$ using a Chebyshev Model with $q = 2$. The exact solution set is shown in blue, the polynomial $\mathcal{P}^2_{\chi,Y_0}(p)$ is shown using the dashed black line and the enclosure $\mathcal{P}^2_{\chi,Y_0}(p) \oplus \mathcal{R}^2_{\chi,Y_0}$ is shown in red.

5.4 Guaranteed Parameter Estimation for an Anaerobic Digestion Process

This section illustrates the benefits of using high-order ODE bounding, optimization-based domain reduction, and CPU-time reduction in the context of guaranteed parameter estimation for a case study in anaerobic digestion. The model was introduced in Chapter 3 (cf. 3.6.2), but for the sake of completeness the equations are presented again,

$$\dot{X}_1 = (\mu_1(S_1) - \alpha D) X_1, \tag{5.16a}$$

$$\dot{X}_2 = (\mu_2(S_2) - \alpha D)\, X_2, \tag{5.16b}$$

$$\dot{S}_1 = D(S_1^{\text{in}} - S_1) - k_1 \mu_1(S_1) X_1, \tag{5.16c}$$

$$\dot{S}_2 = D(S_2^{\text{in}} - S_2) + k_2 \mu_1(S_1) X_1 - k_3 \mu_2(S_2) X_2, \tag{5.16d}$$

$$\dot{Z} = D(Z^{\text{in}} - Z), \tag{5.16e}$$

$$\dot{C} = D(C^{\text{in}} - C) - q_{\text{CO}_2} + k_4 \mu_1(S_1) X_1 + k_5 \mu_2(S_2) X_2\,. \tag{5.16f}$$

with auxiliary equations

$$\mu_1(S_1) := \bar{\mu}_1 \frac{S_1}{S_1 + K_{S_1}}, \tag{5.16g}$$

$$\mu_2(S_2) := \bar{\mu}_2 \frac{S_2}{S_2 + K_{S_2} + S_2^2/K_{I_2}} \tag{5.16h}$$

$$q_{\text{CO}_2} := k_{\text{L}} a (C + S_2 - Z - K_{\text{H}} P_{\text{CO}_2}), \tag{5.16i}$$

$$\text{with} \quad P_{\text{CO}_2} := \frac{\phi_{\text{CO}_2} - \sqrt{\phi_{\text{CO}_2}^2 - 4 K_{\text{H}} P_{\text{t}} (C + S_2 - Z)}}{2 K_{\text{H}}} \tag{5.16j}$$

$$\phi_{\text{CO}_2} := C + S_2 - Z + K_{\text{H}} P_{\text{t}} + \frac{k_6}{k_{\text{L}} a} \mu_2(S_2) X_2\,, \tag{5.16k}$$

Nominal values for all the parameters are taken from [11] (cf. Table 3.1). We apply guaranteed parameter estimation to estimate the kinetic parameters describing biomass growth. These parameters are listed in Table 5.1 with their nominal values and the considered variation ranges. The rest of the parameters as well as the initial conditions used are reported in Table 5.2 for sake of completeness.

Table 5.1 Estimated parameters of the anaerobic digestion model (5.16).

Parameter	Nominal value	Range	Unit
$\bar{\mu}_1$	1.2	[1.15, 1.25]	/day
K_{S_1}	7.1	[6.7, 7.3]	g(COD)/L
$\bar{\mu}_2$	0.74	[0.735, .75]	/day
K_{S_2}	9.28	[9.2, 9.5]	mmol/L
K_{I_2}	256	[235.0, 265.0]	mmol/L

In order to apply guaranteed parameter estimation, pseudo-experimental data are generated by simulating the model (5.16) with nominal parameter values from Tables 5.1 and 5.2 over a four-day period. The profiles used for the dilution rate and for the influent concentrations are those reported in Table 5.3. Moreover, three outputs are considered to carry out

Table 5.2 Constant parameters and initial states of the anaerobic digestion model (5.16).

Parameter	Value	Unit	Parameter	Value	Unit
k_1	42.14	g(COD)/g(cell)	$X_1(0)$	0.5	g(VSS)/L
k_2	116.5	mmol/g(cell)	$X_2(0)$	1.0	g(VSS)/L
k_3	268.0	mmol/g(cell)	$S_1(0)$	1.0	g(COD)/L
k_4	50.6	mmol/g(cell)	$S_2(0)$	5.0	mmol/L
k_5	343.6	mmol/g(cell)	$C(0)$	40.0	mmol/L
k_6	453.0	mmol/g(cell)	$Z(0)$	50.0	mmol/L
$k_{\mathrm{L}}a$	19.8	day^{-1}	P_{t}	1	atm
K_{H}	16	$\mathrm{mmol\,L^{-1}\,atm^{-1}}$	α	0.5	–

the estimation, namely S_1, S_2, and C, with measurements every 4 hours. In order to simulate the effect of measurement noise, the simulated values are rounded up or down to the nearest values by retaining, respectively, 2, 1 and 1 significant digits only; then, measurement error ranges of, respectively, ± 0.01, ± 0.1 and ± 0.1 are added around these values.

Table 5.3 Dilution rate and inlet concentration profiles corresponding to the pseudo-experimental data.

Input	Day 1	Day 2	Day 3	Day 4
D [/day]	0.25	1.00	1.00	0.25
S_1^{in} [g(COD)/L]	2.38	2.38	4.76	2.38
S_2^{in} [mmol/L]	80.0	80.0	160.0	80.0
Z^{in} [mmol/L]	50.0	50.0	100.0	50.0
C^{in} [mmol/L]	5.0	5.0	10.0	5.0

In the remainder of this section, we investigate guaranteed parameter estimation with different bounding techniques and with both optimization-based domain-reduction and CPU-time-reduction strategies in order to demonstrate the proposed improvements on a real-life problem. As previously in the simple case study, a 20% threshold and a maximum of 10 reduction loops are defined for the optimization-based domain-reduction strategy, and an absolute convergence threshold of $\varepsilon_{\mathrm{cvg}} = 10^{-4}$ is defined in connection to the CPU-time-reduction strategy. Problems of increasing complexity with 2, 3, 5 and 7 estimated parameters are addressed in Sect. 5.4.1–Sect. 5.4.3 below.

5.4.1 Case Study 1 – Two-Parameter Guaranteed Parameter Estimation

We consider the estimation of the parameters $\bar{\mu}_1$ and K_{S_1}, while the rest of the parameters from Tables 5.1 and 5.2 are fixed at their nominal values. The set-inversion algorithm is used with continuous-time ODE bounding techniques propagating Taylor models of orders $q = 1,\ldots,4$ with interval or ellipsoidal remainders. The termination criterion is defined as $\varepsilon_{\text{bnd}} = 10^{-4}$, whereas ε_{box} is set to zero, and the maximum number of iterations and maximal computational time are set to 1,000,000 iterations and 10 hours, respectively.

Fig. 5.9 shows both the inner- and outer-approximation of the set of guaranteed parameter estimates for the selected termination criteria. We start by noting that the true parameter values lie inside the approximation of the set P_{e} and that the selected termination criteria appear to be appropriate in view of the approximation level. Such a shape of the guaranteed parameter set is characteristic of the large correlations between the parameters $\bar{\mu}_1$ and K_{S_1}, according to (5.16g), and shows that $\mu_1(S_1) \approx \frac{\bar{\mu}_1}{K_{S_1}} X_1$ in this case.

When the method of differential inequalities is used to bound the reachable set, the algorithm stops after 1,000,000 iterations, without reaching the desired level of approximation—The volume of the partition $\mathbb{P}_{\text{bnd}}$ is about 2.5×10^{-4} then. This behavior is attributed to the inability of the method of differential inequalities to generate tight bounds for the anaerobic digestion model, even for very small parameter uncertainty. In contrast, higher-order ODE bounding techniques enable convergence of the set-inversion algorithm, as summarized in Table 5.4. Using Taylor models in combination with domain reduction, the algorithm is found to converge within a few thousand iterations (2nd and 3rd row), yet this remains insufficient to override the extra computational burden associated with domain reduction (1st row). The use of domain reduction becomes advantageous only when combined with CPU-time reduction (4th row), then leading to dramatic reduction of the run-time down to 49 s. Note that the number of iterations increases in the latter case compared to a run with the same settings but without CPU-time-reduction strategies, a behavior that is indeed expected and attributed to the approximation introduced by the finite convergence threshold ε_{cvg}.

5.4.2 Case Study 2 – Three-Parameter Guaranteed Parameter Estimation

Next, we consider the estimation of the parameters $\bar{\mu}_2$, K_{S_2} and K_{I_2}, while the rest of the parameters from Tables 5.1 and 5.2 are fixed at their nominal values. The set-inversion algorithm is run with the exact same settings as previously in Sect. 5.4.1, to the exception of the termination criterion ε_{bnd} that is now set to 5×10^{-5}.

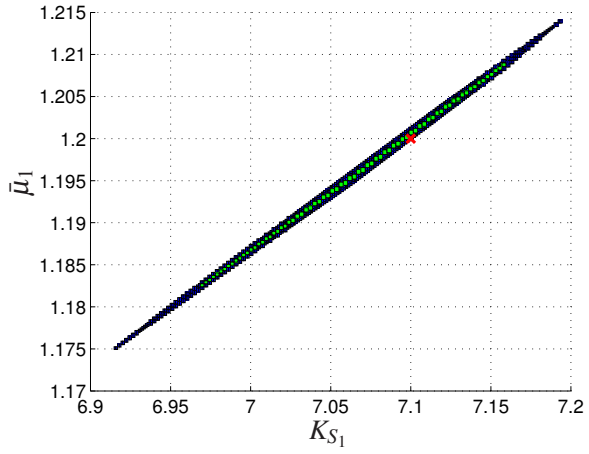

Fig. 5.9 Guaranteed parameter set approximation ($\mathbb{P}_{\text{int}}$ in green, $\mathbb{P}_{\text{bnd}}$ one in blue) for Case Study 1. The red cross indicates the 'true' (nominal) parameter values.

Table 5.4 Iteration counts and run-times of the set-inversion algorithm for Case Study 1.

Bounding method	Domain reduction	CPU-time reduction	Number of iterations	CPU time [s]
TM1+EL	✗	✗	8,738	801
TM1+DI	✓	✗	4,280	1682
TM4+EL	✓	✗	3,690	5748
TM4+EL	✓	✓	5,229	49

Fig. 5.10 shows the outer-approximation of the set of guaranteed parameter estimates for the selected termination criteria. The true parameter values lie inside the approximation of the set P_{e} and the selected termination criteria is deemed appropriate by visual inspection of the approximation level. Here again, the shape of the guaranteed parameter set is expected given the large correlations between the parameters $\bar{\mu}_2$, K_{S_2} and K_{I_2} according to (5.16h).

When the method of differential inequalities is used to bound the reachable set, the algorithm stops after 1,000,000 iterations, without reaching the desired level of approximation—The volume of the partition $\mathbb{P}_{\text{bnd}}$ is about 1.7×10^{-3} then. This behavior is again due to the inability of the method of differential inequalities to generate tight bounds for the anaerobic digestion model, even for very small parametric uncertainty. In contrast, higher-order ODE bounding techniques enable convergence of the set-inversion algorithm, as summarized in Table 5.5. Using first-order Taylor models with ellipsoidal remainders but no other improvement, the set-inversion algorithm takes about 24,000 iterations to converge (1st row). This is to be compared with a few thousand iterations when domain domain reduction is used (2nd, 3rd and 4th rows), similar to the previous 2-parameter case despite the extra parameter. This

Table 5.5 Iteration counts and run-times of the set-inversion algorithm for Case Study 2.

Bounding method	Domain reduction	CPU-time reduction	Number of iterations	CPU time [s]
TM1+EL	✗	✗	23,838	3,584
TM1+DI	✓	✗	4,095	1,154
TM4+EL	✓	✗	3,423	9,593
TM4+EL	✓	✓	3,488	111

suggests that the domain reduction might become more and more advantageous as the number of uncertain parameters increases, a trend that will confirm later on in Sect. 5.4.3. An expected behavior here is the reduction in the number of iterations as higher-order Taylor models are used. Finally, the effect of the CPU-time-reduction strategy is rather dramatic, with a run-time reduction about 2 orders of magnitude lower in the case of fourth-order Taylor models with ellipsoidal remainder bounds. It is noteworthy that the shortest run-time in this case is even lower, down to 41 s when fourth-order Taylor models with interval remainder bounds are used.

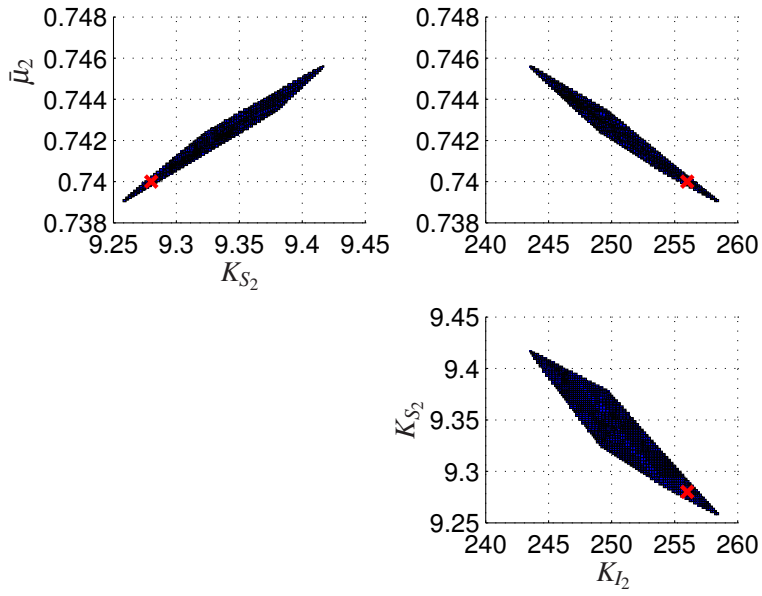

Fig. 5.10 Outer approximation of the set of guaranteed parameter estimates for Case Study 2. Projections onto the subspaces $(K_{S_2},\bar{\mu}_2)$, $(K_{I_2},\bar{\mu}_2)$, and (K_{I_2},K_{S_2}). The red crosses indicate the true (nominal) parameter values.

5.4.3 Case Study 3 – Five- and Seven-Parameter Guaranteed Parameter Estimation

We now consider the estimation of the parameters $\bar{\mu}_1$, K_{S_1}, $\bar{\mu}_2$, K_{S_2} and K_{I_2} simultaneously, leaving the other parameters at their nominal values in Table 5.2. The set-inversion algorithm is run with the exact same settings as previously in Sect. 5.4.1 and Sect. 5.4.2, apart from the termination criterion ε_{bnd} that is now set to 5×10^{-7}.

Fig. 5.11 shows the outer-approximation of the set of guaranteed parameter estimates for the selected termination criterion. The true parameter values lie inside the approximation of the set P_{e} and the approximation level, although coarse, validates the chosen termination criterion. Large correlations between the parameters $\bar{\mu}_1$ and K_{S_1}, on the one hand, and between $\bar{\mu}_2$, K_{S_2} and K_{I_2}, on the other hand, are observed, which is in complete agreement with the results shown earlier in Fig. 5.9 and Fig. 5.10. In contrast, rather small cross-correlations are observed between these two parameter subsets, as illustrated for instance for the parameters K_{S_1} and $\bar{\mu}_2$ in the top-right plot of Fig. 5.11.

Table 5.6 presents a comparison of the performance of various ODE bounding techniques and other improvement strategies. As previously, early termination is obtained with the method of differential inequalities after 1,000,000 iterations (without a single parameter box being fathomed here), and the algorithm now fails to converge after 10 hours with Taylor models as well when domain reduction is not applied. With respect to Taylor models combined with domain reduction, the benefit of higher-order ODE bounds in terms of the number of iterations is becoming more obvious in this 5-parameter problem—for instance, 10 times more iterations are needed with a first-order Taylor model compared to a fourth-order one. Yet, this large reduction is still not enough to overpower the extra computational burden of a single iteration with a higher-order Taylor model. Only when used in combination with CPU-time-reduction strategies are fourth-order Taylor models found to become competitive, with a runtime down to about 2,100 s. Finally, it is noteworthy that the shortest runtime in this case is close to 1,400 s, which is obtained for fourth-order Taylor model with interval remainder bounds and with all the developed reduction strategies.

Concerning the anaerobic digestion application, a more realistic parameter estimation problem should of course consider the initial biomass concentrations to be uncertain as well. Adding both initial concentrations $X_1(0)$ and $X_2(0)$ to the five uncertain kinetic parameters in Table 5.1 yields a total of seven parameters. In order to carry out the computations, only a small level of uncertainty of ± 0.001 g(VSS)/L is considered for $X_1(0)$ and $X_2(0)$ here, and the termination criterion ε_{bnd} is decreased to 5×10^{-11}. The set-inversion algo-

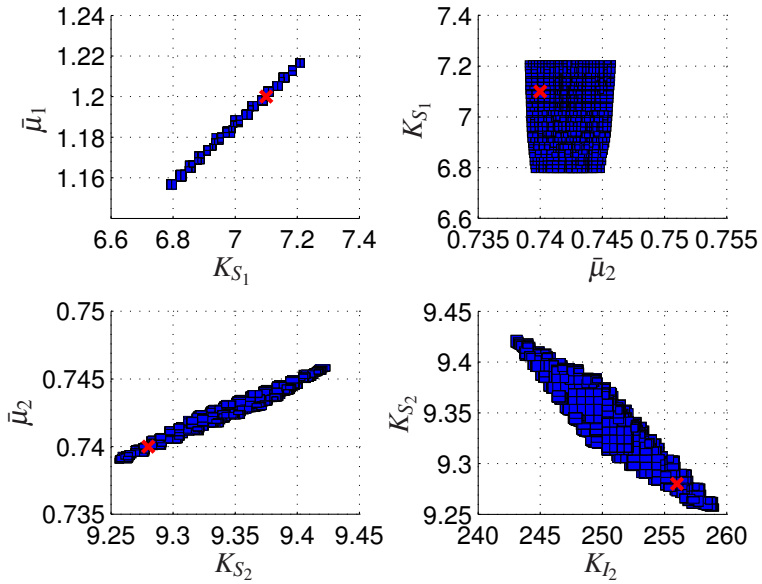

Fig. 5.11 Outer approximation of the set of guaranteed parameter estimates for Case Study 3. Projections onto the subspaces $(K_{S_1},\bar{\mu}_1)$, $(\bar{\mu}_2,K_{S_1})$, $(K_{S_2},\bar{\mu}_2)$, and $(K_{I_2},\bar{\mu}_2)$ only. The red crosses indicate the true (nominal) parameter values.

Table 5.6 Iteration counts and run-times of the set-inversion algorithm for Case Study 3.

Bounding method	Domain reduction	CPU-time reduction	Number of iterations	CPU time [s]
TM1+EL	✗	✗	219,178	36,000[1]
TM1+DI	✓	✗	41,148	8,295
TM4+EL	✓	✗	3,010	18,346
TM4+EL	✓	✓	3,130	2,118

[a]Did not converge to the specified termination criterion within the maximum allowable time.

rithm appears to be tractable only with fourth-order Taylor models (or higher) and only when combined with domain-reduction and CPU-time-reduction strategies—Convergence is achieved after 4,225 iterations and a corresponding runtime just above 19,000 s (5.3 h) in this case. These results confirm the advantage of high-order ODE bounding techniques and improved domain- and CPU-time-reduction strategies in addressing real-life problems possessing complex dynamics and more than a handful of uncertain parameters.

5.5 Guaranteed Asymptotic Analysis for an Anaerobic Digestion Process

In this section we consider again the anaerobic digestion model (5.16), but this time in the context of guaranteed asymptotic analysis. The focus is on varying the dilution rate D as the bifurcation parameter in the range $P_0 := [10^{-3}, 1.5]$, and the state-space domain for $[X_1\ X_2\ S_1\ S_2\ Z\ C\ P_{CO_2}]$ is $X_0 := [0,0.6] \times [0,0.8] \times [0,5.] \times [0,80] \times [0,80] \times [0,80] \times [0,1]$. As previously, the model parameter values are those listed in Table 3.1.

The results produced by the set-inversion algorithm are shown in Fig. 5.12, with an indication of the stable and unstable parts of the equilibrium manifold. In order to compute the enclosures in Step 3 of Algorithm 2, we use affine relaxations derived from 2nd-order Chebyshev models of the equilibrium constraints, Moreover, we perform domain reduction at each node via the solution of auxiliary LPs in order to refine the equilibrium set approximation.

The algorithm terminates after checking 20000 nodes in 121.312 (CPU) seconds. The equilibrium manifold is enclosed using 9149 boxes with a volume of the boundary partition of 3.82×10^{-5}.

After computing the equilibrium manifold, the stability of its branches is determined by first reducing the the Hurwitz matrix (5.8) to an upper triangular form, say U, using a modified Neville elimination algorithm in $O(n_x^2)$ operations [40], and then performing a series of stability test on the elements of U. The following stability tests can be performed sequentially on a subpartition $Y = X \times P$ by evaluating the elements c_i and U_{ii} in a given arithmetic, such as interval arithmetic or Taylor/Chebyshev model arithmetic:

1. If none of the coefficients c_i or diagonal elements U_{ii} are nonnegative, then all the equilibrium points in the considered subpartition are stable;

2. If at least one element c_i or U_{ii} is negative, then all the equilibrium points in the subpartition are unstable;

3. Otherwise (that is, if any one of the elements c_i or U_{ii} have zero in range), the subpartition may contain a bifurcation point.

Like-wise, bifurcation points can be identified by appending to the equilibrium conditions $f(x,p) = 0$ the algebraic conditions:

Steady-State bifurcation: $c_i \neq 0$ and $U_{ii} \neq 0$ for all $i \in \{1,\ldots,n_x\}$.

Hopf bifurcation: $c_{n_x} > 0, U_{n_x-1,n_x-1} = 0$ and $U_{n_x-2} > 0, \ldots, U_{1,1} > 0$.

Note that only steady-state bifurcations occur in this problem, which correspond to a change in stability of the equilibrium manifold. Since we introduce extra constraint in the system, finding the steady state bifurcations takes only 0.695 (CPU) seconds. One such bifurcation occurs here around $D = 1$ [day], where both a stable branch and an unstable branch merge into a single stable branch. Another bifurcation is seen to occur around $D = 1.07$ [day] with a change in stability along a single branch.

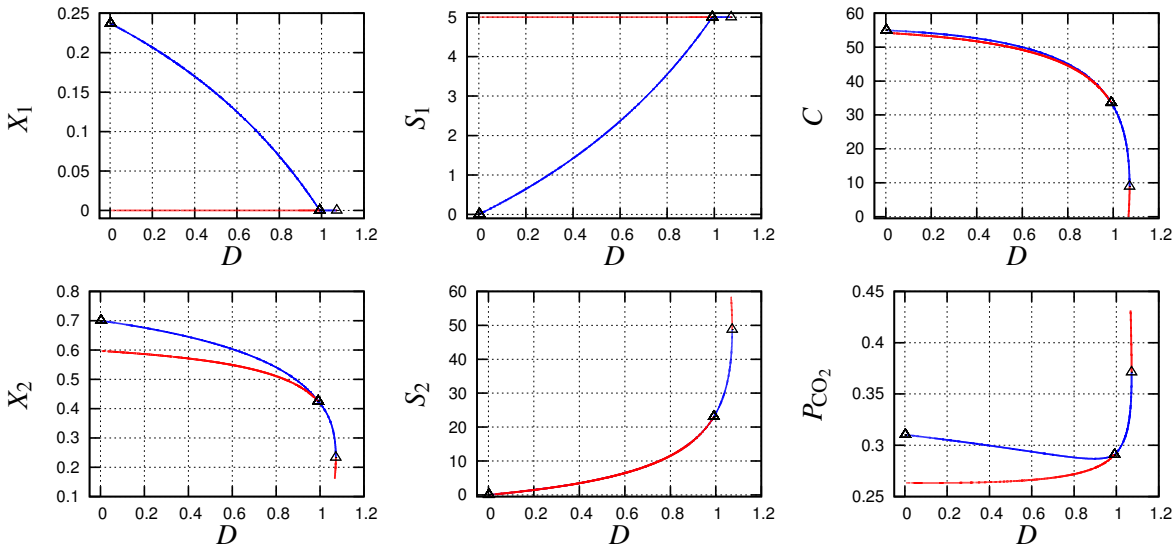

Fig. 5.12 Set of stable (blue) and unstable (red) equilibrium points of the two-step anaerobic digestion model with respect to the dilution rate D.

5.6 Guaranteed Asymptotic Analysis for a Nutrient-Resource-Consumer Model

In this section we consider a nutrient-resource-consumer model (NRC) [64] in the context of guaranteed asymptotic stability analysis. The model is given by the following system of parametric ODEs:

$$\dot{N} = D(N_r - N) - I_{NR}\frac{N}{k_{NR}+N}R \tag{5.17a}$$

$$\dot{R} = \left(\mu_{NR}\frac{N}{k_{NR}+N} - D - m_R\right)R - I_{RC}\frac{R}{k_{RC}+R}C \tag{5.17b}$$

$$\dot{C} = \left(\mu_{RC}\frac{R}{k_{RC}+R} - D - m_C\right)C \tag{5.17c}$$

with parameters $D = 0.02$, $\mu_{NR} = 0.5$ $\mu_{RC} = 0.2$, $I_{NR} = 1.25$, $I_{RC} = 0.333$ $k_{NR} = 8.0$, $k_{RC} = 9.0$, $m_R = 0.025$ and $m_C = 0.01$. The focus here is characterizing the asymptotic behavior of the model with respect to the nutrient density parameter Nr. For the constraint projection problem, the domain of interest is $Nr \in [0,80]$, $N \in [0,30]$, $R \in [0,30]$ and $C \in [0,24]$.

Figure 5.13 shows the bifurcation diagram for the NRC model with respect to the nutrient density parameter Nr. The constraint projection algorithm terminates in 392.54 (CPU) seconds and the solution manifold is enclosed using 271886 boxes for a total volume of the boundary partition of 3.834×10^{-5}.

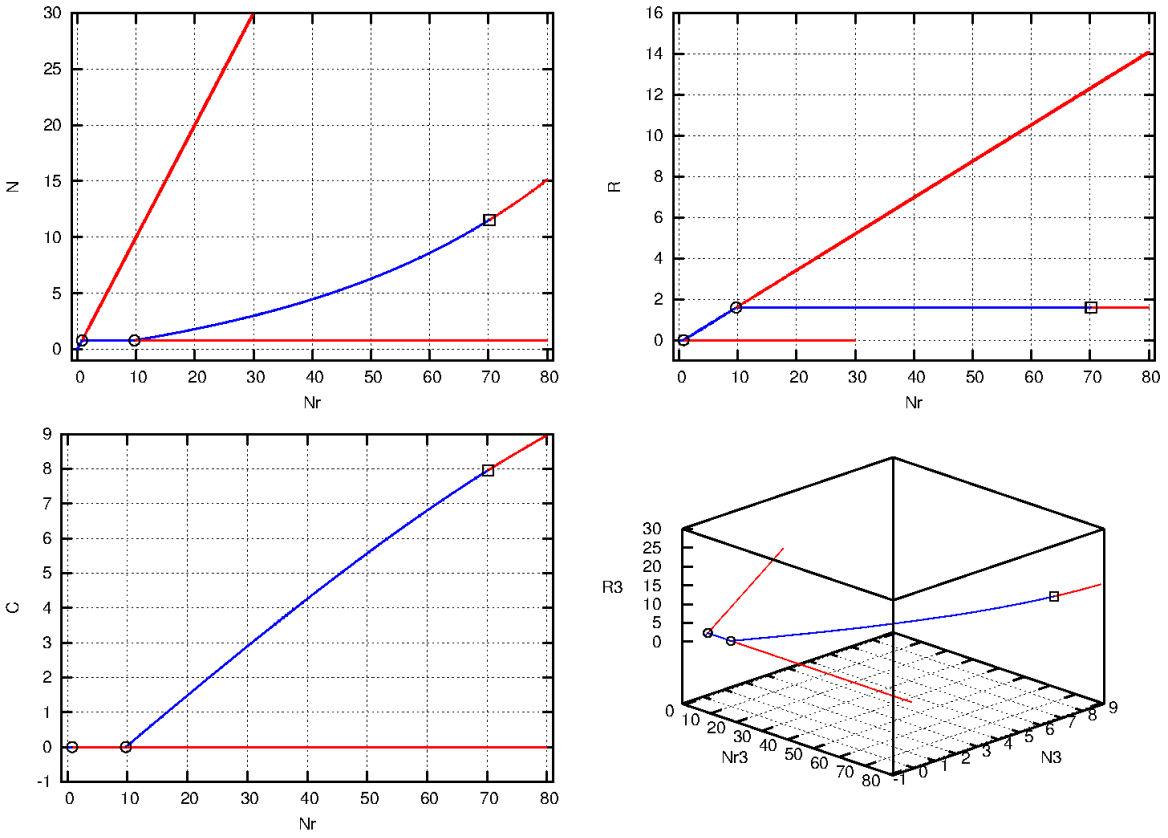

Fig. 5.13 Set of stable (blue) and unstable (red) equilibrium points of nutrient-resource consumer model with respect to nutrient density parameter Nr. Steady-state bifurcations are depicted by circles while the Hopf bifurcation is depicted by a square.

In the given computational domain, the model exhibits two steady state bifurcations — one close to $Nr = 1$, and the other at $Nr = 10$—computed by the algorithm in 0.468 (CPU)

seconds. The Hopf bifurcation is located at $Nr = 70$ and it gives rise to sustained oscillations in the dynamic behavior of the model.

5.7 Conclusion

In this chapter we have presented a constraint projection algorithm for computing enclosures of implicitly defined sets. For this algorithm we have presented and tested the effect of higher order bounding strategies based on polynomial models. We have also exploited the properties of polynomial models in order to reduce the size of the boxes created during the branching step of the algorithm. In order to make the polynomial model bounding more attractive a CPU reduction strategy is presented, where the enclosures of the constraint set can be reused after certain a convergence criterion has been met. For the case when the constraint system includes a system of under determined nonlinear algebraic equations, we have proposed a novel strategy that helps dealing with the clustering effect. This strategy exploits the structure of the boundary partition by representing it as the union of polynomial models using Newton-like methods. The approach has been applied to problems of guaranteed parameter estimation and guaranteed asymptotic analysis of dynamic systems described by parametric ODEs.

Chapter 6

Conclusion and Future Directions

This thesis led to some new contributions in the area of set-theoretical methods for nonlinear systems: first, we have introduced a new framework for the construction and analysis of set-valued arithmetics. Second, we have presented a set-valued integrator based on quadratically convergent affine-set parameterizations that preserves the asymptotic stability of parametric ODEs and thus does not suffer from bound explosion. Third, a framework for the construction and convergence analysis of continuous-time set propagation methods was introduced. This framework is based on the support function representation of convex sets and has also been extended for the case of nonconvex sets represented as the Minkowski sum of a polynomial model and a convex remainder term. The framework has been shown to include as special cases classical methods for continuous-time reachability analysis. Using this framework we have also presented formulations for robust model predictive control using tubes. Fourth, we have contributed to the area of constraint projection by introducing enhancements to classical branch and prune algorithms. Constraint projection was also used to enclose implicitly defined sets arising in guaranteed parameter estimation problems and guaranteed asymptotic analysis of dynamic systems.

In perspective, all of the methods presented in this thesis are intimately related. For example the affine-set arithmetics framework provides means to construct algorithms which are not dependent on a specific parameterization. It also allows us to understand how the convergence properties of a given parameterization affect the development of such algorithms. In particular the affine-set parameterization formalism provides a clear explanation on how the higher-order convergence properties of polynomial get lost when they are used in combination with remainders which are not invariant under affine transformations e.g. intervals. This higher-order convergence property is a key component for our ability to compute stable enclosures (in an infinite time horizon) of the solutions of asymptotically stable

parametric dynamic systems. If we continue with the story told by this thesis, we arrive to the construction of continuous-time methods for set propagation. As we have explained, although we choose not to, continuous time-methods can also be constructed by evaluating the ODE using affine set parameterizations. In this case the link between the chapters is also both way, since the arithmetic used to evaluate the right-hand side of the GDI corresponds to an implementation of a quadratically convergent affine-set arithmetic (Taylor model with ellipsoidal remainder). There is also a link which has been demonstrated empirically, but not proven mathematically: the stability statement for discrete-time integrators extends to continuous-time methods. Intuitively, stability of the enclosures of asymptotically stable dynamic systems is achieved when the enclosure contracts faster than the growth of the reachable set, so as long as we have that condition bound explosion will not occur. The extension of this result for continuous-dynamic systems seems to be correct for various reasons. First, consider constructing the auxiliary ODE system by evaluating the ODE in a particular quadratically convergent arithmetic, then the result would follow by taking the limit as the step-size in the Taylor expansion of the solution goes to zero. Now, the construction of continuous-time methods based on the GDI accounts for facet constraints while the construction directly using affine-set parameterizations does not, hence the GDI-based enclosures are at least as tight as their affine-set parameterization counterpart. Finally, the arithmetic presented in the numerical implementation for continuous-time methods (Taylor models with ellipsoidal remainder) is at least quadratically convergent in the Hausdorff-sense.

The development of bounding methods for both factorable functions and solutions to ODEs, had also a deep impact in the development of 'higher-level' set-theoretical algorithms. Take for example the continuous-time robust model predictive control approach presented at the end of Chapter 4, it is clear that the formulation shows a deep connection between robust forward invariant tubes and reachable tubes for uncertain ODEs. This connection allowed us to provide a formulation which is amenable for practical application (namely roubust MPC based on tubes with ellipsoidal cross-sections). Furthermore, the geometrical insight gained by expressing reachable tubes in terms of support functions enabled the exploitation of the boundary structure of the robust forward invariant tubes for the construction of semi-explicit feedback control laws.

All the methods presented in the earlier chapters come together in the constraint projection methodology. Continuous-time methods are used in order to compute bounds for sets whose implicit constraints depend on ODE solutions, while affine set arithmetics are used when computing bounds of factorable constraint systems and for computing relaxations of

the derived bounds. In here, the convergence and stability properties of the bounds are paramount for the successful implementation of the methods. Having stable bounds for the whole time horizon prevents the algorithm from branching excessively and thus faster termination can be achieved. Convergence rate of the enclosures in constraint projection has also an important role in the termination of constraint projection algorithm, as we keep branching, the overestimation reduces faster which means a smaller number of iterations in the algorithm. In complete-search methods for global optimization, the convergence rate of the enclosures directly affect the clustering problem. In constraint clustering is natural to the geometry of the problem. In the approach we have proposed in this thesis for domain reduction in reduced spaces, the convergence rate of the enclosures is again directly involved in mitigation of the algorithm. The faster the enclosures converge, the faster the Newton-like methods for polynomial models can be used, and the less boxes will be used in the boundary.

A plethora of problems occurring in science and engineering can be formulated using a set-theoretical formalism. As we have seen throughout this thesis, set-theoretical approaches can also be used to solve the problems. Working with sets, often allows us to exploit certain characteristics of the problem, that could be ignored by classical numerical methods. They also provide a robustness and if implemented rigorously they provide rigorous numerical certificates. One of the main reasons for not using these methods is that they are not amenable for computational implementation. Unlike other approaches, the methods presented in these thesis were designed to be general but also amenable for implementation. All of these contributions, have been implemented in CRONOS, a C++ package that builds upon the MC++ library for bounding factorable functions.

Despite these contributions, there is still some road ahead before set-theoretical methods can be used to solve practical problems, in particular large-scale problems. Possible research directions leading towards the implementation of set-theoretical methods include structure exploitation for set-extensions (e.g. sparse implementations of set-valued arithmetics) and investigating and implementing different range bounders for nonconvex parameterizations (e.g. LMI relaxations for polynomial models).

One very interesting research direction is the development of (semi) explicit robust tube-based optimal control and its receding horizon counterpart. The min-max differential inequality is a powerful theoretical tool for the development of such methods, since it allows us to compute (semi) explicit feedback control laws instead of control actions typically computed in nominal optimal control. Consider for example the case of periodic dynamic systems within a fixed horizon optimal control problem framework. In this case a nominal

optimal control problem can be solved in order to compute a trajectory such that a certain (perhaps economic) criterion is optimized. Then, a tube-based optimal control problem can be solved in order to construct a tube (with its associated feedback law) around the nominal solution for a given uncertainty. This effectively renders an open-loop optimal control strategy into a closed loop control without the need of using receding horizon control. Extensions to polytopic tubes, and in general tubes that do not have smooth or positively curved cross-sections can also be constructed using the min-max DI framework.

Finally, the constraint projection framework can be applied to systems of optimality conditions for nonlinear programs or optimal control problems, in order to track changes in the optimal solution or active constraints in the presence of uncertainties. In particular, this approach could be applied to explicit MPC methodologies (e.g. using multi parametric programming). The use of the constraint projection presented in this thesis for multiparametric programming would allow the determination of a polynomial approximation of the boundary between critical regions in nonlinear multi-parametric programs.

In retrospective, set-theoretical methods are not only useful in the formulation of engineering problems—in particular problems involving uncertainties, but they usefulness also extends to the solution of those problems. The set-theoretical formalism often allows for a global understanding of the problem, which translates into efficient algorithms capable of exploiting its geometric structure. These methods work-well in small and medium-scale problems but more research is needed in order to apply them to large-scale problems.

References

[1] Abasaeed, A.E. (2000). "Bifurcation and chaos for a mutating autocatalator in a cstr". *Bioprocess Engineering*, vol. 22 (4), pp. 337–346.

[2] Adjiman, C.S., Dallwig, S., Floudas, C.A. and Neumaier, A. (1998). "A global optimization method, αBB, for general twice-differentiable constrained NLPs - I. Theoretical advances". *Computers & Chemical Enginnering*, vol. 22 (9), pp. 1137–1158.

[3] Alamo, T., Bravo, J.M. and Camacho, E.F. (2005). "Guaranteed state estimation by zonotopes". *Automatica*, vol. 41 (6), pp. 1035–1043.

[4] Allgower, E. and Georg, K. (1980). "Simplicial and continuation methods for approximating fixed points and solutions to systems of equations". *SIAM Review*, vol. 22 (1), pp. 28–85.

[5] Allgower, E. and Georg, K. (2003). *Introduction to Numerical Continuation Methods*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.

[6] Althoff, M., Stursberg, O. and Buss, M. (2008). "Reachability analysis of nonlinear systems with uncertain parameters using conservative linearization". *47th IEEE Conference on Decision and Control, 2008. CDC 2008*. pp. 4042–4048.

[7] Aubin, J.P. (1991). *Viability theory*. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA.

[8] Aubin, J.P. and Cellina, A. (1984). *Differential Inclusions: Set-Valued Maps and Viability Theory*. Grundlehren der mathematischen Wissenschaften, #264. Springer-Verlag, Berlin.

[9] Azagra, D. and Ferrera, J. (2002). "Every closed convex set is the set of minimizers of some C^∞-smooth convex function". *Proceedings of the American Mathematical Society*, vol. 130 (12), pp. 3687–3692.

[10] Berger, M.S. (1977). *Nonlinearity & Functional Analysis: Lectures on Nonlinear Problems in Mathematical Analysis*. Academic Press.

[11] Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A. and Steyer, J.P. (2001). "Dynamical model development and parameter identification for an anaerobic wastewater treatment process". *Biotechnology & Bioengineering*, vol. 75 (4), pp. 424–438.

[12] Bertsekas, D.P. (2005). *Dynamic programming and optimal control. Vol. I*. Athena Scientific, Belmont, MA, 3rd edn.

[13] Berz, M. and Makino, K. (1998). "Verified integration of ODEs and flows using differential algebraic methods on high-order Taylor series". *Reliable Computing*, vol. 4, pp. 361–369.

[14] Berz, M. and Makino, K. (2006). "Performance of taylor model methods for validated integration of ODEs". *Lecture Notes in Computer Science*, vol. 3732, pp. 65–74.

[15] Berz, M. (1997). "From taylor series to taylor models". *AIP Conference Proceedings*, vol. 405. AIP Publishing, pp. 1–23.

[16] Berz, M., Hoffstätter, G. and Atter, G.H. (1998). "Computation and application of taylor polynomials with interval remainder bounds". *Reliable Computing*, vol. 4, pp. 83—97.

[17] Blanchini, F. (1999). "Set invariance in control". *Automatica*, vol. 35 (11), pp. 1747–1767.

[18] Blanchini, F. and Miani, S. (2008). *Set-theoretic methods in control*. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA.

[19] Bompadre, A. and Mitsos, A. (2012). "Convergence rate of McCormick relaxations". *Journal of Global Optimization*, vol. 52 (1), pp. 1–28.

[20] Bompadre, A., Mitsos, A. and Chachuat, B. (2013). "Convergence analysis of Taylor and McCormick-Taylor models". *Journal of Global Optimization*, vol. 57 (1), pp. 75–114.

[21] Cannon, M., Buerger, J., Kouvaritakis, B. and Raković, S. (2011). "Robust tubes in nonlinear model predictive control". *IEEE Trans. Automat. Control*, vol. 56 (8), pp. 1942–1947.

[22] Chachuat, B. and Latifi, M.A. (2003). "A new approach in deterministic global optimization of problems with ordinary differential equations". *Frontiers in Global Optimization*, *Nonconvex Optimization and Its Applications*, vol. 74 (C.A. Floudas and P.M. Pardalos, eds.). Kluwer Academic Publishers, pp. 83–108.

[23] Chachuat, B., Houska, B., Paulen, R., Peri'c, N., Rajyaguru, J. and Villanueva, M.E. (2015). "Set-theoretic approaches in analysis, estimation and control of nonlinear systems". *9th {IFAC} Symposium on Advanced Control of Chemical Processes {ADCHEM}*, IFAC PapersOnline Volume 48, Issue 8. Whistler, Canada, pp. 981 – 995.

[24] Chachuat, B., Singer, A.B. and Barton, P.I. (2006). "Global methods for dynamic optimization and Mixed-Integer dynamic optimization". *Ind. Eng. Chem. Res.*, vol. 45 (25), pp. 8373–8392.

[25] Chachuat, B. and Villanueva, M. (2012). "Bounding the solutions of parametric ODEs: when taylor models meet differential inequalities". *22nd European Symposium on Computer Aided Process Engineering*, *Computer Aided Chemical Engineering*, vol. 30 (I.D.L. Bogle and M. Fairweather, eds.). Elsevier, pp. 1307–1311.

[26] Chutinan, A. and Krogh, B.H. (1999). "Verification of polyhedral-invariant hybrid automata using polygonal flow pipe approximations". *Hybrid Systems: Computation and Control* (F.W. Vaandrager and J.H. van Schuppen, eds.), no. 1569 in Lecture Notes in Computer Science. Springer, Berlin, pp. 76–90.

[27] Clarke, F.H. (1975). "Generalized gradients and applications". *Transactions of the American Mathematical Society*, vol. 205, pp. 247–262.

[28] Coddington, E.A. and Levinson, N. (1955). *Theory of Ordinary Differential Equations*. McGraw-Hill, New York.

[29] Corliss, G.F. and Rihm, R. (1996). "Validating an a priori enclosure using High-Order taylor series". *In Scientific Computing, Computer Arithmetic, and Validated Numerics*. Akademie Verlag, p. 228–238.

[30] Dadhe, K. and Engell, S. (2008). "Robust nonlinear model predictive control: A multi-model nonconservative approach". *International Workshop on assessment and future directions on Nonlinear Model Predictive Control*. Pavia, Italy, pp. –.

[31] Diehl, M. and Björnberg, J. (2004). "Robust dynamic programming for min-max model predictive control of constrained uncertain systems". *IEEE Trans. Automat. Control*, vol. 49 (12), pp. 2253–2257.

[32] Diehl, M., Ferreau, H.J. and Haverbeke, N. (2009). "Efficient numerical methods for nonlinear MPC and moving horizon estimation". *Nonlinear Model Predictive Control: Towards New Challenging Applications*, *Lecture Notes in Control and Information Sciences*, vol. 384. Springer, Heidelberg, pp. 391–417.

[33] Du, K. and Kearfott, R.B. (1994). "The cluster problem in multivariate global optimization". *Journal of Global Optimization*, vol. 5 (3), pp. 253–265.

[34] Dzetkulič, T. (2015). "Rigorous integration of non-linear ordinary differential equations in chebyshev basis". *Numerical Algorithms*, vol. 69 (1), pp. 183–205.

[35] Eijgenraam, P. (1981). *The solution of initial value problems using interval arithmetic: formulation and analysis of an algorithm*, vol. Volume 144 of Mathmematical Centre tracts. Mathematisch Centrum.

[36] Engell, S. (2009). "Online optimizing control: The link between plant economics and process control". *10th International Symposium on Process Systems Engineering: Part A*, *Computer Aided Chemical Engineering*, vol. 27. Elsevier, pp. 79–86.

[37] Evans, M.A., Cannon, M. and Kouvaritakis, B. (2012). "Robust mpc for linear systems with bounded multiplicative uncertainty". *Proc. IEEE 51st Annual Conference on Decision and Control*. Maui (HI), pp. 248–253.

[38] Fiacco, A.V. (1983). *Introduction to sensitivity and stability analysis in nonlinear programming*, *Mathematics in Science and Engineering*, vol. 165. Academic Press, Inc., Orlando, FL.

[39] Friedrichs, K. (1944). "The identity of weak and strong extensions of differential operators". *Transactions of the American Mathematical Society*, vol. 55 (1), pp. 132–151.

[40] Gasca, M. and na, J.P. (1992). "Total positivity and neville elimination". *Linear Algebra and its Applications*, vol. 165, pp. 25–44.

[41] Goulart, P.J. and Kerrigan, E.C. (2007). "Output feedback receding horizon control of constrained systems". *Internat. J. Control*, vol. 80 (1), pp. 8–20.

[42] Goulart, P.J., Kerrigan, E.C. and Maciejowski, J.M. (2006). "Optimization over state feedback policies for robust control with constraints". *Automatica*, vol. 42 (4), pp. 523–533.

[43] Griewank, A. and Walther, A. (2008). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd edn.

[44] Gronwall, T.H. (1919). "Note on the derivatives with respect to a parameter of the solutions of a system of differential equations". *Annals of Mathematics*, vol. 20 (2), pp. 292–296.

[45] Hairer, E., N?rsett, S.P. and Wanner, G. (1993). *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer Science & Business Media.

[46] Helmberg, G., Wagner, P. and Veltkamp, G. (1993). "On faddeev-leverrier's method for the computation of the characteristic polynomial of a matrix and of eigenvectors". *Linear Algebra and its Applications*, vol. 185, pp. 219–233.

[47] Henrion, D. and Korda, M. (2014). "Convex computation of the region of attraction of polynomial control systems". *IEEE Transactions on Automatic Control*, vol. 59 (2), pp. 297–312.

[48] Herceg, M., Kvasnica, M., Jones, C. and Morari, M. (2013). "Multi-Parametric Toolbox 3.0". *Proc. European Control Conference (ECC)*. Zürich, Switzerland, pp. 502–510. http://control.ee.ethz.ch/~mpt.

[49] Hindmarsh, A.C., Brown, P.N., Grant, K.E., Lee, S.L., Serban, R., Shumaker, D.E. and Woodward, C.S. (2005). "SUNDIALS: suite of nonlinear and Differential/Algebraic equation solvers". *ACM Trans. Math. Softw.*, vol. 31 (3), p. 363–396.

[50] Hladík, M. (2013). "Bounds on eigenvalues of real and complex interval matrices". *Applied Mathematics and Computation*, vol. 219 (10), pp. 5584–5591.

[51] Houska, B., Ferreau, H.J. and Diehl, M. (2011). "An auto-generated real-time iteration algorithm for nonlinear MPC in the microsecond range". *Automatica*, vol. 47 (10), pp. 2279–2285.

[52] Houska, B., Mohammadi, A. and Diehl, M. (2015). "A short note on constrained linear control systems with multiplicative ellipsoidal uncertainty". Submitted for publication.

[53] Houska, B., Villanueva, M. and Chachuat, B. (2013). "A validated integration algorithm for nonlinear ODEs using taylor models and ellipsoidal calculus". *2013 IEEE 52nd Annual Conference on Decision and Control (CDC)*. pp. 484–489.

[54] Houska, B., Villanueva, M. and Chachuat, B. (2015). "Stable set-valued integration of nonlinear dynamic systems using affine set-parameterizations". *SIAM Journal on Numerical Analysis*, vol. 53 (5), pp. 2307–2328.

[55] Houska, B., Ferreau, H.J. and Diehl, M. (2011). "ACADO toolkit—An open-source framework for automatic control and dynamic optimization". *Optimal Control Applications and Methods*, vol. 32 (3), pp. 298–312.

[56] Houska, B., Logist, F., Van Impe, J. and Diehl, M. (2012). "Robust optimization of nonlinear dynamic systems with application to a jacketed tubular reactor". *Journal of Process Control*, vol. 22 (6), pp. 1152–1160.

[57] Jaulin, L. (2002). "Nonlinear bounded-error state estimation of continuous-time systems". *Automatica*, vol. 38 (6), pp. 1079 – 1082.

[58] Jaulin, L., Kieffer, M., Didrit, O. and Walter, E. (2001). *Applied Interval Analysis*. Springer-Verlag, London.

[59] Jaulin, L. and Walter, E. (1993). "Set inversion via interval analysis for nonlinear bounded-error estimation". *Automatica*, vol. 29 (4), pp. 1053–1064.

[60] Kahoui, M.E. and Weber, A. (2001). "Deciding hopf bifurcations by quantifier elimination in a software-component architecture". *Journal of Symbolic Computation*, vol. 30 (2), pp. 161–179.

[61] Kieffer, M. and Walter, E. (2011). "Guaranteed estimation of the parameters of nonlinear continuous-time models: contributions of interval analysis". *nternational Journal of Adaptive Control & Signal Processing*, vol. 25 (3), pp. 191–207.

[62] Kieffer, M., Walter, E. and Simeonov, I. (2006). "Guaranteed nonlinear parameter estimation for continuous-time dynamical models". *Proceeding of the 14th IFAC Symposium on System Identification (SYSID)* (B. Ninness and H. Hjalmarsson, eds.). pp. 843–848.

[63] Kletting, M., Kieffer, M. and Walter, E. (2011). *Modeling, Design, and Simulation of Systems with Uncertainties*, *Mathematical Engineering*, vol. 3, chap. Two approaches for guaranteed state estimation of nonlinear continuous-time models. Springer, pp. 199–220.

[64] Kooi, B. (2003). "Numerical bifurcation analysis of ecosystems in a spatially homogeneous environment". *Acta Biotheoretica*, vol. 51 (3), pp. 189–222.

[65] Korda, M., Henrion, D. and Jones, C.N. (2014). "Convex computation of the maximum controlled invariant set for polynomial control systems". *SIAM Journal on Control and Optimization*, vol. 52 (5), pp. 2944–2969.

[66] Kothare, M.V., Balakrishnan, V. and M., M. (1996). "Robust constrained model predictive control using linear matrix inequalities". *Automatica*, vol. 32 (10), pp. 1361–1379.

[67] Kurzhanski, A.B. (2006). "Comparison principle for equations of the Hamilton-Jacobi type in control theory". *Proceedings of the Steklov Institute of Mathematics*, vol. 253 (1), pp. S185–S195.

[68] Kurzhanski, A.B. and Filippova, T.F. (1993). "On the theory of trajectory tubes—a mathematical formalism for uncertain dynamics, viability and control". *Advances in nonlinear dynamics and control: a report from Russia*, *Progr. Systems Control Theory*, vol. 17. Birkhäuser Boston, Boston, MA, pp. 122–188.

[69] Kurzhanski, A.B. and Vályi, I. (1997). *Ellipsoidal calculus for estimation and control*. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA.

[70] Kurzhanski, A.B. and Varaiya, P. (2002). "Reachability analysis for uncertain systems - The ellipsoidal technique". *Dynamics of Continuous, Discrete and Impulsive System, Series B*, vol. 9 (3), pp. 347,368.

[71] Lakshmikantham, V. and Leela, S. (1969). *Differential and integral inequalities: Theory and applications. Vol. I: Ordinary differential equations*. Academic Press, New York-London. Mathematics in Science and Engineering, Vol. 55-I.

[72] Langson, W., Chryssochoos, I., Raković, S.V. and Mayne, D.Q. (2004). "Robust model predictive control using tubes". *Automatica*, vol. 40 (1), pp. 125–133.

[73] Lasserre, J. (2009). *Moments, Positive Polynomials and Their Applications*. World Scientific.

[74] Lee, Y.I., Kouvaritakis, B. and Cannon, M. (2002). "Constrained receding horizon predictive control for nonlinear systems". *Automatica*, vol. 38 (12), pp. 2093–2102.

[75] Li, D. and Xi, Y. (2010). "The feedback robust MPC for LPV systems with bounded rates of parameter changes". *IEEE Trans. Automat. Control*, vol. 55 (2), pp. 503–507.

[76] Limon, D., Bravo, J.M., Alamo, T. and Camacho, E.F. (2005). "Robust MPC of constrained nonlinear systems based on interval arithmetic". *IEE Proceedings–Control Theory and Applications*, vol. 152 (3), pp. 325–332.

[77] Lin, Q. and Rokne, J.G. (1995). "Methods for bounding the range of a polynomial". *Journal of Computational and Applied Mathematics*, vol. 58 (2), pp. 193–199.

[78] Lin, Y. and Stadtherr, M.A. (2007). "Deterministic global optimization of nonlinear dynamic systems". *AIChE Journal*, vol. 53 (4), pp. 866–875.

[79] Lin, Y. and Stadtherr, M.A. (2009). "Rigorous model-based safety analysis for nonlinear continuous-time systems". *Computers & Chemical Engineering*, vol. 33 (2), pp. 493–502.

[80] Lin, Y. and Stadtherr, M.A. (2007). "Guaranteed state and parameter estimation for nonlinear Continuous-Time systems with Bounded-Error measurements". *Industrial & Engineering Chemistry Research*, vol. 46 (22), pp. 7198–7207.

[81] Lohner, R.J., Cash, J.R. and Gladwell, L. (1992). "Computations of guaranteed enclosures for the solutions of ordinary initial and boundary value problems". *Computational Ordinary Differential Equations*, vol. 1 (J.R. Cash and I. Gladwell, eds.). Clarendon Press, p. 425–436.

[82] Lohner, R.J. (2001). "On the ubiquity of the wrapping effect in the computation of error bounds". *Perspectives on Enclosure Methods* (P.D.U. Kulisch, P.D.R. Lohner and D.A. Facius, eds.). Springer Vienna, pp. 201–216.

[83] Lygeros, J. (2004). "On reachability and minimum cost optimal control". *Automatica*, vol. 40 (6), pp. 917–927.

[84] Makino, K. and Berz, M. (1999). "Efficient control of the dependency problem based on Taylor model methods". *Reliable Computing*, vol. 5 (1), pp. 3–12.

[85] Makino, K. and Berz, M. (2003). "Taylor models and other validated functional inclusion methods". *international Journal of Pure and Applied Mathematics*, vol. 4 (4), pp. 379–456.

[86] Maranas, C.D. and Floudas, C.A. (1994). "Global minimum potential energy conformations of small molecules". *Journal of Global Optimization*, vol. 4, pp. 135–170.

[87] Maranas, C.D. and Floudas, C.A. (1994). "A deterministic global optimization approach for molecular structure determination". *The Journal of Chemical Physics*, vol. 100 (2), pp. 1247–1261.

[88] Mayne, D.Q., Raković, S.V., Findeisen, R. and Allgöwer, F. (2009). "Robust output feedback model predictive control of constrained linear systems: time varying case". *Automatica*, vol. 45 (9), pp. 2082–2087.

[89] McCormick, G.P. (1976). "Computability of global solutions to factorable nonconvex programs: Part I – Convex underestimating problems". *Mathematical Programming*, vol. 10, pp. 147–175.

[90] Misener, R. and Floudas, C. (2014). "A framework for globally optimizing mixed-integer signomial programs". *Journal of Optimization Theory and Applications*, vol. 161 (3), pp. 905–932.

[91] Mitchell, I. and Tomlin, C. (2003). "Overapproximating reachable sets by hamilton-jacobi projections". *Journal of Scientific Computing*, vol. 19 (1-3), pp. 323–346.

[92] Mitchell, I.M., Bayen, A.M. and Tomlin, C.J. (2005). "A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games". *IEEE Transactions on Automatic Control*, vol. 50 (7), pp. 947–957.

[93] Mitsos, A., Chachuat, B. and Barton, P.L. (2009). "Towards global bilevel dynamic optimization". *Journal of Global Optimization*, vol. 45 (1), pp. 63–93.

[94] Mönnigmann, M. (2008). "Efficient calculation of bounds on spectra of hessian matrices". *SIAM Journal on Scientific Computing*, vol. 30 (5), p. 2340.

[95] Moore, R.E. (1992). "Parameter sets for bounded-error data". *Mathematics and Computers in Simulation*, vol. 34 (2), pp. 113–119.

[96] Moore, R.E. (1979). *Methods and Applications of Interval Analysis*. SIAM.

[97] Moore, R.E., Kearfott, R.B. and Cloud, M.J. (2009). *Introduction to Interval Analysis*. SIAM.

[98] Naumann, U. (2012). *The Art of Differentiating Computer Programs: An Introduction to Algorithmic Differentiation*. SIAM.

[99] Nedialkov, N.S., Jackson, K.R. and Pryce, J.D. (2001). "An effective high-order interval method for validating existence and uniqueness of the solution of an IVP for an ODE". *Reliable Computing*, vol. 7, pp. 449–465.

[100] Nedialkov, N., Jackson1, K. and Corliss, G. (1999). "Validated solutions of initial value problems for ordinary differential equations". *Applied Mathematics and Computation*, vol. 105 (1), pp. 21–68.

[101] Neher, M., Jackson, K.R. and Nedialkov, N.S. (2007). "On taylor model based integration of ODEs". *SIAM Journal on Numerical Analysis*, vol. 45, p. 236.

[102] Neumaier, A. (2004). "Complete search in continuous global optimization and constraint satisfaction". *Acta Numerica*, vol. 13, pp. 271–369.

[103] Neumaier, A. (2003). "Taylor Forms—Use and limits". *Reliable Computing*, vol. 9 (1), pp. 43–79.

[104] Pannocchia, G., Rawlings, J.B. and Wright, S.J. (2011). "Conditions under which suboptimal nonlinear MPC is inherently robust". *Systems Control Lett.*, vol. 60 (9), pp. 747–755.

[105] Papamichail, I. and Adjiman, C.S. (2004). "Global optimization of dynamic systems". *Computers & Chemical Engineering*, vol. 28 (3), pp. 403–415.

[106] Papamichail, I. and Adjiman, C.S. (2002). "A rigorous global optimization algorithm for problems with ordinary differential equations". *J. of Global Optimization*, vol. 24 (1), p. 1–33.

[107] Parrilo, P.A. (2003). "Semidefinite programming relaxations for semialgebraic problems". *Mathematical Programming*, vol. 96 (2), pp. 293–320.

[108] Paulen, R., Villanueva, M.E., Fikar, M. and Chachuat, B. (2013). "Guaranteed parameter estimation in nonlinear dynamic systems using improved bounding techniques". *Proceedings of the 2013 European Control Conference (ECC'13)*. Zurich, Switzerland, pp. 4514–4519.

[109] Paulen, R., Villanueva, M.E. and Chachuat, B. (2013). "Optimization-based domain reduction in guaranteed parameter estimation of nonlinear dynamic systems". *9th IFAC Symposium on Nonlinear Control Systems (NOLCOS)*, Nonlinear Control Systems. Toulusse, France., pp. 564–569.

[110] Paulen, R., Villanueva, M.E. and Chachuat, B. (2015). "Guaranteed parameter estimation of non-linear dynamic systems using high-order bounding techniques with domain and CPU-time reduction strategies". *IMA Journal of Mathematical Control and Information*, (Accepted).

[111] Raissi, T., Ramdani, N. and Candau, Y. (2004). "Set membership state and parameter estimation for systems described by nonlinear differential equations". *Automatica*, vol. 40, pp. 1771–1777.

[112] Rajyaguru, J. and Chachuat, B. (2013). "Taylor models in deterministic global optimization for Large-Scale systems with few degrees of freedom". *Computer Aided Chemical Engineering*, *23rd European Symposium on Computer Aided Process Engineering*, vol. Volume 32 (A. Kraslawski and I. Turunen, eds.). Elsevier, pp. 973–978.

[113] Rajyaguru, J., Villanueva, M.E., Houska, B. and Chachuat, B. (2014). "Higher-order inclusions of nonlinear systems by chebyshev models". *Proc. AIChE Annual Meeting 2014*. p. 385510.

[114] Raković, S., Munoz-Carpintero, D., Cannon, M. and Kouvaritakis, B. (2012). "Offline tube design for efficient implementation of parameterized tube model predictive control". *Proc. IEEE 51st Annual Conference on Decision and Control*. Maui (HI), pp. 5176–5181.

[115] Raković, S.V. (2009). "Set theoretic methods in model predictive control". *Nonlinear Model Predictive Control: Towards New Challenging Applications*, *Lecture Notes in Control and Information Sciences*, vol. 384. Springer, Heidelberg, pp. 41–54.

[116] Raković, S.V., Kerrigan, E.C., Kouramas, K.I. and Mayne, D.Q. (2005). "Invariant approximation of the minimal robust positively invariant set". *IEEE Trans. Automat. Control*, vol. 50 (3), pp. 406–410.

[117] Raković, S.V., Kouvaritakis, B., Cannon, M. and Panos, C. (2012). "Fully parameterized tube model predictive control". *Internat. J. Robust Nonlinear Control*, vol. 22 (12), pp. 1330–1361.

[118] Raković, S.V., Kouvaritakis, B., Cannon, M., Panos, C. and Findeisen, R. (2012). "Parameterized tube model predictive control". *IEEE Trans. Automat. Control*, vol. 57 (11), pp. 2746–2761.

[119] Ramdani, N., Meslem, N. and Candau, Y. (2009). "A hybrid bounding method for computing an over-approximation for the reachable set of uncertain nonlinear systems". *IEEE Transactions on Automatic Control*, vol. 54 (10), pp. 2352–2364.

[120] Rauh, A., Hofer, E.P. and Auer, E. (2006). "VALENCIA-IVP: A comparison with other initial value problem solvers". *Proceedings of the 12th GAMM-IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics (SCAN'2006)*. Duisburg, Germany, pp. 36–44.

[121] Rauh, A., Westphal, R. and Aschemann, H. (2013). "Verified simulation of control systems with interval parameters using an exponential state enclosure technique". *2013 18th International Conference on Methods and Models in Automation and Robotics (MMAR)*. pp. 241–246.

[122] Rawlings, J.B. and Mayne, D.Q. (2009). *Model Predictive Control: Theory and Design*. Nob Hill Pub.

[123] Rheinboldt, W.C. (2000). "Numerical continuation methods: a perspective". *Journal of Computational and Applied Mathematics*, vol. 124 (1–2), pp. 229–244.

[124] Rubagotti, M., Raimondo, D.M., Ferrara, A. and Magni, L. (2009). "Robust model predictive control of continuous-time sampled-data nonlinear systems with integral sliding mode". *Proceedings of the European Control Conference 2009*. Budapest (HU), pp. 2247–2252.

[125] Sacksteder, R. (1960). "On hypersurfaces with no negative sectional curvatures". *Amer. J. Math.*, vol. 82, pp. 609–630.

[126] Sahlodin, A.M. (2012). *Global Optimization of Dynamic Process Systems using Complete Search Methods*. Ph.D. thesis, McMaster University, Ontario, Canada.

[127] Sahlodin, A.M. and Chachuat, B. (2011). "Convex/concave relaxations of parametric ODEs using Taylor models". *Computers & Chemical Engineering*, vol. 35 (5), pp. 844–857.

[128] Sahlodin, A.M. and Chachuat, B. (2011). "Discretize-then-relax approach for convex/concave relaxations of the solutions of parametric ODEs". *Applied Numerical Mathematics*, vol. 61 (7), pp. 803–820.

[129] Schichl, H. and Neumaier, A. (2005). "Interval analysis on directed acyclic graphs for global optimization". *Journal of Global Optimization*, vol. 33 (4), pp. 541–562.

[130] Scott, J.K. and Barton, P.I. (2010). "Tight, efficient bounds on the solutions of chemical kinetics models". *Computers & Chemical Engineering*, vol. 34 (5), pp. 717 – 731.

[131] Scott, J.K. and Barton, P.I. (2013). "Bounds on the reachable sets of nonlinear control systems". *Automatica*, vol. 49 (1), pp. 93–100.

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

[133] Scott, J.K., Chachuat, B. and Barton, P.I. (2013). "Nonlinear convex and concave relaxations for the solutions of parametric ODEs". *Optimal Control Applications & Methods*, vol. 34 (2), pp. 145–163.

[134] Scott, J.K., Stuber, M. and Barton, P.I. (2011). "Generalized McCormick relaxations". *Journal of Global Optimization*, vol. 51 (4), pp. 569–606.

[135] Sherali, H.D., Dalkiran, E. and Liberti, L. (2012). "Reduced RLT representations for nonconvex polynomial programming problems". *Journal of Global Optimization*, vol. 52 (3), pp. 447–469.

[136] Sherali, H.D. and Fraticelli, B.M.P. (2002). "Enhancing RLT relaxations via a new class of semidefinite cuts". *Journal of Global Optimization*, vol. 22 (1-4), pp. 233–261.

[137] Singer, A.B. (2004). *Global Dynamic Optimization.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

[138] Singer, A.B. and Barton, P.I. (2006). "Global optimization with nonlinear ordinary differential equations". *Journal of Global Optimization*, vol. 34 (2), pp. 159–190.

[139] Smith, E. and Pantelides, C. (1999). "A symbolic reformulation/spatial branch-and-bound algorithm for the global optimisation of nonconvex MINLPs". *Computers & Chemical Engineering*, vol. 23 (4–5), pp. 457–478.

[140] Srinivasan, B., Amrhein, M. and Bonvin, D. (1998). "Reaction and flow variants/invariants in chemical reaction systems with inlet and outlet streams". *AICHE Journal*, vol. 44 (8), pp. 1858–1867.

[141] Streif, S., Kim, K.K.K., Rumschinski, P., Kishida, M., Shen, D.E., Findeisen, R. and Braatz, R.D. (2013). "Robustness analysis, prediction and estimation for uncertain biochemical networks". *Proc. International Symposium on Dynamics and Control of Process Systems (DYCOPS)*. pp. 1–20.

[142] Tawarmalani, M. and Sahinidis, N.V. (2004). "Global optimization of mixed-integer nonlinear programs: A theoretical and computational study". *Mathematical Programming*, vol. 99 (3), pp. 563–591.

[143] Teschl, G. (2012). *Ordinary Differential Equations and Dynamical Systems*. American Mathematical Soc.

[144] Tomlin, C.J. (2011). "Verification and control of hybrid systems using reachability analysis". *19th Mediterranean Conference on Control and Automation (MED)*. Corfu, Greece, p. 150.

[145] Tomlin, C.J., Mitchell, I., Bayen, A.M. and Oishi, M. (2003). "Computational techniques for the verification of hybrid systems". *Proceedings of the IEEE*, vol. 91 (7), pp. 986–1001.

[146] Van Parys, B.P.G., Goulart, P.J. and Morari, M. (2013). "Infinite horizon performance bounds for uncertain constrained systems". *IEEE Trans. Automat. Control*, vol. 58 (11), pp. 2803–2817.

[147] Varaiya, P. and Kurzhanski, A.B. (2006). "Ellipsoidal methods for dynamics and control. Part I". *Journal of Mathematical Sciences*, vol. 139 (5), pp. 6863–6901.

[148] Villanueva, M.E., Houska, B. and Chachuat, B. (2014). "On the stability of set-valued integration for parametric nonlinear {ODEs}". *24th European Symposium on Computer Aided Process Engineering*, *Computer Aided Chemical Engineering*, vol. 33 (J. Klemes, P. Varbanov and P.Y. Liew, eds.). Elsevier, pp. 595 – 600.

[149] Villanueva, M.E., Houska, B. and Chachuat, B. (2015). "Unified framework for the propagation of continuous-time enclosures for parametric nonlinear ODEs". *J. Global Optim.*, vol. 62 (3), pp. 575–613.

[150] Villanueva, M.E., Quirynen, R., Diehl, M., Chachuat, B. and Houska, B. (2016). "Robust mpc via min-max differential inequalities". *Automatica*, ((Submitted)).

[151] Villanueva, M.E., Rajyaguru, J., Houska, B. and Chachuat, B. (2015). "Ellipsoidal arithmetic for multivariate systems". *12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering*, *Computer Aided Chemical Engineering*, vol. 37 (K. Gernaey, J. Huusom and R. Gani, eds.). Elsevier, pp. 767 – 772.

[152] Villanueva, M.E., Paulen, R., Houska, B. and Chachuat, B. (2013). "Enclosing the reachable set of parametric ODEs using taylor models and ellipsoidal calculus". *23rd European Symposium on Computer Aided Process Engineering*, *Computer Aided Chemical Engineering*, vol. 32 (A.K. Turunen and Ilkka, eds.). Elsevier, pp. 979–984.

[153] Walter, E. (1990). "Special issue on parameter identification with error bounds". *Mathematics and Computers in Simulation*, vol. 32 (5–6), pp. 447–607.

[154] Walter, W. (1970). *Differential and integral inequalities*. Springer-Verlag, New York-Berlin. Translated from the German by Lisa Rosenblatt and Lawrence Shampine. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 55.

[155] Wechsung, A., Schaber, S.D. and Barton, P.I. (2014). "The cluster problem revisited". *Journal of Global Optimization*, vol. 58 (3), pp. 429–438.

[156] Yu, S., Reble, M., Chen, H. and Allgöwer, F. (2014). "Inherent robustness properties of quasi-infinite horizon nonlinear model predictive control". *Automatica*, vol. 50 (9), pp. 2269–2280.

[157] Zeilinger, M.N., Raimondo, D.M., Domahidi, A., Morari, M. and Jones, C.N. (2014). "On real-time robust model predictive control". *Automatica*, vol. 50 (3), pp. 683–694.

[158] Zhou, T.S., Zhang, J.J., Yuan, Z.J. and Chen, L.N. (2008). "Synchronization of genetic oscillators". *Chaos*, vol. 18 (3), p. 037126.

[159] Zorn, K. and Sahinidis, N.V. (2014). "Global optimization of general non-convex problems with intermediate bilinear substructures". *Optimization Methods & Software*, vol. 29 (3), pp. 442–462.

Appendix A

On Invariance of the Cycle Time T

In connection to Assumption 3.4, the following result establishes that the cycle time T is independent of t under mild conditions.

Proposition A.1. *Suppose that the right hand-side function f is continuously differentiable and can be rewritten in the form $f(t,x,p) = \hat{f}(\xi(t),x,p)$, where the function ξ is periodic with $\xi(t+T) = \xi(t)$ for all t. Suppose further that there exists a parameter p^* such that the associated limit cycle $\overline{x}(t,p^*)$, defined implicitly by*

$$\overline{x}(t,p^*) := x(t,\overline{x}_0(p^*),p^*) \quad \textit{with } \overline{x}_0 \textit{ such that} \quad \overline{x}_0(p^*) = x(T,\overline{x}_0(p^*),p^*)$$

satisfies condition (3.15)*, that is, is locally asymptotically stable. Then, there exists an open neighborhood $P \subseteq \mathbb{R}^{n_p}$ with $p^* \in P$ and a cycle time $T > 0$ such that the limit cycles $\overline{x}(t,p)$ satisfy* (3.11) *for all $p \in P$. In other words, the cycle time T of the limit cycles is locally independent of p.*

Proof. The proof follows directly by applying the implicit function theorem to the non-linear equation

$$0 \;=\; \overline{x}_0(p^*) - x(T,\overline{x}_0(p^*),p^*)\;,$$

since the associated Jacobian matrix $I - G(T,0,p^*)$ is indeed invertible from the asymptotic stability condition (3.15).

Remark A.1. *Consider Example 3.1. If parameter p_2 is fixed, then the assumptions in Proposition A.1 are satisfied upon choosing $\xi(t) = \sin(p_2 t)$. But if p_2 is allowed to vary, then the function ξ depends on an unknown parameter, which is not allowed in Proposition A.1.*

Appendix B

Technical Lemmata

The following two lemmata are used in the proof of Theorem 4.3. Although variants of these results can be found in the literature [39], we provide short proofs for the sake of completeness.

Lemma B.1. *Let $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function. For any compact set $D \subset \mathbb{R}^n$ and any finite tolerance $\varepsilon > 0$, there exists a smooth function $\varphi_\varepsilon : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\forall x \in D\,, \quad \|\varphi_\varepsilon(x) - \varphi(x)\| \leq \alpha(\varepsilon)\,, \tag{B.1}$$

for some continuous function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ with $\alpha(0) = 0$.

Proof. The proof follows by applying well-known standard analysis techniques [39], and we only summarize the main idea here. Let $\sigma_\varepsilon : \mathbb{R}^n \to \mathbb{R}$, $\varepsilon > 0$ be a family of smooth functions parameterized in $\varepsilon > 0$, such that $\sigma_\varepsilon(x) = 0$ for all x with $\|x\| \geq \varepsilon$ and $\int_{\mathbb{R}^n} \sigma_\varepsilon(x)\,\mathrm{d}x = 1$. Of the alternatives for constructing such a family of 'mollifier' functions, we consider the function

$$\sigma_\varepsilon(x) := \begin{cases} C(\varepsilon) \exp\left(\dfrac{1}{\|x\|^2 - \varepsilon^2}\right) & \text{if } \|x\| < \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

with

$$C(\varepsilon) := \int_{\|x\| \leq \varepsilon} \exp\left(\frac{1}{\|x\|^2 - \varepsilon^2}\right) \mathrm{d}x\,.$$

In turn, the function φ_ε can be defined as the convolution

$$\forall x \in \mathbb{R}^{n_x}, \quad \varphi_\varepsilon(x) \;:=\; \int_{\mathbb{R}^{n_x}} \sigma_\varepsilon(x-y)\varphi(y)\,\mathrm{d}y,$$

which is smooth by construction for any $\varepsilon > 0$. Observe that the function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$\forall \varepsilon \geq 0, \quad \alpha(\varepsilon) \;:=\; \max_{x\in D} \max_{y} \left\{ \|\varphi(x)-\varphi(y)\| \mid \|y-x\| \leq \varepsilon \right\},$$

is is continuous since φ is itself continuous and D is compact, and such that $\alpha(0) = 0$. In particular, this choice of α satisfies the condition (B.1). □

Lemma B.2. *Let $Y : [0,T] \to \mathbb{K}^{n_x}_{\mathrm{C}}$ be a set-valued function such that $V[Y(\cdot)](c)$ is differentiable and $\dot{V}[Y(\cdot)](c)$ is bounded for all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$. Then, there exists a family of functions $g_\varepsilon : [0,T] \times \mathbb{R}^{n_x} \to \mathbb{R}$ parameterized by $\varepsilon \geq 0$, such that $g_\varepsilon(t,\cdot)$ is strictly convex and smooth for all $\varepsilon > 0$ and all $t \in [0,T]$, and the associated sets $Y_\varepsilon(t) := \{x \in \mathbb{R}^{n_x} \mid g_\varepsilon(t,x) \leq 0\}$ satisfy for all $\varepsilon \geq 0$ and all $t \in [0,T]$*

$$\begin{aligned} & Y(t) \subseteq Y_\varepsilon(t) \quad \text{and} \quad d_{\mathrm{H}}(Y(t), Y_\varepsilon(t)) \leq \alpha(\varepsilon)\,, \\ \forall c \in \mathbb{R}^{n_x} \text{ with } c^\top c = 1, \quad & \dot{V}[Y_\varepsilon(t)](c) \geq \dot{V}[Y(t)](c) + L\,\alpha(\varepsilon) \end{aligned} \tag{B.2}$$

for some continuous function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ with $\alpha(0) = 0$, and any constant $0 \leq L < \frac{1}{T}$.

Proof. A proof can be obtained by passing through two steps.

S1 We start with any smooth function $\nu_\varepsilon(t,\cdot) : \mathbb{R}^{n_x} \to \mathbb{R}$ such that

$$\begin{aligned} & \forall \varepsilon > 0,\ \forall c \in \mathbb{R}^{n_x} \text{ with } c^\top c = 1,\ \forall t \in [0,T]\,: \\ & \quad \nu_\varepsilon(t,c) \geq \dot{V}[Y(t)](c) \quad \text{and} \quad \left\| \nu_\varepsilon(t,c) - \dot{V}(t,c) \right\| \leq \alpha_1(\varepsilon)\,, \end{aligned}$$

for some continuous function $\alpha_1 : \mathbb{R}_+ \to \mathbb{R}_+$ with $\alpha_1(0) = 0$. Such a function is guaranteed to exist by Lemma B.1. Then, we define the set-valued function $Z_\varepsilon : [0,T] \to \mathbb{K}^{n_x}_{\mathrm{C}}$ such that

$$\forall c \in \mathbb{R}^{n_x} \text{ with } c^\top c = 1, \quad V[Z_\varepsilon(t)](c) \;:=\; V[Y_(0)](c) + \int_0^t \nu_\varepsilon(\tau,c)\,\mathrm{d}\tau\,.$$

The following properties hold by construction of Z_ε, for every $\varepsilon > 0$:

a) For all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$, the function $V[Z_\varepsilon(\cdot)](c)$ is differentiable on $[0,T]$, and we have

$$\forall c \in \mathbb{R}^{n_x} \text{ with } c^\top c = 1,\ \forall t \in [0,T]\,, \quad \dot{V}[Z_\varepsilon(t)](c) \ \geq\ \dot{V}[Y(t)](c)\,.$$

b) $d_{\mathrm{H}}(Z_\varepsilon(t), Y(t)) \leq T\alpha_1(\varepsilon)$, by property (2).

S2 We construct the set-valued function

$$Y_\varepsilon(t) \ :=\ Z_\varepsilon(t) \oplus [T\alpha_1(\varepsilon) + tL\alpha(\varepsilon)]\,\mathcal{B}^{n_x} \quad \text{with} \quad \alpha(\varepsilon) \ :=\ \frac{2T}{1-TL}\alpha_1(\varepsilon)\,,$$

with $0 \leq L < \frac{1}{T}$. Note that the function α is continuous and non-negative and it satisfies $\alpha(0) = 0$ by definition. Therefore, we have $Y_\varepsilon(t) \supseteq Y(t)$ since $Y_\varepsilon(t) \supseteq Z_\varepsilon(t)$ and $d_{\mathrm{H}}(Y_\varepsilon(t), Z_\varepsilon(t)) \geq d_{\mathrm{H}}(Z_\varepsilon(t), Y(t))$. It follows from Property a) that

$$\begin{aligned}\dot{V}[Y_\varepsilon(t)](c) &= \dot{V}[Z_\varepsilon(t)](c) + L\alpha(\varepsilon) \\ &\geq \dot{V}[Y(t)](c) + L\alpha(\varepsilon)\,,\end{aligned}$$

for all $t \in [0,T]$, all $c \in \mathbb{R}^{n_x}$ with $c^\top c = 1$, and all $\varepsilon \geq 0$. Moreover, by Property b), we have

$$\begin{aligned}d_{\mathrm{H}}(Y_\varepsilon(t), Y(t)) &\leq d_{\mathrm{H}}(Z_\varepsilon(t), Y(t)) + \alpha_0(\varepsilon) + T\alpha_1(\varepsilon) + TL\alpha(\varepsilon) \\ &\leq 2\alpha_0(\varepsilon) + 2T\alpha_1(\varepsilon) + TL\alpha(\varepsilon) \\ &= \alpha(\varepsilon)\,,\end{aligned}$$

for all $t \in [0,T]$, and all $\varepsilon \geq 0$. Finally, by Theorem 1 in [9], there exists a functions $g_\varepsilon : [0,T] \times \mathbb{R}^{n_x} \to \mathbb{R}$ such that $Y_\varepsilon(t) =: \{x \in \mathbb{R}^{n_x} \mid g_\varepsilon(t,x) \leq 0\}$ and $g_\varepsilon(t,\cdot)$ is convex and smooth for all $\varepsilon \geq 0$ and all $t \in [0,T]$. In order for $g_\varepsilon(t,\cdot)$ to be strictly convex for all $\varepsilon > 0$, one can always add a strictly convex and smooth term of order $\mathbf{O}\,(()\,\varepsilon)$ that is negative on the compact sets $\bigcup_{t\in[0,T]} Y_\varepsilon(t)$. □

The following lemma is used in the proof of Theorem 4.4. The result allows one to bound the solution of a particular parametric differential inequality and can be regarded as a generalization of Gronwall's lemma [44].

Lemma B.3. *Let $v \in \mathbb{R}_+$ and let $u : [0,T] \to \mathbb{R}$ be a Lipschitz-continuous function satisfying the parametric differential inequality*

$$a.e.\ t \in [0,T]\,, \quad \frac{\mathrm{d}}{\mathrm{d}t}[u(t)]^n \;\leq\; n \sum_{i=0}^{n} L_i\, u(t)^{n-i} v^{im} \quad \textit{with} \quad u(0) \leq C_0 v^m\,, \tag{B.3}$$

for some integers $m,n \geq 1$ and a set of constants $0 \leq L_0, \ldots, L_n < \infty$ and $C_0 \geq 1$. Then, $u(t) \leq C_0 \exp\left(\sum_{i=0}^{n} L_i t\right) v^m$, for all $t \in [0,T]$.

Proof. The proof proceeds in two steps. It is assumed first that the function u is differentiable on $[0,T]$. Then, it is argued that the result still holds in extending this class of functions to Lipschitz-continuous.

Assuming that u is differentiable on $[0,T]$ and discretizing the differential inequality (B.3) with a step-size $h := \frac{T}{N}$ for a large enough $N \in \mathbb{N}$ gives

$$\forall k \in \{0, N-1\}\,, \quad [u((k+1)h)]^n \leq [u(kh)]^n + hn \sum_{i=0}^{n} L_i\, (u(kh))^{n-i} v^{im} + h\,\alpha(h)\,,$$

for some continuous function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ with $\alpha(0) = 0$. Now, supposing that $u(kh) \leq C_k v^m$ with $C_k \geq 1$, we have

$$[u((k+1)h)]^n \leq (C_{k+1})^n v^{nm} + h\,\alpha(h) \quad \text{with} \quad (C_{k+1})^n := \left(1 + hn \sum_{i=0}^{n} L_i\right) (C_k)^n \geq 1\,.$$

In particular, the definition of $(C_{k+1})^n$ uses the result that

$$(C_k)^n + hn \sum_{i=0}^{n} L_i (C_k)^{n-i} \;\leq\; \left(1 + hn \sum_{i=0}^{n} L_i\right) (C_k)^n\,,$$

for all $C_k \geq 1$. It follows by induction that $u(kh) \leq C_k v^m + h\,\alpha(h)$ for each $k = 0, \ldots, N$, with

$$C_k \;=\; \left(1 + hn \sum_{i=0}^{n} L_i\right)^{k/n} C_0\,.$$

Let $\bar{t} \in [0,T]$ be such that $\bar{t} := \frac{k_0}{N_0} T$ for given $0 \leq k_0 \leq N_0$, and consider the sequence $\{\overline{C}_j\}$ given by

$$\overline{C}_j \;:=\; \left(1 + \frac{nT}{jN_0} \sum_{i=0}^{n} L_i\right)^{jk_0/n} C_0\,,$$

so that $u(\bar{t}) \leq \overline{C}_j\, v^m + \frac{T}{jN_0}\,\alpha(\frac{T}{jN_0})$ for all $j \geq 1$. It follows from the definition of the exponential function as $\exp(x) := \lim_{j\to\infty}(1+\frac{x}{j})^j$ that this sequence is convergent, and we have

$$\lim_{j\to\infty} \overline{C}_j \;=\; \exp\left(\sum_{i=0}^{n} L_i \bar{t}\right) C_0 \,.$$

As u is continuous on $[0,T]$, and since the rationals are a dense subset of the real numbers, it follows that

$$\forall t \in [0,T]\,, \quad u(t) \;\leq\; C(t)\, v^m \quad \text{with} \quad C(t) := \exp\left(\sum_{i=0}^{n} L_i t\right) C_0 \,.$$

In a second step, the assumption of differentiability for u can be relaxed to Lipschitz-continuity, by a similar argument as in part S3 of the proof of Theorem 4.3, namely that any (locally) Lipschitz-continuous function is differentiable almost everywhere. □

The following lemma enables the construction of inner approximation of centered ellipsoids. Similar constrictions have been derived by Kurzhanski and co-workers [see, e.g. 69], but we provide a simple construction for the sake of consistency.

Lemma B.4. *Let $z \in \mathcal{E}(q_z, Q_z) \in \mathbb{K}_C^{n_z}$, and let*

$$R_z = (1-\gamma) Q_z + \left(1 - \frac{1}{\gamma}\right)(z-q_z)(z-q_z)^\mathsf{T} \,,$$

for some $\gamma \in (0,1)$. If $R_z \in \mathbb{S}_+^{n_z}$, then $\mathcal{E}(z,R_z) \subseteq \mathcal{E}(q_z,Q_z)$ or, equivalently $\mathcal{E}(R_z) \subseteq \mathcal{E}(z-q_z,Q_z)$.

Proof. Let us introduce the rank-1 ellipsoid $\mathcal{E}\left(q_z, (z-q_z)(z-q_z)^\mathsf{T}\right)$. Notice that if

$$\mathcal{E}\left(q_z, (z-q_z)(z-q_z)^\mathsf{T}\right) \oplus \mathcal{E}(R_z) \subseteq E(q_z, Q_z) \tag{B.4}$$

then $\mathcal{E}(z,R_z) \subseteq \mathcal{E}(q_z,Q_z)$. Using the standard formula for the ellipsoidal bounding of the Minkowski sum of ellipsoids [68], we find that (B.4) holds if

$$Q_u = \frac{1}{\gamma}(z-q_z)(z-q_z)^\mathsf{T} + \frac{1}{1-\gamma} R_z \,, \tag{B.5}$$

for any $\gamma \in (0,1)$. Solving Eq. (B.5) for R_z gives

$$R_z = (1-\gamma) Q_z + \left(1 - \frac{1}{\gamma}\right)(z-q_z)(z-q_z)^\mathsf{T} \,,$$

which proves the first result. The second claim follows trivially. □

The following lemma enables the construction of smooth nonlinearity bounders for twice-continuously differentiable functions.

Lemma B.5. *Consider a twice-continuously-differentiable function $g : \mathbb{R}^{n_y} \to \mathbb{R}^{n_y}$, and define the remainder function $n : \mathbb{R}^{n_y} \to \mathbb{R}^{n_y}$ such that*

$$g(y) = g(q_y) + \frac{\partial g}{\partial y}(q_y)\delta_y + n(\delta_y),$$

with $\delta_y := y - q_y$. Let $D_y \in \mathbb{K}^{n_y}$ and $(q_y, Q_y) \in \mathbb{R}^{n_y} \times \mathbb{S}_+^{n_y}$ such that $\mathcal{E}(q_y, Q_y) \subseteq D_y$. Suppose that there exist constants $\bar{F}_1, \ldots, \bar{F}_{n_y} \in \mathbb{R}_+$ satisfying

$$\forall y \in \mathcal{D}_y, \quad \bar{F}_i \geq \left\| \frac{\partial^2 g_i}{\partial y^2}(y)\, S_i \right\|_{\mathrm{F}},$$

for certain invertible matrices $S_1, \ldots, S_{n_y} \in \mathbb{R}^{n_y \times n_y}$. Then, $n(\delta_y) \in \mathcal{E}(Q_n)$, for all $\delta_y \in \mathcal{E}(Q_y)$ with

$$Q_n := \frac{1}{4} \operatorname{diag}\left(\bar{F}_i^2 \left\| S_i^{-1} Q_y \right\|_{\mathrm{F}}^2 \right)_{1 \leq i \leq n_y}.$$

Proof. From Taylor's theorem, the remainder function n_i corresponding to g_i, for each $i = 1, \ldots, n_y$, is given by

$$n_i(\delta_y) = \frac{1}{2} \delta_y^{\mathsf{T}} \frac{\partial^2 g_i}{\partial y^2}(\xi_i)\, \delta_y,$$

for some $\xi_i \in \operatorname{conv}(\{y, q_y\})$. Then, for all $\delta_y \in \mathcal{E}(Q_y)$, we have

$$\begin{aligned}
n_i(\delta_y) &= \frac{1}{2} \operatorname{Tr}\left(\frac{\partial^2 g_i}{\partial y^2}(\xi_i)\, S_i S_i^{-1} \delta_y \delta_y^{\mathsf{T}} \right) \\
&\leq \frac{1}{2} \operatorname{Tr}\left(\frac{\partial^2 g_i}{\partial y^2}(\xi_i)\, S_i S_i^{-1} Q_y \right) \\
&= \frac{1}{2} \left\| \frac{\partial^2 g_i}{\partial y^2}(\xi_i)\, S_i \left(S_i^{-1} Q_y \right) \right\|_{\mathrm{F}} \\
&\leq \frac{1}{2} \left\| \frac{\partial^2 g_i}{\partial y^2}(\xi_i)\, S_i \right\|_{\mathrm{F}} \left\| S_i^{-1} Q_y \right\|_{\mathrm{F}} \\
&\leq \frac{1}{2} \bar{F}_i \left\| S_i^{-1} Q_y \right\|_{\mathrm{F}}.
\end{aligned}$$

Therefore, n is bounded on $\mathcal{E}(Q_y)$ by an ellipsoid centered at the origin and with semi-axes of length $\frac{1}{2}\bar{F}_i \| S_i^{-1} Q_y \|_{\mathrm{F}}$. □

Appendix C

Generalized Rotational Inertia For Ellipsoidal Sets

For a set $X \in \mathbb{K}^n$ the generalized rotational inertia (GRI) with respect to a reference $x_{\text{ref}} \in \mathbb{R}^n$ is given by

$$\ell(X) = \frac{\int_X (x - x_{\text{ref}})^\intercal D (x - x_{\text{ref}}) \, \mathrm{d}x}{\int_X 1 \, \mathrm{d}x}, \tag{C.1}$$

with $D \in \mathbb{S}^n_+$ denoting any weighting matrix.

In the following we present a derivation for an explicit expresion of the GRI of ellipsoidal sets.

Proposition C.1. *Let $X := \mathcal{E}(q, Q)$ with $q \in \mathbb{R}^n$ and $Q \in \mathbb{S}^n_+$, then the generalized rotational inertia of X with respect to a reference point $x_{\text{ref}} \in \mathbb{R}^n$ and a weighting matrix $D \in \mathbb{S}^n_+$ is given by*

$$\begin{aligned} \ell(\mathcal{E}(q,Q)) &= \frac{\int_{\mathcal{E}(q,Q)} (x - x_{\text{ref}})^\intercal D (x - x_{\text{ref}}) \, \mathrm{d}x}{\int_{\mathcal{E}(q,Q)} 1 \, \mathrm{d}x} \\ &= (q - x_{\text{ref}})^\intercal D (q - x_{\text{ref}}) + \frac{n}{n+2} \operatorname{Tr}(DQ) \ . \end{aligned} \tag{C.2}$$

Before proceeding with the proof of Proposition C.1, we will introduce the following trivial but useful identity

Lemma C.1. *Let $A \in \mathbb{R}^{n \times n}$ and $x \in \mathcal{B}^n$, we have*

$$\int_{\mathcal{B}^n} x^\intercal A x \, \mathrm{d}x = \frac{V_n}{n+2} \operatorname{Tr}(A) \tag{C.3}$$

Proof. First, notice that the volume of the unit ball in hyperspherical coordinates can be written as

$$\begin{aligned} V_n &= \int_0^1 \int_{\mathcal{S}^{n-1}} r^{n-1} \mathrm{d}r \, |J(\sigma)| \, \mathrm{d}\sigma \\ &= \frac{1}{n} \int_{\mathcal{S}^{n-1}} |J(\sigma)| \, \mathrm{d}\sigma \, . \end{aligned}$$

where $|J(\sigma)| \, \mathrm{d}\sigma$ includes all the angle contributions and the determinant of the Jacobian of the coordinate transformation. Alternatively, the last identity can be expressed as

$$nV_n = A_n = \int_{\mathcal{S}^{n-1}} |J(\sigma)| \, \mathrm{d}\sigma \, . \tag{C.4}$$

Decomposing the quadratic form $x^\intercal A x$, we have

$$\begin{aligned} \int_{\mathcal{B}^n} x^\intercal A x \, \mathrm{d}x &= \int_{\mathcal{B}^n} \sum_{i=1}^n x_i \left(\sum_{j=1}^n A_{i,j} x_j \right) \mathrm{d}x \\ &= \int_{\mathcal{B}^n} \sum_{i=1}^n x_i \left(\sum_{j=i}^n A_{i,j} x_j + \sum_{j \neq i}^n A_{i,j} x_j \right) \mathrm{d}x \\ &= \int_{\mathcal{B}^n} \sum_{i=1}^n x_i \sum_{j=i}^n A_{i,j} x_j \, \mathrm{d}x + \int_{\mathcal{B}^n} \sum_{i=1}^n x_i \sum_{j \neq i}^n A_{i,j} x_j \, \mathrm{d}x \end{aligned}$$

By symmetry of the unit ball the second integrand vanishes, hence we have

$$\begin{aligned} &= \frac{1}{n} \int_{\mathcal{B}^n} \sum_{i=1}^n A_{i,i} x_i^2 \, \mathrm{d}x \\ &= \frac{\sum_{i=1}^n A_{i,i}}{n} \int_{\mathcal{B}^n} x_i^2 \, \mathrm{d}x \end{aligned}$$

Where the last equality follows from the application of Fubini's Theorem, as the integral is bounded. Rewritting the last integral in hyperspherical coordinates we get

$$\begin{aligned} &= \frac{1}{n} \sum_{i=1}^n A_{i,i} \int_0^1 \int_{\mathcal{S}^{n-1}} r^2 r^{n-1} \, \mathrm{d}r \, |J(\sigma)| \, \mathrm{d}\sigma \\ &= \frac{1}{n} \sum_{i=1}^n A_{i,i} \int_0^1 \int_{\mathcal{S}^{n-1}} r^{n+1} \, \mathrm{d}r \, |J(\sigma)| \, \mathrm{d}\sigma \\ &= \frac{1}{n(n+2)} \operatorname{Tr}(A) \int_{\mathcal{S}^{n-1}} |J(\sigma)| \, \mathrm{d}\sigma \end{aligned}$$

substituting Eq. (C.4)

$$= \frac{V_n}{n+2} \operatorname{Tr}(A)$$

which proves the identity . □

Now, let us proceed with the proof of Proposition C.1

Proof. Using the change of variables $x = q + Q^{\frac{1}{2}} v$ we can change the domain of integration from an ellipsoid to an unit ball, obtaining

$$\begin{aligned}
\ell(\mathcal{E}(q,Q)) &= \frac{\displaystyle\int_{\mathcal{E}(q,Q)} (x - x_{\text{ref}})^{\mathsf{T}} D (x - x_{\text{ref}}) \, \mathrm{d}x}{\displaystyle\int_{\mathcal{E}(q,Q)} 1 \, \mathrm{d}x} \\
&= \frac{\displaystyle\int_{\mathcal{B}_n} \left((q - x_{\text{ref}}) + Q^{\frac{1}{2}} v \right)^{\mathsf{T}} \left(D(q - x_{\text{ref}}) + D Q^{\frac{1}{2}} v \right) \operatorname{Det}(Q^{\frac{1}{2}}) \, \mathrm{d}v}{\displaystyle\int_{\mathcal{B}_n} \operatorname{Det}(Q^{\frac{1}{2}}) \, \mathrm{d}v} \\
&= \frac{(q - x_{\text{ref}})^{\mathsf{T}} D (q - x_{\text{ref}}) \displaystyle\int_{\mathcal{B}_n} \mathrm{d}v + 2 \int_{\mathcal{B}_n} (q - x_{\text{ref}})^{\mathsf{T}} D v \, \mathrm{d}v + \int_{\mathcal{B}_n} v^{\mathsf{T}} D Q v \, \mathrm{d}v}{V_n} ,
\end{aligned}$$

at this point we notice that the second integral in the numerator vanishes due to symmetry of the domain of integration and the integrand being odd. Integrating the constant function in the first integral and applying Identity (C.3) to the third integral with $A := DQ$ yields

$$\ell(\mathcal{E}(q,Q)) = (q - x_{\text{ref}})^{\mathsf{T}} D (q - x_{\text{ref}}) + \frac{1}{n+2} \operatorname{Tr}(DQ) \ ,$$

which proves the proposition. □

Appendix D

Brief Tutorial Overview of the MC++ and CRONOS Libraries

The purpose of this appendix is to relate the theoretical concepts and constructions presented in this Thesis to their implementation in the MC++ (http://omega-icl.bitbucket.org/mcpp/) and CRONOS (http://omega-icl.bitbucket.org/cronos/) libraries.

MC++ is an open-source C++ library for the construction of enclosures for the image-set of factorable functions. It is a header-only library developed by the OMEGA (Optimization Methods for Green Applications) group at the Centre for Process Systems Engineering (Imperial College London). It provides a collection of classes for constructing enclosures using different arithmetics such as interval, McCormick relaxation, eigenvalue, ellipsoidal and polynomial (Taylor and Chebyshev) model arithmetics.

The MC++ library makes use of operator overloading and templates. This makes the evaluation of set-extensions as intuitive as evaluating functions in real numbers. Adopting a template-based design facilitates the use of other arithmetic implementations, e.g. interval libraries such as PROFIL (http://www.ti3.tu-harburg.de/profil_e) or FILIB++ (http://www2.math.uni-wuppertal.de/~xsc/software/filib.html) can be used in place of the non-verified interval library. Apart from these optional dependencies, MC++ relies on the LAPACK (http://www.netlib.org/lapack/) and BLAS (http://www.netlib.org/blas/) libraries and the CPPLAPACK wrapper (http://www.cpplapack.sourceforge.net/) for linear algebra computations and FADBAD++ (http://www.fadbad.com/fadbad.html) for automatic differentiation.

CRONOS (Complete seaRch sOlutions for NOnlinear Systems) is an open-source C++ library for the analysis, estimation and optimization of nonlinear systems using complete-search methods. Like MC++ is a header-only library developed within the OMEGA group

and provides a collection of classes tailored for analysing different aspects of nonlinear systems.

The CRONOS library, relies on a number of libraries to work

- MC++ library for constructing and bounding factorable functions
- GSL (GNU Scientific Library, http://www.gnu.org/software/gsl/) or SUNDIALS (SUite of Nonlinear and DIfferential/ALgebraic equation Solvers, https://computation.llnl.gov/casc/sundials/main.html) integrators for set propagation through ODEs
- MC13 and MC21 from the Harwell Soubrutine Library (http://www.hsl.rl.ac.uk/) for sparse matrix computations used in implicit equation solvers
- CPLEX (http://www-03.ibm.com/software/products/en/ibmilogcpleoptistud) or GUROBI (http://www.gurobi.com/) optimizers for solving linear programming problems involved in domain reduction routines
- IPOPT (Interior Point OPTimizer, https://projects.coin-or.org/Ipopt) for local NLP optimization.

Finally, this tutorial is by no means an exhaustive guide of the libraries. Instructions to obtain, install and use MC++ and CRONOS can be found at (http://omega-icl.bitbucket.org/mcpp/) and (http://omega-icl.bitbucket.org/cronos/ respectively. The main objective of this appendix is–as stated earlier– to relate the theoretical concepts presented in this thesis to their implementation and use within the MC++ and CRONOS libraries.

The next section illustrates how to construct bounds for a factorable function using MC++.

D.1 Bounding Factorable Functions Using MC++

The first building block for bounding the image set of a factorable function is a way of constructing and representing such function. As defined in Chapter 2, a function is factorable whenever it can be represented recursively using a finite number of so called atom operations. These operations are specified in a computational list and includes binary sums, binary products and univariate operations such as square root, power, logarithms, and trigonometric functions.

In MC++, factorable functions are stored and manipulated using the classes contained in the `ffunc.hpp` header file. Figure D.1 depicts a minimal working example to represent

the function

$$f(x) = \begin{pmatrix} \sqrt{x_0 + x_1} + x_0 x_1 \\ (x_0 - x_1)^2 + 3x_1 \end{pmatrix} . \tag{D.1}$$

where $x =^{\mathsf{T}} (x_0 x_1)$.

Construction of Directed Acyclic Graph. Firstly, factorable functions are stored and manipulated by the `mc::FFGraph` environment. Variables participating in a factorable function (`X[0]`, `X[1]` for independent and `F[0]`, `F[1]` for dependent in our example) are represented by the `mc::FFVar` data type. MC++ makes use of operator overloading of the atom operations for most data types –in particular `mc::FFVar`. This allows to obtain dependent variables `F[0]` , `F[1]` as the result of some atom operation overloaded on `X[0]`, `X[1]`.

The decomposition of a factorable function into its atoms can be represented in a variety of ways. For example using the lifted representation (cf. Section 2.2), the decomposition of Equation (D.1) into its factors can be written as:

$$u^0(x) = x \tag{D.2a}$$

$$u^1(x) = g_1\left(u^0(x)\right) = \begin{pmatrix} u^0(x) \\ a_1(u^0(x)) \end{pmatrix} = \begin{pmatrix} u^0(x) \\ u_0^0(x) u_1^0(x) \end{pmatrix} \tag{D.2b}$$

$$u^2(x) = g_2\left(u^1(x)\right) = \begin{pmatrix} u^1(x) \\ a_2(u^1(x)) \end{pmatrix} = \begin{pmatrix} u^1(x) \\ u_0^1(x) + u_1^0(x) \end{pmatrix} \tag{D.2c}$$

$$u^3(x) = g_3\left(u^2(x)\right) = \begin{pmatrix} u^2(x) \\ a_3(u^2(x)) \end{pmatrix} = \begin{pmatrix} u^2(x) \\ \sqrt{u_3^2(x)} \end{pmatrix} \tag{D.2d}$$

$$u^4(x) = g_4\left(u^3(x)\right) = \begin{pmatrix} u^3(x) \\ a_4(u^3(x)) \end{pmatrix} = \begin{pmatrix} u^3(x) \\ u_2^3(x) + u_4^3 \end{pmatrix} \tag{D.2e}$$

$$u^5(x) = g_5\left(u^4(x)\right) = \begin{pmatrix} u^4(x) \\ a_5(u^4(x)) \end{pmatrix} = \begin{pmatrix} u^4(x) \\ 3u_1^4(x) \end{pmatrix} \tag{D.2f}$$

$$u^6(x) = g_6\left(u^5(x)\right) = \begin{pmatrix} u^5(x) \\ a_6(u^5(x)) \end{pmatrix} = \begin{pmatrix} u^5(x) \\ u_0^5(x) - u_1^5(x) \end{pmatrix} \tag{D.2g}$$

$$u^7(x) = g_7\left(u^6(x)\right) = \begin{pmatrix} u^6(x) \\ a_7(u^6(x)) \end{pmatrix} = \begin{pmatrix} u^6(x) \\ \left(u_7^6(x)\right)^2 \end{pmatrix} \tag{D.2h}$$

```
#include <fstream>
#include "ffunc.hpp"

int main()
{
/////////  Representing a Factorable Function  /////////

  // Define enviroment to store factorable function FF
  mc::FFGraph FF;
  // Define independent vars X participating in FF
  const unsigned int NX = 2; // No. of independent vars
  mc::FFVar X[NX];
  // Set independent vars X in factorable function FF
  for( unsigned int i=0; i<NX; i++ ) X[i].set( &FF );
  // Define dependent variables F participating in FF
  const unsigned int NF = 2; // No. of dependent vars
  mc::FFVar F[NF] = { mc::sqrt(X[0]+X[1]) + X[0]*X[1],
                      mc::pow(X[0]-X[1],2) + 3*X[1] };
  // Output FF decomposition represented as code list
  std::cout << FF << std::endl;
  // Output FF decomposition represented as DAG
  std::ofstream output_F( "F.dot", std::ios_base::out );
  FF.dot_script( NF, F, output_F ); // Generates DOT fig
  output_F.close();

  return 0;
}
```

Fig. D.1 Minimal working code for representing a factorable function (Eq. D.1) in MC++.

$$u^8(x) = g_8\left(u^7(x)\right) = \begin{pmatrix} u^7(x) \\ a_8(u^7(x)) \end{pmatrix} = \begin{pmatrix} u^7(x) \\ u_6^7(x) + u_8^7(x) \end{pmatrix}. \tag{D.2i}$$

Notice that $u_j^i(x)$ denotes the j-th component of the i-th iteration and that we are adopting for the time being the C++ index notation for vectors–indices starting with 0.

In MC++ the decomposition of a factorable function is represented as a directed acyclic graph or as a code list. The code list representation is obtained by using the (overloaded) standard output operator (`std::out «`) on the `mc::FFGraph` object. The directed acyclic graph representation is through the `dot_script` method in the `mc::FFGraph` class. The `dot_script` method takes as arguments the array of dependent variables (`F`), its dimension (*NF*) and an output stream (`std::ofstream`) object (`output_F` in this example) and re-

turns a DOT script file. The directed acyclic graph (DAG) can then visualized using the DOT program. Figure D.2 shows both the code list output and the DAG obtained using MC++ and DOT. Here, we can see that MC++ introduces auxiliary variables `Z1`,...,`Z8` to represent the the atom operations in the factored decomposition. We can see that the three representations are equivalent by associating the auxiliary variables in both the code list and DAG representations to the atom operations $a_i(u_j^{i-1})$ in the lifted representation.

```
DAG VARIABLES:
  X0  =>  { Z0 Z1 Z6 }
  X1  =>  { Z0 Z1 Z5 Z6 }

DAG INTERMEDIATES:
  Z0 <= X0 * X1      => { Z3 }
  Z1 <= X0 + X1      => { Z2 }
  Z2 <= SQRT( Z1 )   => { Z3 }
  Z3 <= Z0 + Z2      => { }
  Z5 <= X1 * Z4      => { Z8 }
  Z6 <= X0 - X1      => { Z7 }
  Z7 <= SQR( Z6 )    => { Z8 }
  Z8 <= Z5 + Z7      => { }
  Z4 <= 3(D)         => { Z5 }
```

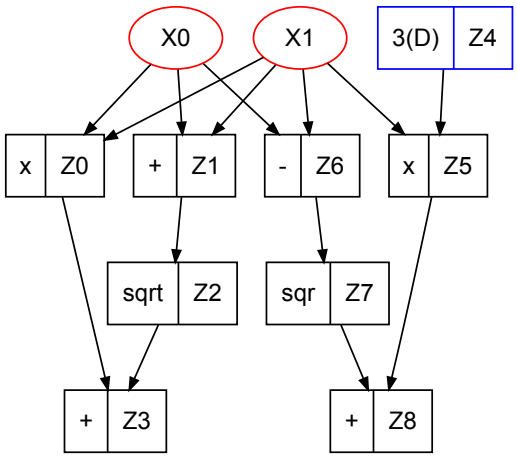

Fig. D.2 Code list (left) and DAG (right) representation of a factorable function using MC++.

Once the factorable function has been defined in MC++ it can also be evaluated in different arithmetics. As an example, we will evaluate Eq. (D.1) in interval, Chebyshev model and ellipsoidal arithmetic.

DAG Evaluation in Interval Arithmetic. Figure D.3 presents an MC++ code snippet that can be added to that presented in Fig. D.1 for the evaluation of a factorable function in interval arithmetic. In this example evaluation is done using the MC++ interval class (`mc::Interval`) and requires to include the `interval.hpp` header file. Evaluation in intervals is very intuitive, first intervals for the independent variables are instantiated using the class constructor with the lower and upper endpoints as arguments. Then, interval variables are defined for the dependent variables and finally the `eval` method of the `mc::FFGraph` class is called to perform the evaluation. This method is templated, so different arithmetics can be used in place of intervals. Figure D.4 presentes the output of the code after compiling and running.

```
////////////  Evaluation of FF in Intervals  //////////

  // Define interval domain for independent variables
  mc::Interval IX[NX] = { mc::Interval(1.585,4.415),
                          mc::Interval(2.585,5.415) };
  // Define interval variables to hold image set
  mc::Interval IF[NF];
  // Evaluation of F in intervals (results stored in IF)
  FF.eval(NF,F,IF,NX,X,IX);
  // Output results
  std::cout << "Interval enclosure of F" << std::endl;
  for( unsigned int i=0; i<NF; i++ )
        std::cout << "IF[" << i << "]= "<< IF[i] <<"\n";
  std::cout << std::endl;
```

Fig. D.3 Code snippet for the evaluation of a factorable function in interval arithmetic using the MC++ interval class.

```
Interval enclosure of F
IF[0]= [  6.13928e+00 :  2.70425e+01 ]
IF[1]= [  7.75500e+00 :  3.09139e+01 ]
```

Fig. D.4 Example of an interval enclosure as produced by the evaluation of a factorable function using MC++.

DAG Evaluation in Chebyshev Model Arithmetic. Evaluating a factorable function in Chebyshev model arithmetic using MC++ can be achieved through the templated classes `mc::CModel<T>` and `mc::CVar<T>` contained in the `cmodel.hpp` header file (See Figure D.5 for a small code snippet). From the perspective of an MC++ user, the main differences between interval and Chebyshev model arithmetic is the fact that the latter has both a model environment and a lightweight variable class –while in the former everything is contained in the same class. In our example, the Chebyshev model environment is constructed with `mc::Interval` as a template parameter and requires the number of independent variables (`NX`) and the order (2 in our case) as parameters. Chebyshev variables for the independent variables and their interval domains need to be associated to the Chebyshev model environment using the `set` method. Once this hass been done and Chebyshev variables for the dependent variables have been created the evaluation can proceed as in the interval case.

Figure D.6 shows the output of the evaluation of Eq. (D.1) in Chebyshev model arithmetic. The output for each dependent variable (`CVF[0]` and `CVF[1]`) is a list of coefficients

```
////////  Evaluation of FF in Chebyshev Models  ////////

  // Define Chebyshev model environment
  mc::CModel<mc::Interval> CM( NX, 2 );
  // Define Chebyshev independent variables
  mc::CVar<mc::Interval> CVX[NX];
  for( unsigned int i=0; i<NX; i++ )
        CVX[i].set( &CM, i, IX[i] );
  // Define Chebyshev dependent variables
  mc::CVar<mc::Interval> CVF[NF];
  // Evaluation of F in Chebyshev model arithmetic
  // (results stored in CVF)
  FF.eval(NF,F,CVF,NX,X,CVX);
  //Output results
  std::cout << "Chebyshev model enclosure of F" << "\n";
  for( unsigned int i=0; i<NF; i++ )
        std::cout <<"CVF[" << i << "]:"<< CVF[i] <<"\n";
```

Fig. D.5 Code snippet for the evaluation of a factorable function in Chebyshev model arithmetic using MC++.

(in this case `a0` to `a5`) together with the interval remainder (`R`) and the interval bounds for the model (`B`). Next to the numerical values of the coefficients are `NX` columns with integers from 0 up to the order of the model (2 in our case). Let $T_i(x_j)$ be the ith Chebyshev polynomial on the jth variable, then column number denotes the index of the variable while the integer in that colum denotes the Chebyshev polynomial. For example the line `a2 = 5.93174e+00 1 0` denotes a term of the form $a_2 T_1(x_0)$ in the polynomial model.

DAG Evaluation in Ellipsoidal Arithmetic. The first step towards the evaluation in ellipsoidal arithmetic is to define the parameters for the ellipsoidal domain where the independent variables lie. This is achieved by defining the center vector and the shape matrix through `CPPL::dcovector` and `CPPL::dsymatrix` objects. Then, the ellipsoidal image environment is defined using the templated class `mc::EllImg<T>` from the `ellimage.hpp` header file. Since each auxiliary variable has an interval range, the template parameter should be an interval class (in this case, the `mc::Interval`) class provided with MC++. The ellipsoidal domain is then associated to the ellipsoidal image object using the `set` method. Independent and dependent (ellipsoidal) variables are created using the templated class `mc::EllVar<T>` (also contained in `ellimage.hpp`). The `mc::EllVar<T>` class uses also an interval class as the template parameter. Its usage is similar to that of the `CVar<T>`

```
Chebyshev model enclosure of F
CVF[0]:
   a0     =  1.46317e+01      0  0
   a1     =  4.51674e+00      0  1
   a2     =  5.93174e+00      1  0
   a3     = -7.03380e-03      0  2
   a4     =  1.97409e+00      1  1
   a5     = -7.03380e-03      2  0
   R      =  [ -3.93409e-03 :  3.93409e-03 ]
   B      =  [  2.19110e+00 :  2.70441e+01 ]

CVF[1]:
   a0     =  1.50022e+01      0  0
   a1     =  7.07500e+00      0  1
   a2     = -2.83000e+00      1  0
   a3     =  1.00111e+00      0  2
   a4     = -4.00445e+00      1  1
   a5     =  1.00111e+00      2  0
   R      =  [  0.00000e+00 :  0.00000e+00 ]
   B      =  [  2.92277e+00 :  3.09139e+01 ]
```

Fig. D.6 Example of a Chebyshev model as produced by the evaluation of a factorable function using MC++.

class and it needs also to be associated to the ellipsoidal image environment using its `set` method. Finally, the dependent variables are also defined as functions of the ellipsoidal independent variables defined before. As explained in Chapter 2, the ellipsoidal arithmetic uses a sequence of liftings, so its shape matrix holds all the first and second order dependencies between the original variables and the auxiliary variables created. In order to obtain the enclosure for the image set of the function, the dependent variables have to be projected using the `get` method of the `mc::EllImg<T>` class. Finally, the projected ellipsoid can be obtained by using the standard output on the `Ellimage<T>` object. Figure D.8 presents an example of the output of the evaluation of Eq. (D.1) in ellipsoidal arithmetic. The standard output applied to a projected ellipsoidal image gives prints the center vector and the symmetric shape matrix.

```
////////////  Evaluating a Factorable Function  //////////
////////////      in Ellipsoidal Arithmetic     //////////

  // Define parameters for ellipsoidal domain
  CPPL::dcovector q0(NF); CPPL::dsymatrix Q0(NF);
  q0(0) = 3.;      Q0(0,0) = 2.;
  q0(1) = 4.;      Q0(1,0) = 1.; Q0(1,1) = 2.;

  //CPPL::dssmatrix depmap = FF.depmap(NF,F,NX,X);
  // Construct ellipsoidal domain
  mc::EllImg<I> Img;
  mc::Ellipsoid::options.PSDCHK = false;
  Img.set( Q0, q0 );//, depmap );
  Img.options.CHEBUSE  = false;
  // Set independent variables in the ellipsoidal model
  mc::EllVar<I> EX[NX];
  for( long i=0; i<NX; ++i ) EX[i].set( Img, i );
  // Define dependent variables to hold result
  mc::EllVar<I> EF[NF];
  FF.eval( NF, F, EF, NX, X, EX );
  mc::EllImg<I> ImgProj = Img.get( NF, EF );
  std::cout << "Ellipsoidal␣Enclosure:";
  std::cout << ImgProj << std::endl;
```

Fig. D.7 Code snippet for evaluating a factorable function (Eq. D.1) in ellipsoidal arithmetic in MC++ using operator overloading.

```
Ellipsoidal Enclosure:
center:
 1.52526e+01
 1.44325e+01
shape:
 1.52960e+02 {4.55765e+01}
 4.55765e+01  9.03034e+01
```

Fig. D.8 Example of an ellipsoidal enclosure as produced by the evaluation of a factorable function using MC++.

D.2 Bounding the Reachable Set of Parametric ODEs using CRONOS

The focus of this section is the construction of bounds for parametric ODEs using CRONOS. In particular, the implementation will be demonstrated with the following Lotka-Volterra system:

$$\begin{aligned} t \in [0,10]: \\ \dot{x}_0(t,p) &= p x_0(t,p)(1-x_1(t,p)) \\ \dot{x}_1(t,p) &= p x_1(t,p)(x_0(t,p)-1) \end{aligned} \tag{D.3}$$

with $p \in [2.95, 3.05]$.

In the following, we will focus on continuous-time methods, (cf. Chapter 4) since the application of the discrete-time integrator presented in Chapter 3 is analogous. Figure D.9 presents a minimal working example for the definition of the IVP in Eq. (D.3).

The problem definition is very similar to that of bounding a factorable function. The problem is defined as a directed acyclic graph via an `mc::FFGraph` object. Then parameters and state variables are defined as `mc::FFVar` objects and set in the DAG –just as we did with independent variables in the previous section. Finally, the right-hand side and initial value functions are defined as `mc::FFVar` objects depending on states and parameters –it can also be constant like the initial value functions in our example.

Once the problem has been defined, the reachable set at a certain time point can be computed using either continuous or discrete-time method. Figure D.10 presents a code-snippet that can be added after the problem definition for computing enclosures for reachable sets using continuous-time methods and Chebyshev models with ellipsoidal remainder. The first step is to define the method, in our example we are using GSL integrators through the class `mc::ODEBND_GSL` contained in the `odebnd_gsl.hpp` header file. This class is templated in the specific arithmetic we want to use, in our example it requires an interval class, a Chebyshev model and a Chebyshev variable class (also with intervals as template parameters). In order to bound the ODE, the DAG, states, parameters, right-hand side and initial value functions have to be set using the `set_dag`, `set_state`, `set_parameter`, `set_differential` and `set_initial` methods respectively. Different wrapping mitigation strategies –as defined in Chapter 4– can be set using the class option `WRAPMIT` depending on whether we want to use differential inequalities `DINEQ`, the ellipsoidal method `ELLIPS` or just avoiding the exploitation of facet constraints `NONE`.

```
#include <fstream>
#include "odebnd_gsl.hpp"
#include "interval.hpp"
#include "cmodel.hpp"
typedef mc::Interval I;
typedef mc::CModel<I> CM;
typedef mc::CVar<I> CV;

int main()
{
//////// Bounding the Reachable Set of pODEs //////////
////////          Problem Definition          //////////

  mc::FFGraph IVP;  // DAG describing the parametric ODE
  // Define parameter (NP) and state (NX) dimension
  const unsigned int NP = 1, NX = 2;
  mc::FFVar P[NP];  // Parameter array
  for( unsigned int i=0; i<NP; i++ ) P[i].set( &IVP );
  mc::FFVar X[NX];  // State array
  for( unsigned int i=0; i<NX; i++ ) X[i].set( &IVP );
  mc::FFVar RHS[NX];  // Right-hand side function
  RHS[0] = P[0] * X[0] * ( 1. - X[1] );
  RHS[1] = P[0] * X[1] * ( X[0] - 1. );
  mc::FFVar IC[NX] = { 1.2, 1.1 };  // Initial values

  return 0;
}
```

Fig. D.9 Minimal working code for defining the bounding problem for the reachable set of a parametric ODE system (Eq. D.3).

The next step is the definition of the numeric and set-valued environments needed for evaluating the factorable functions that define the reachability problem. Here we are interested in computing the reachable set at `NS = 20` equidistant points between the initial and final times. These time stages are defined as a real-valued (`double`) array. Then the set-valued environments are set, starting by the interval domain of the parameter and then the Chebyshev model environment. The Chebyshev model environment is defined with respect to the parameters and in this example we are using a third order model. Since we are interested in bounding the reachable sets at different time points, `NS+1` arrays of `NX` Chebyshev variables have to be created. It is important to notice that the arrays are allo-

cated dynamically using the `new` operator and thus they have to be deallocated using the `delete[]` operator at the end of the `main()` function.

Computing Chebyshev model bounds for the state variables `CVxk` at `NS` time points `tk` with respect to the parameter bounds `CVp` is achieved through the `bounds` method of the `mc::ODEBND_GSL` class. These bounds are outputted to the terminal and can also be recorded to an output stream through the `bndrec` method.

Computing enclosures using other set-valued arithmetics can be easily done by calling the `mc::ODEBND_GSL` with other template parameters such as Taylor models (`mc::TModel<mc::Interval>`) with Taylor variables (`mc::TVar<mc::Interval>`). Interval (differential inequalities) and the Ellipsoidal methods can be used without polynomial models through the `mc::ODEBND_GSL<T>` with an interval class as the single template parameter and using `DINEQ` and `ELLIPS` respectively as the wrapping mitigation strategy (`options.WRAPMIT`). In a similar manner, SUNDIALS integrators can be used in place of the GSL integrators just by changing `mc::ODEBND_GSL` to the `mc::ODEBND_SUNDIALS` class. The discrete integrator presented in Chapter 3 is implemented in CRONOS as the `mc::ODEBND_VAL` class and its usage is analogous to the example presented here.

```
///////// Bounding the Reachable Set of pODEs //////////
/////////   Using GSL and Chebyshev Models   ///////////

  // Define Bounding Problem
  mc::ODEBND_GSL<I,CM,CV> LV; // Use GSL
  LV.set_dag( &IVP );
  LV.set_state( NX, X );
  LV.set_parameter( NP, P );
  LV.set_differential( NX, RHS );
  LV.set_initial( NX, IC );
  LV.options.RESRECORD = true;
  // Define wrapping mitigation strategy:
  // NONE, DINEQ or ELLIPS
  LV.options.WRAPMIT  =
  mc::ODEBND_GSL<I,CM,CV>::Options::ELLIPS;
  // Define time horizon and time stages
  double t0 = 0., tf = 10.;  // Time horizon
  const unsigned int NS = 20;  // Time stages
  double tk[NS+1]; tk[0] = t0;
  for( unsigned k=0; k<NS; k++ )
        tk[k+1] = tk[k] + (tf-t0)/(double)NS;
  // Define parameter domain
  mc::Interval Ip[NP] = { I(2.95,3.05) };
  // Define Chebyshev model environment and variables
  // to bound RHS and IC w.r.t. parametersgnu
  CM CMenv( NP, 3 ); // Environment
  CV CVp[NP]; // Variables
  for( unsigned i=0; i<NP; i++ )
        CVp[i].set( &CMenv, i, Ip[i] );
  // Define Chebyshev model variables to hold state bounds
  // at each time stage
  CV* CVxk[NS+1];
  for( unsigned k=0; k<=NS; k++ )
        CVxk[k] = new CV[NX];
  // Evaluate ODE bounds in
  LV.bounds( NS, tk, CVp, CVxk );
  std::ofstream bndrec( "Bounds.out", std::ios_base::out );
  LV.record( bndrec );
```

Fig. D.10 Code snippet for computing bounds of a parametric ODE system (continuous-time method) using Chebyshev models.

D.3 Bounding Systems of Nonlinear Algebraic Equations using CRONOS

This section deals with the computation of enclosures for sets defined implicitly by systems of nonlinear equations using CRONOS. As an example, consider the steady-state manifold of a mutating autocatalator model [1], given by

$$\begin{aligned} 0 &= -\frac{x}{\theta} + (1+\alpha)\gamma_1(1-x)y^2 + \beta\gamma_1(1-x)z^2 \\ 0 &= \frac{y_f - y}{\theta} + (1-\alpha)\gamma_1(1-x)y^2 - \gamma_2 y \\ 0 &= \frac{z_f - z}{\theta} + \beta\gamma_1(1-x)z^2 + 2\alpha\gamma_1(1-x)y^2 - \frac{\gamma_2}{\beta}z\,, \end{aligned} \tag{D.4}$$

with $\theta \in [10^{-6}, 1]$ and $(x, y, z) \in [0, 1]^3$.

In CRONOS, the solution of nonlinear algebraic equations is addressed via the `mc::NLCP_GUROBI<T>` class, where the template parameter is any interval class. Figure D.11 presents a minimal working code to find the equilibrium manifold of Equation D.4.

The code requires the `nlcp_gurobi.hpp` header file. The problem is defined just as the ODE and factorable function bounding –through a directed acyclic graph. Once the DAG, variables and functions have been defined, the constraint projection problem is instantiated through the `mc::NLCP_GUROBI` class. The problem is set in a similar manner as the ODE problem from the previous section, the DAG, variables and constraints are set through the `set_dag`, `set_var` and `add_ctr`. If the system is underdetermined (as in our example) the `set_var(NP-NF,P)` method can be used to set the (`NP-NF`) degrees of freedom variables (in our case `P[0] = theta`), while `set_dep(NF,P+NP-NF,F)` is used to set the remaining dependent variables. Finally, the problem can be solved through the `solve` method on the initial interval box.

```
#include <fstream>
#include "nlcp_gurobi.hpp"
#include "interval.hpp"
typedef mc::Interval I;
int main()
{
/////////     Bounding the Solution Set of     ///////////
/////////       Nonlinear parametric AEs        //////////
  mc::FFGraph AE;  // DAG describing the AE System
  // Define variable (NP) and equation (NX) dimension
  const unsigned int NP = 4, NF = 3;
  mc::FFVar T, mc::FFVar P[NP];  // Variables array
  for( unsigned int i=0; i<NP; i++ ) P[i].set( &AE );
  // Define model Parameters
  double  gamma1 = 450., gamma2 = 11.25, Yf = 0.067;
  double   alpha = 0.22,   beta = 0.66,  Zf = 0.0;
  // Define equilibrium Parameters
  mc::FFVar theta = P[0]; X = P[1], Y = P[2], Z = P[3];
  //Define algebraic equation system (dependent vars)
  mc::FFVar F[NF];
  F[0] =  - (X/theta)+(1.+alpha)*gamma1*(1.-X)*mc::pow(Y,2)
          + beta*gamma1*(1.-X)*mc::pow(Z,2);
  F[1] =    (Yf-Y)/theta
          + (1.-alpha)*gamma1*(1.-X)*mc::pow(Y,2)-gamma2*Y;
  F[2] =    (Zf-Z)/theta + beta*gamma1*(1.-X)*mc::pow(Z,2)
          + 2.*alpha*gamma1*(1.-X)*mc::pow(Y,2)
          - (gamma2/beta)*Z;
  // Define constraint projection problem
  mc::NLCP_GUROBI<I> CP;
  CP.set_dag( &AE ); CP.set_var( NP, P );
  for( unsigned i=0; i<NF; i++ )
    CP.add_ctr( mc::BASE_OPT::EQ, F[i] );
  // Define initial search domain
  const I Ip[NP] = {I(1.e-6,1.0), I(0.,1.),
                    I(0., 1.), I(0.,1.)};
  // Solve problem
  CP.setup(); CP.solve( Ip );
  return 0;
}
```

Fig. D.11 Minimal working code for defining the bounding problem for the reachable set of a parametric ODE system (Eq. D.3).