Advances in Delays and Dynamics 9

Alban Quadrat
Eva Zerz *Editors*

Algebraic and Symbolic Computation Methods in Dynamical Systems

Springer

Advances in Delays and Dynamics

Volume 9

Delay systems are largely encountered in modeling propagation and transportation phenomena, population dynamics and representing interactions between interconnected dynamics through material, energy and communication flows. Thought as an open library on delays and dynamics, this series is devoted to publish basic and advanced textbooks, explorative research monographs as well as proceedings volumes focusing on delays from modeling to analysis, optimization, control with a particular emphasis on applications spanning biology, ecology, economy and engineering. Topics covering interactions between delays and modeling (from engineering to biology and economic sciences), control strategies (including also control structure and robustness issues), optimization and computation (including also numerical approaches and related algorithms) by creating links and bridges between fields and areas in a delay setting are particularly encouraged.

More information about this series at http://www.springer.com/series/11914

Alban Quadrat · Eva Zerz
Editors

Algebraic and Symbolic Computation Methods in Dynamical Systems

Springer

Editors
Alban Quadrat
Inria Paris, Institut de Mathématiques de Jussieu-Paris Rive Gauche
Sorbonne University
Paris, France

Eva Zerz
Lehrstuhl D fuer Mathematik
RWTH Aachen University
Aachen, Germany

ISSN 2197-117X ISSN 2197-1161 (electronic)
Advances in Delays and Dynamics
ISBN 978-3-030-38355-8 ISBN 978-3-030-38356-5 (eBook)
https://doi.org/10.1007/978-3-030-38356-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to the memory of Isabelle Périer

Preface

This book aims at reviewing recent progress in the direction of algebraic and symbolic computation methods for functional systems—systems of equations whose unknowns are functions (e.g., systems of ordinary or partial differential equations, of differential time-delay equations, of difference equations, of integro-differential equations)—and for their controls.

In the nineties, modern algebraic theories (differential algebra, formal theory of systems of partial differential equations, *D*-modules, algebraic analysis, etc.) were introduced in mathematical systems theory and in control theory. Combined with real algebraic geometry, which was previously introduced in control theory, the past years have seen a flourishing development of algebraic methods in control theory. One of the strengths of algebraic methods lies in their close connections to computations. The use of the above-mentioned algebraic theories in control theory has been an important source of motivation to develop effective versions of these theories (when possible). With the development of computer algebra and computer algebra systems, symbolic methods for control theory have been developed over the past years.

The goal of this book is to propose a partial state-of-the-art in this direction based on articles written for the invited sessions *Algebraic and Symbolic Methods in Mathematical Systems Theory* of the 5th Symposium on System Structure and Control, IFAC, Grenoble, France, 2013, and *Algebraic Methods and Symbolic-Numeric Computation in Systems Theory* of the 21st International Symposium on Mathematical Theory of Networks and Systems (MTNS 2014), Groningen, the Netherlands, 2014, organized by the editors of the book. To make recent results more easily accessible to a large audience, these articles have been largely extended and the chapters include materials which survey the main mathematical methods and results, we hope in an accessible manner, and illustrated with explicit examples. The combination of pure mathematics, mathematical systems theory, control theory, computer algebra and implementation is demanding but, we believe, rewarding.

The book is divided into the following chapters:

Part I. Effective Algebraic Methods for Linear Functional Systems

The first part of the book focusses on the algebraic analysis approach to linear functional systems and their controls. The first chapter by *Thomas Cluzeau*, *Christoph Koutschan*, *Alban Quadrat* and *Maris Tõnso* gives a state-of-the-art of this theory and explains the connections with Willems' behavioral approach. Both the mathematical and computer algebra aspects are given and a recent `Mathematica` implementation of these results is illustrated with examples. Within the algebraic analysis approach, the second chapter, written by *Thomas Cluzeau* and *Alban Quadrat*, studies the equivalence problem, namely, the problem of recognizing when two linear functional systems are equivalent in the sense that a one-to-one transformation between the system solutions exists. Conditions are obtained based on isomorphic modules. They generalize different standard results of linear systems theory. The last chapter by *Alban Quadrat* and *Georg Regensburger* initiates the effective study of the noncommutative ring of ordinary integro-differential operators with polynomial coefficients. Based on the computation of the polynomial solutions of integro-differential operators, the explicit computation of compatibility conditions of such operators is shown. This is the first step towards an algebraic analysis approach to linear systems of integro-differential equations and to their applications in control theory.

Part II. Symbolic Methods for Nonlinear Dynamical Systems and for Applications to Observation and Estimation Problems

The second part is first dedicated to effective methods for nonlinear systems of differential equations based on Thomas decomposition technique or differential algebra methods (differential elimination), and their applications to nonlinear control theory and particularly to observability. The second part of the chapter is dedicated to the parameter estimation problem for nonlinear ordinary differential systems and linear partial differential systems based on integro-differential algebras, elimination methods and computation of annihilators of polynomials. In the first chapter, written by *Markus Lange-Hegermann* and *Daniel Robertz*, Thomas decomposition method for algebraic and differential systems is introduced and illustrated with explicit examples and an implementation in `Maple`. The authors then explain how to use Thomas decomposition method to solve (differential) elimination problems and finally to study classical control problems for nonlinear differential systems. In the second chapter by *Sette Diop*, the differential algebraic approach to polynomially nonlinear systems is first reviewed and then differential elimination techniques are used to study observation and the sensor selection problems. The third chapter, written by *François Boulier*, *François Lemaire*, *Markus Rosenkranz*, *Rosane Ushirobira* and *Nathalie Verdière*, studies the parameter estimation problem for nonlinear control systems based on integral

input-output representations, integro-differential equations and the first steps towards an extension of Ritt-Kolchin differential algebra to integro-differential algebra. Finally, in the last chapter by *Rosane Ushirobira*, *Anja Korporal* and *Wilfrid Perruquetti*, the parameter estimation and the numerical differentiation problems are first reviewed for linear ordinary differential equations. Then, they are extended to the case of linear partial differential equations based on the computation of annihilators of multivariate Laurent polynomials over the Weyl algebra of partial differential operators with polynomial coefficients.

Part III. Algebraic Geometry Methods for Systems and Control Theory

The third part of the book is concerned with applications of (real) algebraic geometry methods to multidimensional systems, differential time-delay systems and nonlinear systems. The first chapter by *Yacine Bouzidi* and *Fabrice Rouillier* first gives an overview on recent computational aspects of real algebraic geometry and then applies them to the study of the structural stability of multidimensional systems. The second chapter, written by *Islam Boussaada* and *Silviu-Iulian Niculescu*, reviews recent results on the characterization of the multiplicity of imaginary roots of quasipolynomials associated with linear differential time-delay systems and on Birkhoff matrices related to the latter problem. A bound for the multiplicity of a crossing imaginary root is obtained and compared with the standard Pólya-Szegő generic bound. The third chapter by *Christian Schilli*, *Eva Zerz* and *Viktor Levandovskyy* fully characterizes when an algebraic variety is controlled and conditioned invariant with respect to a polynomially nonlinear state-space system and a polynomial feedback. The condition can be effectively checked by means of Gröbner basis techniques. The extension of this result to single output systems with rational feedback is also solved based on the concept of fractional modules. The different results are implemented in `Singular`. Finally, in the last chapter by *Ricardo Pereira* and *Paula Rocha*, controller invariance is first introduced in the context of *n*D behaviors and then characterized. The case where the controller is regular is completely characterized and the controllers achieving invariance are explicitly obtained. Preliminary results for 1D systems are finally obtained in the non-regular case.

Paris, France — Alban Quadrat
Aachen, Germany — Eva Zerz
September 2017

Contents

Part I
Effective Algebraic Methods for Linear Functional Systems

Chapter 1
Effective Algebraic Analysis Approach to Linear Systems over Ore Algebras

T. Cluzeau, C. Koutschan, A. Quadrat and M. Tõnso

Abstract The purpose of this chapter is to present a survey on the effective algebraic analysis approach to linear systems theory with applications to control theory and mathematical physics. In particular, we show how the combination of effective methods of computer algebra—based on Gröbner basis techniques over a class of noncommutative polynomial rings of functional operators called Ore algebras—and constructive aspects of module theory and homological algebra enables the characterization of structural properties of linear functional systems. Algorithms are given and a dedicated implementation, called OREALGEBRAICANALYSIS, based on the Mathematica package HOLONOMICFUNCTIONS, is demonstrated.

Keywords Linear systems theory · Control theory · Algebraic analysis · Computer algebra · Implementation

Supported by the PHC Parrot CASCAC (29586NG) and by the Austrian Science Fund (FWF): W1214.

T. Cluzeau
CNRS, XLIM UMR 7252, Université de Limoges, 123 avenue Albert Thomas, 87060 Limoges Cedex, France
e-mail: thomas.cluzeau@unilim.fr

C. Koutschan
RICAM, Austrian Academy of Sciences, Altenberger Straße 69, 4040 Linz, Austria
e-mail: christoph.koutschan@ricam.oeaw.ac.at

A. Quadrat (✉)
Inria Paris, Ouragan Project-Team, Institut de Mathématiques de Jussieu-Paris Rive Gauche, Sorbonne University, 4 Place Jussieu, 75252 Paris Cedex 05, France
e-mail: alban.quadrat@inria.fr

M. Tõnso
Institute of Cybernetics Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia
e-mail: maris@cc.ioc.ee

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_1

1.1 Introduction

To introduce the *algebraic analysis approach to linear systems over Ore algebras*, we use explicit examples. The model of a *stirred tank* studied in [32] on page 7 is defined by the following mass balance equations

$$\begin{cases} \dfrac{d\,V(t)}{dt} = -k\sqrt{\dfrac{V(t)}{S}} + F_1(t) + F_2(t), \\ \dfrac{d\,(c(t)\,V(t))}{dt} = -c(t)\,k\sqrt{\dfrac{V(t)}{S}} + c_1\,F_1(t) + c_2\,F_2(t), \end{cases}$$

where F_1 and F_2 denote the flow rates of two incoming flows feeding the tank, c_1 and c_2 two constant concentrations of dissolved materials, c the concentration in the tank, V the volume, k an experimental constant, and S the constant cross-sectional area. The algebraic analysis approach can only handle linear systems. See [7] for a first attempt to extend the algebraic analysis approach to particular classes of nonlinear systems. We refer to [35] for the use of differential elimination techniques for studying this nonlinear system. If V_0 is a constant volume, c_0 a constant concentration, and

$$\begin{aligned} F_{10} &:= \frac{(c_2 - c_0)}{(c_2 - c_1)}\,k\sqrt{\frac{V_0}{S}}, & F_{20} &:= \frac{(c_0 - c_1)}{(c_2 - c_1)}\,k\sqrt{\frac{V_0}{S}}, \\ V(t) &:= V_0 + x_1(t), & c(t) &:= c_0 + x_2(t), \\ F_1(t) &:= F_{10} + u_1(t), & F_2(t) &:= F_{20} + u_2(t), \end{aligned}$$

then the linearized model around the steady-state equilibrium is defined by

$$\begin{cases} \dot{x}_1(t) = -\dfrac{1}{2\,\theta}\,x_1(t) + u_1(t) + u_2(t), \\ \dot{x}_2(t) = -\dfrac{1}{\theta}\,x_2(t) + \left(\dfrac{c_1 - c_0}{V_0}\right) u_1(t) + \left(\dfrac{c_2 - c_0}{V_0}\right) u_2(t), \end{cases} \tag{1.1}$$

with the notation $\theta := V_0/F_0$ (the holdup time of the tank), where $F_0 := k\sqrt{V_0/S}$. See pages 8–9 of [32]. The linear OD system (1.1) can then be studied by means of the standard analysis and synthesis techniques developed for linear OD systems.

Now, if a transport delay of amplitude $\tau > 0$ occurs in the pipe, then we obtain the following linear differential time-delay (DTD) system:

$$\begin{cases} \dot{x}_1(t) = -\dfrac{1}{2\,\theta}\,x_1(t) + u_1(t) + u_2(t), \\ \dot{x}_2(t) = -\dfrac{1}{\theta}\,x_2(t) + \left(\dfrac{c_1 - c_0}{V_0}\right) u_1(t - \tau) + \left(\dfrac{c_2 - c_0}{V_0}\right) u_2(t - \tau). \end{cases} \tag{1.2}$$

For more details, see pages 449–451 of [32]. Then, (1.2) can be studied by means of methods dedicated to linear DTD systems.

Following [32], if the valve settings are commanded by a process control computer which can only be changed at discrete instants and remain constant in between, the following discrete-time model of (1.1) can then be derived

$$\left\{\begin{array}{l} x_1(n+1) = e^{-\frac{\Delta}{2\theta}}\, x_1(n) + 2\,\theta\,(1 - e^{-\frac{\Delta}{2\theta}})\,(u_1(n) + u_2(n)), \\ x_2(n+1) = e^{-\frac{\Delta}{\theta}}\, x_2(n) + \dfrac{\theta\,(1 - e^{-\frac{\Delta}{\theta}})}{V_0}\,((c_1 - c_0)\,u_1(n) + (c_2 - c_0)\,u_2(n)), \end{array}\right. \tag{1.3}$$

where Δ is the constant length of time intervals. For more details, see page 449 of [32]. Again, (1.3) can then be studied by means of standard techniques developed for linear discrete-time systems.

As shown above, a physical system can be modeled by means of different systems of *functional equations*, namely, systems whose unknowns are functions (e.g., OD systems, DTD systems, discrete-time systems). Moreover, the "same" system can be defined by means of different representations (e.g., state-space, input-output, polynomial, behaviors, geometric, systems over a ring, implicit, … representations). These representations are defined by different numbers of unknowns and equations. Linear systems are usually studied by means of dedicated mathematical methods which usually depend on the representations. The equivalences between different representations and different formulations of system-theoretic properties (e.g., controllability à la Kalman, controllability for polynomial systems, controllability à la Willems) are known for certain classes of linear functional systems.

We can wonder whether or not a unique mathematical approach to linear systems exists which satisfies the following two important requirements:

(a) The approach can handle the standard classes of linear functional systems studied in control theory by means of common mathematical concepts, methods, theorems, algorithms, and implementations.
(b) The approach does not depend on particular representations of the linear systems.

The goal of this paper is to show that the algebraic analysis approach satisfies these two points. *Algebraic analysis* (also called *D-module theory*) is a mathematical theory developed by B. Malgrange, J. Bernstein, M. Sato and his school in the sixties to study linear systems of partial differential (PD) equations by means of module theory, homological algebra, and sheaf theory (see [25, 28, 40] and the references therein). In the nineties, algebraic analysis techniques were introduced in mathematical systems theory and control theory by U. Oberst, M. Fliess, and J.-F. Pommaret. For more details, see [21, 23, 43, 45, 46, 57] and the references therein.

Within the algebraic version of algebraic analysis, a linear system is studied by means of a *finitely presented left module M* [52] over a ring D of functional operators, and its $\mathcal{F}$-solutions are defined by the homomorphisms (namely, the left D-linear maps) from M to $\mathcal{F}$, where $\mathcal{F}$ is a left D-module. We recall that a module is an alge-

braic structure which is defined by the same properties as the ones for a vector space but its scalars belong to a ring and not a field. Equivalent representations of a linear system yield isomorphic modules. These isomorphic modules are finitely presented by the different presentations, i.e., by the different matrices of functional operators defined by these representations. Hence, up to isomorphism, a linear system defines uniquely a finitely presented module. Structural (built-in) properties of linear systems, i.e., properties which do not depend on the representation of the system, then correspond to module properties (e.g., torsion elements, torsion-freeness, projectiveness, freeness). To study these module properties, we use *homological algebra* methods since they depend only on the underlying modules (up to isomorphism) and not on the presentations of the modules, i.e., not on the representations of the linear systems. Therefore, we have a way to study structural properties of linear systems independently of their representations. A second benefit of using homological algebra techniques is that large classes of linear functional systems can be studied by means of the same techniques, results, and algorithms since the standard rings of functional operators share the same properties. Only the "arithmetic" of the functional operators can be different. Based on *Gröbner or Janet basis techniques* [1, 33] for classes of noncommutative polynomial rings of functional operators, effective studies of module theory and homological algebra have recently been developed (see [9, 12, 47, 51] and the references therein). Dedicated symbolic packages such as OREMODULES, OREMORPHISMS and CLIPS have been developed [10, 13, 58].

The purpose of this paper is two-fold. We first give a brief overview of the algebraic analysis approach to linear systems defined over Ore algebras. We then show how a recent implementation of Gröbner bases for large classes of Ore algebras in a `Mathematica` package called HOLONOMICFUNCTIONS [29, 30] can be used to extend the classes of linear functional systems we can effectively study within the algebraic analysis approach. In particular, using the recent OREALGEBRAICANALYSIS package, we can now handle generic linearizations of explicit nonlinear functional systems or linear systems containing transcendental function (e.g., sin, cos, tanh) or special function coefficients (e.g., Airy or Bessel functions). These classes could not be studied by the OREMODULES, OREMORPHISMS or CLIP packages.

The paper is organized as follows. In Sect. 1.2, we explain that standard linear functional systems encountered in control theory can be studied by means of a polynomial approach over Ore algebras of functional operators, i.e., over a certain class of noncommutative polynomial rings. In Sect. 1.3, we shortly explain the concept of a Gröbner basis for left ideals and left modules over certain Ore algebras, and give algorithms to compute kernel and left/right inverses of matrices with entries in these Ore algebras. In Sect. 1.4, we introduce the algebraic analysis approach to linear systems theory and, using homological algebra techniques, we explain that this approach is an intrinsic polynomial approach to linear systems theory and we characterize standard system-theoretic properties in terms of module properties and homological algebra concepts that are shown to be computable. Finally, in Sect. 1.5, these results are illustrated on explicit examples which are studied by means of the OREALGEBRAICANALYSIS package. This package is based on the `Mathematica` package HOLONOMICFUNCTIONS which contains Gröbner basis techniques for general classes of Ore algebras.

1.2 Linear Systems over Ore Algebras

In this section, we introduce the concept of a *skew polynomial ring*, an *Ore extension* and an *Ore algebra* [16] which will play important roles in what follows. To motivate the abstract definitions, let us start with standard examples of functional operators. In his treatises on differential equations, G. Boole used the idea of representing a linear OD equation $\sum_{i=0}^{r} a_i\, y^{(i)}(t) = 0$, where $a_i \in \mathbb{R}$, by means of the operator $P := \sum_{i=0}^{r} a_i\, \frac{d^i}{dt^i}$, where $\frac{d}{dt}\, y(t) := y^{(1)}(t) = \dot{y}(t)$ is the first derivative of the function y. Note that $\frac{d^i}{dt^i}$ is the ith composition of the operator $\frac{d}{dt}$. If the composition of operators is simply denoted by the standard product, we then have $\frac{d^i}{dt^i} = \left(\frac{d}{dt}\right)^i$. Hence, P can be rewritten as the polynomial $P = \sum_{i=0}^{r} a_i\, \partial^i$ in $\partial := \frac{d}{dt}$ with coefficients in $\mathbb{R}$. It is important to note that the element $a_i \in \mathbb{R}$ in the expression of P is seen as the multiplication operator $y \longmapsto a_i\, y$, and $a_i\, \partial^i$ stands for the composition of the two operators a_i and ∂^i. As understood by G. Boole, the set of OD operators forms the commutative polynomial ring $\mathbb{R}[\partial]$. Algebraic techniques (e.g., Euclidean division) can then be used to study linear OD equations with constant coefficients.

More generally, if $\mathbb{A}$ is a *differential ring*, namely a ring equipped with a derivation $\frac{d}{dt} : \mathbb{A} \longrightarrow \mathbb{A}$ satisfying the additivity condition and Leibniz's rule, namely,

$$\forall\, a_1,\, a_2 \in \mathbb{A}, \quad \frac{d}{dt}(a_1 + a_2) = \frac{da_1}{dt} + \frac{da_2}{dt}, \quad \frac{d}{dt}(a_1\, a_2) = \frac{da_1}{dt}\, a_2 + a_1\, \frac{da_2}{dt},$$

such as, for instance, the ring (resp., field) $k[t]$ (resp., $k(t)$) of polynomials (resp., rational functions) in t with coefficients in a field k (e.g., $k = \mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$) or $C^\infty(\mathbb{R})$, then we can define the set of all the OD operators of the form $\sum_{i=0}^{r} a_i\, \partial^i$ with $a_i \in \mathbb{A}$. This set inherits a ring structure if the composition of OD operators is still an OD operator, i.e., if we have

$$\left(\sum_{j=0}^{m} b_j\, \partial^j\right) \left(\sum_{i=0}^{n} a_i\, \partial^i\right) = \sum_{k=0}^{l} c_k\, \partial^k, \tag{1.4}$$

for a certain l and for some $c_k \in \mathbb{A}$. In particular, such an identity should hold for $m = 1$ and $n = 0$, i.e., the composition of the two operators $b_1\, \partial$ and a_0 has to be an OD operator. Since operators are understood by their actions on functions, we get

$$\begin{aligned}\forall\, y \in \mathbb{A}, \quad (b_1\, \partial\, a_0)\, y &= b_1\, \partial\, (a_0\, y) = b_1\, \frac{d}{dt}\, (a_0\, y) = b_1\, \left(a_0\, \frac{dy}{dt} + \frac{da_0}{dt}\, y\right) \\ &= \left(b_1\, \left(a_0\, \partial + \frac{da_0}{dt}\right)\right)\, y.\end{aligned}$$

Hence, on the OD operator level, we have the following commutation rule:

$$\forall\, a \in \mathbb{A}, \quad \partial\, a = a\, \partial + \dot{a}. \tag{1.5}$$

It can be shown below that this commutation rule is enough to define a ring structure on the set of all the OD operators with coefficients in $\mathbb{A}$. Note that the above commutation rule shows that this ring is usually noncommutative apart from the case where $\dot{a} = 0$ for all $a \in \mathbb{A}$, i.e., the case where $\mathbb{A}$ is a ring of constants.

If we consider the case of a time-delay operator S defined by $S\, y(t) = y(t-h)$, where $h > 0$, then to understand $S\, a$ as an operator, where a is an element of a *difference ring* $\mathbb{A}$ of functions, namely, a commutative ring $\mathbb{A}$ of functions of t equipped with the endomorphism $a(t) \in \mathbb{A} \longmapsto a(t-h) \in \mathbb{A}$, we have to apply it to a function y. We get

$$(S\, a(t))\, y(t) = S\, (a(t)\, y(t)) = a(t-h)\, y(t-h) = (a(t-h)\, S)\, y(t),$$

i.e., on the operator level, we have the following commutation rule:

$$S\, a(t) = a(t-h)\, S. \tag{1.6}$$

We note that in (1.5) and (1.6) the "degree" in ∂ or in S is 1 in both sides of the equalities. More generally, we can consider an operator ∂ which satisfies

$$\forall\, a \in \mathbb{A}, \quad \partial\, a = \sigma\, \partial + \delta,$$

where $0 \neq \sigma,\ \delta \in \mathbb{A}$, so that both sides of the above expression have degree 1 in ∂. Clearly, σ and δ depend on a, i.e., $\sigma(a)$ and $\delta(a)$. If we want to define a ring formed by elements which can uniquely be represented as $\sum_{i=0}^{r} a_i\, \partial^i$, we must have

$$\begin{aligned} \forall\, a_1,\, a_2 \in \mathbb{A}, \quad \partial\, (a_1 + a_2) &= \sigma(a_1 + a_2)\, \partial + \delta(a_1 + a_2) \\ &= \partial\, a_1 + \partial\, a_2 = \sigma(a_1)\, \partial + \delta(a_1) + \sigma(a_2)\, \partial + \delta(a_2) \\ &= (\sigma(a_1) + \sigma(a_2))\, \partial + \delta(a_1) + \delta(a_2), \end{aligned}$$

which yields the following identities:

$$\sigma(a_1 + a_2) = \sigma(a_1) + \sigma(a_2), \quad \delta(a_1 + a_2) = \delta(a_1) + \delta(a_2).$$

Similarly, using the associativity of operators, we obtain

$$\begin{aligned} \forall\, a_1,\, a_2 \in \mathbb{A}, \quad \partial\, (a_1\, a_2) &= \sigma(a_1\, a_2)\, \partial + \delta(a_1\, a_2) \\ &= (\partial\, a_1)\, a_2 = (\sigma(a_1)\, \partial + \delta(a_1))\, a_2 \\ &= \sigma(a_1)\, (\sigma(a_2)\, \partial + \delta(a_2)) + \delta(a_1)\, a_2 \\ &= \sigma(a_1)\, \sigma(a_2)\, \partial + \sigma(a_1)\, \delta(a_2) + \delta(a_1)\, a_2, \end{aligned}$$

which yields the following identities:

$$\sigma(a_1\,a_2) = \sigma(a_1)\,\sigma(a_2), \quad \delta(a_1\,a_2) = \sigma(a_1)\,\delta(a_2) + \delta(a_1)\,a_2.$$

We also have that $\partial = \partial\,1 = \sigma(1)\,\partial + \delta(1)$, which yields:

$$\sigma(1) = 1, \quad \delta(1) = 0.$$

The conditions on σ show that σ is an endomorphism of the ring $\mathbb{A}$ and δ is called a *σ-derivation* (if $\sigma = \mathrm{id}_{\mathbb{A}}$, we find again the above definition of a derivation).

The concept of an *Ore extension* of a ring $\mathbb{A}$ was introduced by Ore [44] in 1933 to develop a unified mathematical framework to represent linear functional operators such as differential operators, difference and shift operators, q-shift and q-differential operators, and many more. Nowadays, this concept is widely used to state results and algorithms about linear functional operators in a concise and general form. For applications of this framework, for instance, to the problem of factoring operators or *creative telescoping*, see [4, 8] and the references therein.

Definition 1 ([16]) Let $\mathbb{A}$ be a ring. An *Ore extension* $\mathbb{O} := \mathbb{A}[\partial; \sigma, \delta]$ of $\mathbb{A}$ is the noncommutative ring formed by all polynomials of the form $\sum_{i=0}^{n} a_i\,\partial^i$, where $n \in \mathbb{N}$ and $a_i \in \mathbb{A}$, obeying the following commutation rule

$$\forall\, a \in \mathbb{A}, \quad \partial\,a = \sigma(a)\,\partial + \delta(a), \tag{1.7}$$

where σ is an endomorphism of $\mathbb{A}$, namely, $\sigma : \mathbb{A} \longrightarrow \mathbb{A}$ satisfies

$$\forall\, a,\, b \in \mathbb{A}, \quad \begin{cases} \sigma(1) = 1, \\ \sigma(a+b) = \sigma(a) + \sigma(b), \\ \sigma(a\,b) = \sigma(a)\,\sigma(b), \end{cases}$$

and δ is a σ-derivation of $\mathbb{A}$, namely, $\delta : \mathbb{A} \longrightarrow \mathbb{A}$ satisfies:

$$\forall\, a,\, b \in \mathbb{A}, \quad \begin{cases} \delta(a+b) = \delta(a) + \delta(b), \\ \delta(a\,b) = \sigma(a)\,\delta(b) + \delta(a)\,b. \end{cases} \tag{1.8}$$

The Ore extension $\mathbb{A}[\partial; \sigma, \delta]$ is also called a *skew polynomial ring*.

Let $\mathbb{O} := \mathbb{A}[\partial; \sigma, \delta]$ be a skew polynomial ring, $P := \sum_{i=0}^{n} a_i\,\partial^i \in \mathbb{O}$, where $a_n \neq 0$, and $Q := \sum_{i=0}^{m} b_i\,\partial^i \in \mathbb{O}$, where $b_m \neq 0$. If $\mathbb{A}$ is a *domain*, i.e., $\mathbb{A}$ does not contain non-trivial zero divisors, then we have

$$P\,Q = (a_n\,\partial^n + \cdots)\,(b_m\,\partial^m + \cdots) = a_n\,\sigma^n(b_m)\,\partial^{n+m} + \cdots,$$

where $\cdots$ represents lower degree terms. Moreover, if σ is injective, we can define the *degree* of P to be n and the *degree* of Q to be m since we have:

$$\forall\, P,\, Q \in \mathbb{O}, \quad \deg_{\partial}(P\,Q) = \deg_{\partial}(P) + \deg_{\partial}(Q).$$

A skew polynomial ring $\mathbb{A}[\partial; \sigma, \delta]$ has the structure of an $\mathbb{A}-\mathbb{A}$-*bimodule*, namely, $\mathbb{O}$ has a left module structure defined by

$$\forall\, a \in \mathbb{A}, \quad \forall\, P = \sum_{i=0}^{r} a_i\, \partial^i \in \mathbb{O} : \quad a\, P = \sum_{i=0}^{r} (a\, a_i)\, \partial^i,$$

and a right $\mathbb{A}$-module structure defined by

$$\forall\, a \in \mathbb{A}, \quad \forall\, P = \sum_{i=0}^{r} a_i\, \partial^i \in \mathbb{O} : \quad P\, a = \sum_{i=0}^{r} a_i\, \partial^i\, a,$$

and they satisfy the following associativity condition:

$$\forall\, a_1,\, a_2 \in \mathbb{A}, \quad \forall\, P \in \mathbb{O}, \quad (a_1\, P)\, a_2 = a_1\, (P\, a_2).$$

Example 1 Let us give a few examples of skew polynomial rings.

(a) If $(\mathbb{A}, \delta)$ is a differential ring, i.e., $\mathbb{A}$ is a ring and δ is a derivation of $\mathbb{A}$, i.e., δ satisfies (1.8) with $\sigma = \mathrm{id}_{\mathbb{A}}$, then we can define the skew polynomial ring $\mathbb{A}[\partial; \mathrm{id}_{\mathbb{A}}, \delta]$ of OD operators with coefficients in $\mathbb{A}$. Then, (1.7) yields (1.5). For instance, if we consider again (1.1), then we can define the algebra $\mathbb{O} := \mathbb{Q}(\theta, c_0, c_1, c_2, V_0)[\partial; \mathrm{id}_{\mathbb{A}}, \delta]$ of OD operators with coefficients in the field $\mathbb{A} := \mathbb{Q}(\theta, c_0, c_1, c_2, V_0)$ of rational functions in the system parameters θ, c_0, c_1, c_2, and V_0, where $\delta := \frac{d}{dt}$ is the trivial derivation of $\mathbb{A}$, i.e., $\delta(a) = 0$ for all $a \in \mathbb{A}$. Thus, (1.5) implies that $\partial\, a = a\, \partial$ for all $a \in \mathbb{A}$, which shows that $\mathbb{O}$ is a commutative polynomial ring. Then, (1.1) can be rewritten as $R\, \eta = 0$, where:

$$R := \begin{pmatrix} \partial + \dfrac{1}{2\,\theta} & 0 & -1 & -1 \\ 0 & \partial + \dfrac{1}{\theta} & -\dfrac{c_1 - c_0}{V_0} & -\dfrac{c_2 - c_0}{V_0} \end{pmatrix} \in \mathbb{O}^{2\times 4}, \quad \eta := \begin{pmatrix} x_1(t) \\ x_2(t) \\ u_1(t) \\ u_2(t) \end{pmatrix}.$$

If one of the parameters is now a smooth function of t, then δ is no more the trivial derivation of $\mathbb{A} := C^\infty(\mathbb{R})$, and thus $\mathbb{O}$ is then a noncommutative polynomial ring in ∂ with coefficients in $\mathbb{A}$.
A simple example of a noncommutative polynomial ring of OD operators is given by $\mathbb{O} := \mathbb{R}[x][\partial; \mathrm{id}, \delta]$, where $\delta := \frac{d}{dx}$ is the standard derivation on $\mathbb{R}[x]$. The error function $\mathrm{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ satisfies the following ODE:

$$\left(\partial^2 + 2\, x\, \partial\right) \mathrm{erf}(x) = 0.$$

(b) If we consider the algebra $\mathbb{A} := \mathbb{Q}(\theta, c_0, c_1, c_2, V_0, \Delta, n)$ and the endomorphism $\sigma(a(n)) := a(n+1)$ of $\mathbb{A}$, then we can define the skew polynomial ring $\mathbb{O} := \mathbb{A}[S; \sigma, 0]$ of forward shift operators, which encodes the commutation rule $S\,a(n) = a(n+1)\,S$ for $a \in \mathbb{A}$. Then, (1.3) can be written as $R\,\eta = 0$, where:

$$R := \begin{pmatrix} S - e^{-\frac{\Delta}{2\theta}} & 0 & -2\,\theta\,(1 - e^{-\frac{\Delta}{2\theta}}) & -2\,\theta\,(1 - e^{-\frac{\Delta}{2\theta}}) \\ 0 & S - e^{-\frac{\Delta}{\theta}} & -\alpha\,(c_1 - c_0) & -\alpha\,(c_2 - c_0) \end{pmatrix} \in \mathbb{O}^{2\times 4},$$

$$\alpha := \frac{\theta\,(1 - e^{-\frac{\Delta}{\theta}})}{V_0}, \quad \eta := (x_1(n) \quad x_2(n) \quad u_1(n) \quad u_2(n))^T.$$

Since no entry of R is a (rational) function of n, we can only consider the algebra $\mathbb{A} := \mathbb{Q}(\theta, c_0, c_1, c_2, V_0, \Delta)$ and $\sigma = \mathrm{id}_A$. We then get $S\,a = a\,S$ for all $a \in \mathbb{A}$, i.e., the ring of shift operators with constant coefficients is commutative.
A simple example of a noncommutative polynomial ring of shift operators is $\mathbb{Q}[n][S; \sigma, 0]$, where $\sigma(a(n)) = a(n+1)$ for all $a \in \mathbb{Q}[n]$. The Gamma function $\Gamma(z) := \int_0^{+\infty} t^{z-1}\,e^{-t}\,dt$ for $\Re(z) > 0$ satisfies the following recurrence relation:

$$(S - n)\,\Gamma(n) = 0.$$

(c) Similarly as the previous case, if $h \in \mathbb{R}_{\geq 0}$ and $\mathbb{A}$ is a difference ring of functions of t with $\sigma(a(t)) = a(t-h)$ for all $a \in \mathbb{A}$ as an endomorphism, then we can define the ring $\mathbb{O} := \mathbb{A}[S; \sigma, 0]$ of TD operators in S with coefficients in $\mathbb{A}$. We then have $S\,a(t) = a(t-h)\,S$, which is exactly (1.6).

(d) If we want to reformulate (1.2) within the language of Ore extensions, we have to define the ring of DTD operators. To do that, we can first consider a difference-differential ring $(\mathbb{A}, \sigma, \delta)$ and the skew polynomial ring $\mathbb{B} := \mathbb{A}[\partial; \mathrm{id}_{\mathbb{A}}, \delta]$ defined in (a) and then define the Ore extension $\mathbb{O} := \mathbb{B}[S; \sigma, 0]$ of $\mathbb{B}$, where σ is the endomorphism of $\mathbb{B}$ defined by $\sigma(a(t)) = a(t-h)$ for all $a \in \mathbb{A}$ and $\sigma(\partial) = \partial$ so that $\sigma(\sum_{i=0}^r a_i(t)\,\partial^i) = \sum_{i=0}^r a_i(t-h)\,\partial^i$. In particular, we have $S\,\partial = \sigma(\partial)\,S = \partial\,S$, i.e., the two operators ∂ and S commute. This last identity encodes the following identity:

$$(\partial\,S)(y(t)) = \partial(y(t-h)) = \dot{y}(t-h) = (S\,\partial)(y(t)). \tag{1.9}$$

Then, (1.2) can be rewritten as $R\,\eta = 0$, where:

$$R := \begin{pmatrix} \partial + \dfrac{1}{2\,\theta} & 0 & -1 & -1 \\ 0 & \partial + \dfrac{1}{\theta} & -\dfrac{(c_1 - c_0)}{V_0}\,S & -\dfrac{(c_2 - c_0)}{V_0}\,S \end{pmatrix} \in \mathbb{O}^{2\times 4},$$

$$\eta := (x_1(t) \quad x_2(t) \quad u_1(t) \quad u_2(t))^T.$$

(e) If we consider the difference (resp., divided difference) operator

$$a(t) \longmapsto a(t+1) - a(t) \quad \left(\text{resp., } a(t) \longmapsto \frac{a(t) - a(t_0)}{t - t_0}\right),$$

for a fixed $t_0 \in \mathbb{R}$ and for all a belonging to a field $\mathbb{A}$ of real-valued functions of t, then we can form the skew polynomial ring $\mathbb{A}[\partial; \sigma, \delta]$ of difference (resp., divided difference) operators with coefficients in $\mathbb{A}$ by respectively considering:

$$\forall\, a \in \mathbb{A}, \quad \begin{cases} \sigma(a(t)) = a(t+1), \\ \delta(a(t)) = a(t+1) - a(t), \end{cases} \quad \begin{cases} \sigma(a(t)) = a(t_0), \\ \delta(a(t)) = \dfrac{a(t) - a(t_0)}{t - t_0}. \end{cases}$$

If $\mathbb{A}$ is a (skew) field, then the *right Euclidean division* can be performed, i.e., the algebra $\mathbb{O}$ is a *right Euclidean domain*, and thus a *principal left ideal domain*, namely, every left ideal of $\mathbb{O}$ is finitely generated (see, e.g., [4, 16]). Finally, if σ is also invertible, i.e., is an *automorphism* of $\mathbb{A}$, then the *left Euclidean division* can also be performed, i.e., $\mathbb{O}$ is a *left Euclidean domain*, and thus a *principal right ideal domain*. More details on skew polynomial rings can be found in [16]. A left and right Euclidean domain is simply called a *Euclidean domain*.

Theorem 1 ([16]) *Let $D := \mathbb{A}[\partial; \sigma, \delta]$ be a left skew polynomial ring over a ring $\mathbb{A}$. Then, we have:*

(a) If $\mathbb{A}$ is a domain*, i.e., $\mathbb{A}$ does not have non-trivial zero divisors, and σ is an injective endomorphism of $\mathbb{A}$, then D is a domain.*

(b) If $\mathbb{A}$ is a left Ore domain*, i.e., a domain $\mathbb{A}$ which satisfies the* left Ore property *which states that for a_1, $a_2 \in \mathbb{A} \setminus \{0\}$, there exist b_1, $b_2 \in \mathbb{A} \setminus \{0\}$ such that $b_1\, a_1 = b_2\, a_2$, and α is injective, then D is a left Ore domain.*

(c) If $\mathbb{A}$ is a left *(resp., right)* noetherian ring*, i.e., every left (resp., right) ideal of $\mathbb{A}$ is finitely generated, and α is an automorphism of $\mathbb{A}$, then D is a left (right) noetherian ring. Moreover, if $\mathbb{A}$ is a domain, then D is a left Ore domain.*

As shown in (d) of Example 1, we can iterate the construction of an Ore extension to obtain a multivariate noncommutative polynomial ring:

$$\mathbb{A}[\partial_1; \sigma_1, \delta_1] \cdots [\partial_m; \sigma_m, \delta_m] := (\cdots((\mathbb{A}[\partial_1; \sigma_1, \delta_1])[\partial_2; \sigma_2, \delta_2]) \cdots)[\partial_m; \sigma_m, \delta_m].$$

If $\mathbb{B} := \mathbb{A}[\partial_1; \sigma_1, \delta_1] \cdots [\partial_{m-1}; \sigma_{m-1}, \delta_{m-1}]$, then $\mathbb{O} := \mathbb{B}[\partial_m; \sigma_m, \delta_m]$, where σ_m is an endomorphism of $\mathbb{B}$ and δ_m is a σ_m-derivation of $\mathbb{B}$. In particular, we get:

$$\forall\, i = 1, \ldots, m-1, \quad \forall\, a \in \mathbb{A}, \quad \begin{cases} \partial_m\, \partial_i = \sigma_m(\partial_i)\, \partial_m + \delta_m(\partial_i), \\ \partial_m\, a = \sigma_m(a)\, \partial_m + \delta_m(a). \end{cases}$$

Similarly, we have:

$$1 \leqslant i < j \leqslant m, \quad \forall\, a \in \mathbb{A}, \quad \begin{cases} \partial_j\, \partial_i = \sigma_j(\partial_i)\, \partial_j + \delta_j(\partial_i) \\ \partial_j\, a = \sigma_j(a)\, \partial_j + \delta_j(a). \end{cases} \tag{1.10}$$

If we want that ∂_j commutes with ∂_i, σ_j and δ_j must satisfy the conditions:

$$1 \leqslant i < j \leqslant m, \quad \sigma_j(\partial_i) = \partial_i, \quad \delta_j(\partial_i) = 0. \tag{1.11}$$

Moreover, let us assume that $\sigma_j(\mathbb{A}) \subseteq \mathbb{A}$ and $\delta_j(\mathbb{A}) \subseteq \mathbb{A}$. Then, we have:

$$\begin{aligned}
\partial_j\,(\partial_i\,a) &= \partial_j(\sigma_i(a)\,\partial_i + \delta_i(a)) \\
&= \sigma_j(\sigma_i(a)\,\partial_i)\,\partial_j + \delta_j(\sigma_i(a)\,\partial_i) + \sigma_j(\delta_i(a))\,\partial_j + \delta_j(\delta_i(a)) \\
&= \sigma_j(\sigma_i(a))\,\sigma_j(\partial_i)\,\partial_j + \sigma_j(\sigma_i(a))\,\delta_j(\partial_i) + \delta_j(\sigma_i(a))\,\partial_i \\
&\quad + \sigma_j(\delta_i(a))\,\partial_j + \delta_j(\delta_i(a))
\end{aligned}$$

Using (1.11), the above identity reduces to:

$$\partial_j\,(\partial_i\,a) = \sigma_j(\sigma_i(a))\,\partial_i\,\partial_j + \delta_j(\sigma_i(a))\,\partial_i + \sigma_j(\delta_i(a))\,\partial_j + \delta_j(\delta_i(a)).$$

Since $\sigma_j(a) \in \mathbb{A}$ and $\delta_j(a) \in \mathbb{A}$, we also have:

$$\begin{aligned}
\partial_i\,(\partial_j\,a) &= \partial_i(\sigma_j(a)\,\partial_j + \delta_j(a)) \\
&= \sigma_i(\sigma_j(a))\,\partial_i\,\partial_j + \delta_i(\sigma_j(a))\,\partial_j + \sigma_i(\delta_j(a))\,\partial_i + \delta_i(\delta_j(a)).
\end{aligned}$$

If we have $\sigma_j(\sigma_i(a)) = \sigma_i(\sigma_j(a))$, $\delta_j(\sigma_i(a)) = \sigma_i(\delta_j(a))$, $\sigma_j(\delta_i(a)) = \delta_i(\sigma_j(a))$, and $\delta_j(\delta_i(a)) = \delta_i(\delta_j(a))$ for all $a \in \mathbb{A}$, then we get $\partial_j\,\partial_i\,a = \partial_i\,\partial_j\,a$ for all $a \in \mathbb{A}$.

Definition 2 Let $\mathbb{k}$ be a field. If $\mathbb{A}$ is a $\mathbb{k}$-algebra, then an Ore extension of $\mathbb{A}$ of the form $\mathbb{A}[\partial_1; \sigma_1, \delta_1] \cdots [\partial_m; \sigma_m, \delta_m]$ is called an *Ore algebra* if $\sigma_j(\mathbb{A}) \subseteq \mathbb{A}$ and $\delta_j(\mathbb{A}) \subseteq \mathbb{A}$ for $j = 1, \ldots, m$, and:

$$1 \leqslant i < j \leqslant m, \quad \sigma_j(\partial_i) = \partial_i, \quad \delta_j(\partial_i) = 0,$$

$$1 \leqslant i, j \leqslant m, \quad i \neq j, \quad \begin{cases} (\sigma_j \circ \sigma_i)_{|\mathbb{A}} = (\sigma_i \circ \sigma_j)_{|\mathbb{A}}, \\ (\delta_j \circ \sigma_i)_{|\mathbb{A}} = (\sigma_i \circ \delta_j)_{|\mathbb{A}}, \\ (\delta_j \circ \delta_i)_{|\mathbb{A}} = (\delta_i \circ \delta_j)_{|\mathbb{A}}. \end{cases}$$

We then have $\partial_j\,\partial_i\,a = \partial_i\,\partial_j\,a$ for $1 \leqslant i < j \leqslant m$ and for all $a \in \mathbb{A}$.

Finally, an Ore algebra $\mathbb{A}[\partial_1; \sigma_1, \delta_1] \cdots [\partial_m; \sigma_m, \delta_m]$ with $\mathbb{A} := \mathbb{k}[x_1, \ldots, x_n]$ (resp., $\mathbb{A} := \mathbb{k}(x_1, \ldots, x_n)$) is called a *polynomial* (resp., *rational*) *Ore algebra*.

Remark 1 In Definition 2, the numbers m and n can be different. For instance, considering again (d) of Example 1, i.e., the Ore algebra $\mathbb{O} := \mathbb{A}[\partial; \mathrm{id}_A, \delta][S; \sigma, 0]$, where, for instance, $\mathbb{A} := \mathbb{k}[t]$, then we have $m = 2$ and $n = 1$.

If $\mathbb{O} := \mathbb{A}[\partial_1; \sigma_1, \delta_1] \cdots [\partial_m; \sigma_m, \delta_m]$ is an Ore extension of a ring $\mathbb{A}$, then $P \in \mathbb{O}$ can be expressed as $P = \sum_{0 \leqslant |\nu| \leqslant r} p_\nu\,\partial^\nu$, where $r \in \mathbb{N}$, $p_\nu \in \mathbb{A}$, $\nu := (\nu_1 \ \ldots \ \nu_m)^T \in \mathbb{N}^m$, $|\nu| := \nu_1 + \cdots + \nu_m$, and $\partial^\nu := \partial_1^{\nu_1} \cdots \partial_m^{\nu_m}$.

Example 2 (a) If $\mathbb{A} := \Bbbk[x_1, \ldots, x_n]$ (resp. $\mathbb{A} := \Bbbk(x_1, \ldots, x_n)$), then the Ore algebra $\mathbb{O} := \mathbb{A}[\partial_1; \sigma_1, \delta_1] \cdots [\partial_n; \sigma_n, \delta_n]$, where $\sigma_i := \mathrm{id}$ and $\delta_i := \frac{\partial}{\partial x_i}$ for $i = 1, \ldots, n$, is called the *polynomial* (resp., *rational*) *Weyl algebra* of PD operators with coefficients in $\mathbb{A}$. It is denoted by $A_n(\Bbbk)$ (resp., $B_n(\Bbbk)$).

(b) We can combine the two skew polynomial algebras defined in (a) and (b) of Example 1 to obtain the Ore algebra $\mathbb{O} := \mathbb{Q}(n, t)[\partial; \mathrm{id}, \delta][S; \sigma, 0]$ of differential-shift operators with coefficients in $\mathbb{Q}(n, t)$. The Bessel function of the first kind $J_n(t)$ satisfies the following functional equation:

$$\frac{d}{dt} J_n(t) = n\, t^{-1}\, J_n(t) - J_{n+1}(t).$$

This equation can be rewritten as $P\, J_n(t) = 0$, where $P := \partial + S - n\, t^{-1} \in \mathbb{O}$.

(c) If $\mathbb{A}$ is a $\Bbbk$-algebra equipped with the following endomorphisms

$$\forall\, a \in \mathbb{A}, \quad \sigma_k(a(i_1, \ldots, i_n)) := a(i_1, \ldots, i_k + 1, \ldots, i_n), \quad k = 1, \ldots n,$$

(e.g., $\mathbb{A} := \Bbbk[i_1, \ldots, i_n]$, $\Bbbk(i_1, \ldots, i_n)$, or the algebra of real-valued sequences in $(i_1, \ldots, i_n) \in \mathbb{Z}^n$), then $\mathbb{A}[S_1; \sigma_1, 0] \cdots [S_n; \sigma_n, 0]$ is the Ore algebra of multi-shift operators with coefficients in $\mathbb{A}$.

(d) The ring of differential time-varying delay operators with $S\, y(t) = y(t - h(t))$, where h is a smooth function satisfying $h(t) < t$ for all t larger than or equal to a certain $T > 0$, does not usually form an Ore algebra since we have

$$(\partial\, S)(y(t)) = \partial\, y(t - h(t)) = (1 - \dot{h}(t))\, \dot{y}(t - h(t)) = (1 - \dot{h}(t))\, (S\, \partial)(y(t)),$$

i.e., $\partial\, S = (1 - \dot{h})\, S\, \partial$. It is an Ore algebra if and only if h is a constant function and we find then again (1.9). In [50], it is shown that the ring of differential time-varying delay operators can be defined as an Ore extension and its properties are studied in terms of the function h.

For more examples of Ore algebras of functional operators and their uses in combinatorics and in the study of special functions, see [11] and the references therein.

Theorem 1 can be used to prove that the Ore algebras defined in Example 2 are both left and right noetherian domains. We say that they are *noetherian domains*.

Finally, let us introduce the concept of an *involution* of a ring which will be used in Sects. 1.3.2 and 1.4.2.

Definition 3 Let $\mathbb{O}$ be an Ore algebra over a base field $\Bbbk$. An *involution of* $\mathbb{O}$ is an anti-automorphism of order two of $\mathbb{O}$, i.e., a $\Bbbk$-linear map $\theta : \mathbb{O} \longrightarrow \mathbb{O}$ satisfying:

$$\forall P_1,\ P_2 \in \mathbb{O}, \quad \theta(P_1\, P_2) = \theta(P_2)\, \theta(P_1), \quad \theta \circ \theta = \mathrm{id}_{\mathbb{O}}.$$

Let us give a few examples of involutions.

Example 3 (a) If $\mathcal{O}$ is a commutative ring (e.g., $\mathcal{O} := \Bbbk[x_1, \ldots, x_n]$), then $\theta = \mathrm{id}_{\mathcal{O}}$ is an involution of $\mathcal{O}$.

(b) Let $\mathcal{O} := A_n(\Bbbk)$ the polynomial Weyl algebra over $\Bbbk$. Then, an involution of $\mathcal{O}$ is defined by $\theta(x_i) := x_i$ and $\theta(\partial_i) := -\partial_i$ for $i = 1, \ldots, n$. More generally, if $\mathcal{O} := \mathbb{A}[\partial_1; \mathrm{id}, \delta_1] \cdots [\partial_n; \mathrm{id}, \delta_n]$ is a ring of PD operators with coefficients in the differential ring $(\mathbb{A}, \{\delta_1, \ldots, \delta_n\})$, where $\delta_i := \frac{\partial}{\partial x_i}$ for $i = 1, \ldots, n$, then an involution θ of $\mathcal{O}$ is defined by:

$$\forall\, a \in \mathbb{A}, \quad \theta(a) := a, \quad \theta(\partial_i) := -\partial_i, \quad i = 1, \ldots, n.$$

(c) Let $\mathcal{O} := \Bbbk(n)[S; \sigma, 0]$ be the skew polynomial ring of forward shift operators considered in (b) of Example 1. Then, an involution of $\mathcal{O}$ can be defined by $\theta(n) := -n$ and $\theta(S) := -S$.

(d) Let $\mathcal{O} := \Bbbk[t][\partial; \mathrm{id}, \delta][S; \sigma, 0]$ be the Ore algebra of differential time-delay operators defined by $\delta := \frac{d}{dt}$, and $\sigma(a(t)) := a(t-1)$, where $a \in \Bbbk[t]$. Then, an involution of $\mathcal{O}$ can be defined by $\theta(t) := -t$, $\theta(\partial) := \partial$, and $\theta(S) := S$.

1.3 Gröbner Basis Techniques

In Sect. 1.2, we explain how standard linear functional systems can be defined by means of matrices of functional operators, i.e., by means of matrices with entries in noncommutative polynomial rings such as skew polynomial rings, Ore extensions, or Ore algebras. The idea of studying linear functional systems by means of the algebraic properties of their representations is well-developed in the polynomial approach [27]. If the ring of functional operators is a Euclidean domain, then *Smith normal forms* [27] can be extended to this noncommutative framework by considering the so-called *Jacobson normal forms*. For more details, implementations, and applications of Jacobson normal forms, see [38] and the references therein. If the ring of functional operators is not a Euclidean domain (e.g., if the ring is usually defined by more than one functional operators), then such normal forms do not exist. But the Euclidean algorithm of multivariate (noncommutative) polynomials can still be used if the set of monomials appearing in the polynomials can be ordered in a particular way. This idea yields the concept of a *Gröbner basis* for a set of polynomials (i.e., for an ideal) or for a matrix (i.e., for a module).

In the next sections, we will state algorithms for the study of built-in properties of linear functional systems. These algorithms will be based on elimination techniques such as Gröbner basis techniques over noncommutative Ore algebras. Before doing so, we first motivate their uses by an explicit example.

Example 4 In fluid mechanics, *Stokes equations*, which describe the flow of a viscous and incompressible fluid at low Reynolds number, are defined by

$$\begin{cases} -\nu\,\Delta\,u + c\,u + \nabla\,p = 0, \\ \nabla\,.\,u = 0, \end{cases}$$

where $u \in \mathbb{R}^n$ is the velocity, p the pressure, ν the viscosity, and c the reaction coefficient. For simplicity reasons, let us consider the special case $n = 2$, i.e.

$$\begin{cases} E_1 := -\nu\,(\partial_x^2\,u_1 + \partial_y^2\,u_1) + c\,u_1 + \partial_x\,p = 0, \\ E_2 := -\nu\,(\partial_x^2\,u_2 + \partial_y^2\,u_2) + c\,u_2 + \partial_y\,p = 0, \\ E_3 := \partial_x\,u_1 + \partial_y\,u_2 = 0, \end{cases} \tag{1.12}$$

with the standard notations $\partial_x := \frac{\partial}{\partial x}$ and $\partial_y := \frac{\partial}{\partial y}$.

We can wonder if the pressure p satisfies a system of PDEs by itself, i.e., if the components u_1 and u_2 of the speed can be eliminated from the equations of (1.12) to get PDEs only on p. Differentiating E_1 (resp., E_2) with respect to x (resp., y), we first obtain:

$$\begin{cases} \partial_x\,E_1 = -\nu\,\partial_x\,(\partial_x^2\,u_1 + \partial_y^2\,u_1) + c\,\partial_x\,u_1 + \partial_x^2\,p = 0, \\ \partial_y\,E_2 = -\nu\,\partial_y\,(\partial_x^2\,u_2 + \partial_y^2\,u_2) + c\,\partial_y\,u_2 + \partial_y^2\,p = 0. \end{cases}$$

Similarly, we have:

$$\begin{cases} \nu\,(\partial_x^2\,E_3 + \partial_y^2\,E_3) = \nu\,\partial_x\,(\partial_x^2\,u_1 + \partial_y^2\,u_1) + \nu\,\partial_y\,(\partial_x^2\,u_2 + \partial_y^2\,u_2) = 0, \\ -c\,E_3 = -c\,(\partial_x\,u_1 + \partial_y\,u_2) = 0. \end{cases}$$

Adding all the new differential consequences of the equations of (1.12), we get

$$\partial_x\,E_1 + \partial_y\,E_2 + \nu\,(\partial_x^2\,E_3 + \partial_y^2\,E_3) - c\,E_3 = \partial_x^2\,p + \partial_y^2\,p = 0,$$

i.e., (1.12) yields $\Delta\,p = 0$, where $\Delta := \partial_x^2 + \partial_y^2$ is the *Laplacian operator*. This is an important result in hydrodynamics: the pressure must satisfy $\Delta\,p = 0$.

Gröbner basis techniques can be used for automatically eliminating (if possible) fixed unknowns. To do that, we first have to recast the above computations within a polynomial framework. Let us first consider the commutative polynomial ring $D := \mathbb{Q}(\nu, c)\left[\partial_x, \partial_y\right]$ of PD operators in ∂_x and ∂_y with coefficients in $\mathbb{Q}(\nu, c)$. The operators ∂_x and ∂_y commute, i.e., $\partial_x\,\partial_y = \partial_y\,\partial_x$, because of Schwarz's theorem and (1.12) has only constant coefficients. An element $P \in D$ is of the form $P = \sum_{0 \leqslant \mu_x + \mu_y \leqslant r} a_\mu\,\partial_x^{\mu_x}\,\partial_y^{\mu_y} \in D$, where $r \in \mathbb{N}$, $a_\mu \in \mathbb{Q}(\nu, c)$, and $\mu := (\mu_x \quad \mu_y)^T \in \mathbb{N}^2$. Then, (1.12) can be rewritten as $R\,\eta = 0$, where:

$$R := \begin{pmatrix} -\nu\,\Delta + c & 0 & \partial_x \\ 0 & -\nu\,\Delta + c & \partial_y \\ \partial_x & \partial_y & 0 \end{pmatrix} \in D^{3\times 3}, \quad \eta := \begin{pmatrix} u_1 \\ u_2 \\ p \end{pmatrix}.$$

Then, the above computations correspond to the following matrix computations

$$(\partial_x \quad \partial_y \quad \nu\,\Delta - c)\begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} = \big((\partial_x \quad \partial_y \quad \nu\,\Delta - c)\,R\big)\begin{pmatrix} u_1 \\ u_2 \\ p \end{pmatrix} = \Delta\,p$$

and using the fact that $\Delta\,p = (0 \quad 0 \quad \Delta)\,\eta$, we obtain:

$$(0 \quad 0 \quad \Delta) = (\partial_x \quad \partial_y \quad \nu\,\Delta - c)\,R \in D^{1\times 3}\,R := \{\mu\,R \mid \mu \in D^{1\times 3}\}.$$

We note that the D-submodule $D^{1\times 3}\,R$ of $D^{1\times 3}$ is formed by all the D-linear combinations of the rows of R. These combinations correspond to all the linear differential consequences of the equations of (1.12). Within the operator framework, the fact that the pressure satisfies $\Delta\,p = 0$ can be rewritten as $(0 \quad 0 \quad \Delta) \in D^{1\times 3}\,R$.

If $R \in D^{q\times p}$, then the (left) D-submodule $L := D^{1\times q}\,R$ of $D^{1\times p}$ is generated by the rows of R. If D is a (noncommutative) polynomial ring, then a Gröbner basis of L is another set of generators of L, i.e., we have $L = D^{1\times q'}\,R'$ for a certain matrix $R' \in D^{q'\times p}$, for which the so-called *membership problem* can easily be checked. The membership problem aims at deciding whether or not $\lambda \in D^{1\times p}$ belongs to $D^{1\times q}\,R$. If D is a commutative polynomial ring with coefficients in a computable field, then Buchberger's algorithm [5] computes a Gröbner basis for a fixed monomial order. This result can be extended for some classes of noncommutative polynomial rings where the algorithm is proved to terminate. If a Gröbner basis R' of L is known, then we can reduce any $\lambda \in D^{1\times p}$ with respect to this Gröbner basis in a unique way, i.e., there exists a unique $\overline{\lambda} \in D^{1\times p}$, called the *normal form* of λ, such that $\lambda = \overline{\lambda} + \mu'\,R'$ for a certain $\mu' \in D^{1\times q'}$. Hence, we obtain that $\lambda \in L$ if and only if we have $\overline{\lambda} = 0$.

In the next sections, we first define the concept of a Gröbner basis for a finitely generated left ideal and then for a finitely generated left module.

1.3.1 Gröbner Bases for Ideals over Ore Algebras

We first explain the basics of Gröbner bases using the standard commutative setting, i.e., for the case of a polynomial ring in several commuting variables, and then shortly explain how the theory can be extended to noncommutative Ore algebras.

Let $\boldsymbol{x} := x_1, \ldots, x_n$ be a collection of variables, and let us denote by $\Bbbk[\boldsymbol{x}]$ the ring of multivariate polynomials in $x_1, \ldots, x_n$ with coefficients in the field $\Bbbk$. For $\alpha \in \mathbb{N}^n$, we define the monomial $\boldsymbol{x}^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. Unlike for univariate polynomials, there is no natural ordering of the monomials $\boldsymbol{x}^\alpha$ in a multivariate polynomial $\sum_{\alpha\in\mathbb{N}^n} c_\alpha\,\boldsymbol{x}^\alpha$. This is the reason for introducing the notion of *monomial order*, that is a *total order* $\prec$ on the set $\{\boldsymbol{x}^\alpha \mid \alpha \in \mathbb{N}^n\}$ of $\boldsymbol{x}$-monomials, namely an order $\prec$ which is total (i.e., we have either $\boldsymbol{x}^\alpha \prec \boldsymbol{x}^\beta$ or $\boldsymbol{x}^\beta \prec \boldsymbol{x}^\alpha$ for all $\alpha,\ \beta \in \mathbb{N}^n$, $\alpha \neq \beta$).

Definition 4 A monomial order on the set $\{\boldsymbol{x}^\alpha \mid \alpha \in \mathbb{N}^n\}$ of $\boldsymbol{x}$-monomials is called *admissible* if it satisfies the following conditions:

$$\begin{aligned} &(a)\quad 1 \prec \boldsymbol{x}^\alpha && \forall\ \alpha \in \mathbb{N}^n \setminus \{(0,\ldots,0)\},\\ &(b)\quad \boldsymbol{x}^\alpha \prec \boldsymbol{x}^\beta \implies \boldsymbol{x}^\alpha \boldsymbol{x}^\gamma \prec \boldsymbol{x}^\beta \boldsymbol{x}^\gamma && \forall\ \alpha,\ \beta,\ \gamma \in \mathbb{N}^n. \end{aligned}$$

It follows that the set of monomials is *well-founded* with respect to any admissible monomial order, i.e., that each strictly decreasing sequence of monomials is finite. This is a crucial property for proving the termination of Buchberger's algorithm which computes a Gröbner basis of a polynomial ideal.

Example 5 We identify a monomial $\boldsymbol{x}^\alpha$ with the multi-index $\alpha \in \mathbb{N}^n$.

(a) The *lexicographic order* on $\boldsymbol{x}$-monomials is defined by $\alpha \prec_{\text{lex}} \beta$ whenever the first nonzero entry of $\beta - \alpha$ is positive. For instance, if we consider $\mathbb{Q}[x_1, x_2, x_3]$, then we have:

$$\begin{aligned} 1 \prec_{\text{lex}} x_3 \prec_{\text{lex}} x_3^2 \prec_{\text{lex}} x_2 \prec_{\text{lex}} x_2 x_3 \prec_{\text{lex}} x_2^2 \prec_{\text{lex}} x_1 \prec_{\text{lex}} x_1 x_3 \\ \prec_{\text{lex}} x_1 x_2 \prec_{\text{lex}} x_1^2. \end{aligned}$$

(b) The *total degree order* (also called *degree reverse lexicographic order* or *graded reverse lexicographic order*) on $\boldsymbol{x}$-monomials is defined by $\alpha \prec_{\text{tdeg}} \beta$ whenever $|\alpha| < |\beta|$ or if we have $|\alpha| = |\beta|$, then the last nonzero entry of $\beta - \alpha$ is negative. It is also denoted $\prec_{\text{degrevlex}}$. For instance, if we consider $\mathbb{Q}[x_1, x_2, x_3]$, then we have:

$$\begin{aligned} 1 \prec_{\text{tdeg}} x_3 \prec_{\text{tdeg}} x_2 \prec_{\text{tdeg}} x_1 \prec_{\text{tdeg}} x_3^2 \prec_{\text{tdeg}} x_2 x_3 \prec_{\text{tdeg}} x_1 x_3 \\ \prec_{\text{tdeg}} x_2^2 \prec_{\text{tdeg}} x_1 x_2 \prec_{\text{tdeg}} x_1^2. \end{aligned}$$

(c) Let $\boldsymbol{x} := x_1, \ldots, x_n$ and $\boldsymbol{y} := y_1, \ldots, y_m$ be two collections of variables. Assume that an admissible monomial order $\prec_X$ (resp., $\prec_Y$) on $\boldsymbol{x}$-monomials (resp., on $\boldsymbol{y}$-monomials) is given. An *elimination order* is then defined by

$$u\,v \prec w\,t \iff u \prec_X w \ \text{ or } \ u = w \ \text{ and } \ v \prec_Y t,$$

where $u,\ w$ (resp., $v,\ t$) are $\boldsymbol{x}$-monomials (resp., $\boldsymbol{y}$-monomials). An elimination order serves to eliminate the x_i's. The elimination order, which will be used in what follows, is the one induced by the total degree orders on $\boldsymbol{x}$-monomials and $\boldsymbol{y}$-monomials. This is a very common order called *lexdeg*. For instance, if we consider $\mathbb{Q}[x_1, x_2, x_3]$, $\boldsymbol{x} = x_1,\ x_2$, $\boldsymbol{y} = x_3$, $\prec_X = \prec_{\text{tdeg}}$ and $\prec_Y = \prec_{\text{tdeg}}$, then we have:

$$\begin{aligned} 1 \prec_{\text{lexdeg}} x_3 \prec_{\text{lexdeg}} x_3^2 \prec_{\text{lexdeg}} x_2 \prec_{\text{lexdeg}} x_2 x_3 \prec_{\text{lexdeg}} x_1 \prec_{\text{lexdeg}} x_1 x_3 \\ \prec_{\text{lexdeg}} x_2^2 \prec_{\text{lexdeg}} x_1 x_2 \prec_{\text{lexdeg}} x_1^2. \end{aligned}$$

Definition 5 Let $P \in \mathbb{k}[\boldsymbol{x}] \setminus \{0\}$ and $\prec$ be an admissible monomial order. We can then define:

- The *leading monomial* $\mathrm{lm}_{\prec}(P)$ of P to be the $\prec$-maximal monomial that appears in P with nonzero coefficient.
- The *leading coefficient* $\mathrm{lc}_{\prec}(P)$ of P to be the coefficient of $\mathrm{lm}_{\prec}(P)$.
- The *leading term* $\mathrm{lt}_{\prec}(P)$ of P to be the product $\mathrm{lc}_{\prec}(P)\,\mathrm{lm}_{\prec}(P)$.

When no confusion can arise, we skip the explicit mentioning of the monomial order in the subscripts. Hence, we can write $P = \mathrm{lc}(P)\,\mathrm{lm}(P) + Q = \mathrm{lt}(P) + Q$, where all monomials in the expanded expression of Q are strictly smaller (with respect to the chosen monomial order) than $\mathrm{lm}(P)$.

Next, the concept of *polynomial reduction* is introduced, also called multivariate polynomial division, as it generalizes Euclidean division of univariate polynomials. For this purpose, we fix an admissible monomial order $\prec$ and use it in the following without any explicit mentioning. For nonzero polynomials $P,\ Q \in \Bbbk[\boldsymbol{x}]$, one says that P is *reducible* by Q if $\mathrm{lm}(P)$ is divisible by $\mathrm{lm}(Q)$. In other words, one can *reduce* P with respect to Q, and the result of the reduction is denoted by

$$\mathrm{red}_{\prec}(P,\ Q) = \mathrm{red}(P,\ Q) := P - \frac{\mathrm{lt}(P)}{\mathrm{lt}(Q)}\,Q.$$

It is important to notice that $\mathrm{red}(P,\ Q) = 0$ or $\mathrm{lm}\big(\mathrm{red}(P,\ Q)\big) \prec \mathrm{lm}(P)$. If $\mathcal{G} := \{G_1, \ldots, G_s\} \subseteq \Bbbk[\boldsymbol{x}] \setminus \{0\}$ is a set of polynomials, then $\mathrm{red}(P,\ \mathcal{G})$ denotes a polynomial obtained by iteratively reducing P with some elements of $\mathcal{G}$ until no such reduction is possible any more, i.e., the result is *irreducible* with respect to all elements of $\mathcal{G}$. Note that $\mathrm{red}(P,\ \mathcal{G})$ is usually not uniquely defined since it may depend on the choice of the polynomial G_i that is used in a certain reduction step as demonstrated in the following example.

Example 6 Let us consider $\mathbb{Q}[x_1, x_2]$ endowed with a total degree order (see (b) of Definition 4). Choosing $\mathcal{G} := \{G_1, G_2\}$ with $G_1 := x_1\,x_2 - 1$ and $G_2 := x_1^2 + x_2 + 1$, the monomial $x_1^2\,x_2$ can be reduced in two different ways yielding the two different irreducible polynomials $x_1^2\,x_2 - x_1\,G_1 = x_1$ and $x_1^2\,x_2 - x_2\,G_2 = -x_2^2 - x_2$.

Definition 6 Let $\langle \mathcal{G} \rangle$ denote the *ideal* generated by $G_1, \ldots, G_s \in \Bbbk[\boldsymbol{x}]$, i.e.:

$$\langle \mathcal{G} \rangle = \big\langle G_1, \ldots, G_s \big\rangle := \big\{ P_1\,G_1 + \cdots + P_s\,G_s \mid P_1, \ldots, P_s \in \Bbbk[\boldsymbol{x}] \big\}.$$

Then, $\mathcal{G}$ is called a *Gröbner basis* with respect to the admissible monomial order $\prec$ if one of the following equivalent statements holds:

(a) $P \in \langle \mathcal{G} \rangle$ if and only if $\mathrm{red}_{\prec}(P,\ \mathcal{G}) = 0$.
(b) $\mathrm{red}_{\prec}(P,\ \mathcal{G})$ is unique for any $P \in \Bbbk[\boldsymbol{x}]$.
(c) If $P \in \langle \mathcal{G} \rangle \setminus \{0\}$ then there exists $G_i \in \mathcal{G}$ such that $\mathrm{lm}_{\prec}(G_i)$ divides $\mathrm{lm}_{\prec}(P)$.
(d) $\big\langle \big\{ \mathrm{lm}_{\prec}(P) \mid P \in \langle \mathcal{G} \rangle \setminus \{0\} \big\} \big\rangle = \big\langle \mathrm{lm}_{\prec}(G_1), \ldots, \mathrm{lm}_{\prec}(G_s) \big\rangle$.

Condition (a) highlights one of the most important applications of Gröbner bases, namely the algorithmic decision of the *ideal membership problem*, i.e., given $P,\ G_1, \ldots, G_s \in \Bbbk[\boldsymbol{x}]$ decide whether $P \in \langle G_1, \ldots, G_s \rangle$. Having a Gröbner basis

at hand, this problem is solved by reducing P and checking whether the final reduction, called the *normal form* of P, is zero.

Example 7 We consider again Example 6 where we now set $\mathcal{G} := \{G_1, G_2, G_3\}$, where $G_3 := x_2^2 + x_1 + x_2$. Since $G_3 = x_2\, G_2 - x_1\, G_1$, we have $\langle \mathcal{G} \rangle = \langle G_1, G_2 \rangle$. We claim that $\mathcal{G}$ is a Gröbner basis (see below for an algorithm which computes a Gröbner basis). Now, the monomial $x_1^2\, x_2$ reduces to x_1 since the polynomial $-x_2^2 - x_2 = \mathrm{red}(x_1^2\, x_2,\ G_2)$ is now reducible by G_3 yielding $\mathrm{red}(-x_2^2 - x_2,\ G_3) = x_1$. No further reductions can be done. Hence, we obtain $x_1^2\, x_2 \notin \langle \mathcal{G} \rangle$.

Let us now shortly explain the principle of *Buchberger's algorithm* for computing a Gröbner basis of a polynomial ideal. Let $\mathrm{lcm}(m_1, m_2)$ denote the least common multiple of the two monomials m_1 and m_2. Buchberger's algorithm is based on the computation of the so-called *S-polynomials*.

Definition 7 Given $P,\ Q \in \Bbbk[\boldsymbol{x}] \setminus \{0\}$ and a monomial order $\prec$, we can define the *S-polynomial* $\mathrm{S}(P,\ Q)$ by:

$$\mathrm{S}(P,\ Q) := \frac{\mathrm{lcm}(\mathrm{lm}(P),\ \mathrm{lm}(Q))}{\mathrm{lt}(P)}\, P - \frac{\mathrm{lcm}(\mathrm{lm}(P),\ \mathrm{lm}(Q))}{\mathrm{lt}(Q)}\, Q.$$

Given a finite set $\{P_1, \ldots, P_r\}$ of elements of $\Bbbk[\boldsymbol{x}]$ and an admissible monomial order $\prec$ on $\boldsymbol{x}$-monomials, Buchberger's algorithm, which computes a Gröbner basis $\mathcal{G} := \{Q_1, \ldots, Q_s\}$ of the ideal $\langle P_1, \ldots, P_r \rangle$ of $\Bbbk[\boldsymbol{x}]$, can be sketched as follows:

(a) Set $\mathcal{G} := \{P_1, \ldots, P_r\}$ and let $\mathcal{P}$ be the set of pairs of distinct elements of $\mathcal{G}$;
(b) While $\mathcal{P} \neq \emptyset$, do:

- Choose $(P_i,\ P_j) \in \mathcal{P}$ and remove it from $\mathcal{P}$;
- Compute $S(P_i,\ P_j)$ and its reduction $R_{ij} := \mathrm{red}_{\prec}(S(P_i, P_j),\ \mathcal{G})$ by $\mathcal{G}$;
- If $R_{ij} \neq 0$, then:
  - Add $\{(P,\ R_{ij}) \mid P \in \mathcal{G}\}$ to $\mathcal{P}$;
  - Add R_{ij} to $\mathcal{G}$;

(c) Return $\mathcal{G}$.

One can prove that the latter process terminates with a Gröbner basis $\mathcal{G}$ of the ideal $\langle P_1, \ldots, P_r \rangle$. For more details, we refer to [1, 5, 17, 24, 26]. While an ideal admits many different Gröbner bases with respect to the same monomial order, one can achieve uniqueness by means of the following definition.

Definition 8 A Gröbner basis $\mathcal{G} := \{G_1, \ldots, G_s\}$ is said to be *reduced* if it satisfies the following two conditions:

- $\mathrm{lc}(G_i) = 1$ for $i = 1, \ldots, s$.
- Each monomial in G_i is irreducible with respect to $\mathcal{G} \setminus \{G_i\}$ for all $i = 1, \ldots, s$.

Example 8 The Gröbner basis in Example 7 is a reduced one.

Remark 2 For an ideal of $\Bbbk[\boldsymbol{x}]$ defined by a finite set of generators and a given monomial order $\prec$, one can compute a Gröbner basis, using, e.g., Buchberger's algorithm [5]. Algorithms for computing Gröbner bases are implemented in most of the computer algebra systems such as `Maple`, `Mathematica`, and `Magma`, or in dedicated computer algebra systems such as `Singular` and `Macaulay2`. However, in practice, such computations can be very costly, and it is still a topic of ongoing research to design faster algorithms for computing Gröbner bases. See the recent survey article [19] and the references therein.

Let us shortly state a few applications of Gröbner bases. Using the concept of a reduced Gröbner basis, we obtain a procedure to test whether or not two ideals of a commutative polynomial ring over a field, defined by different sets of generators, are equal: we check whether or not they have the same reduced Gröbner basis.

Solving a system of polynomial equations is an important application of Gröbner bases. For this purpose, we use the lexicographic order (see (a) of Example 5) which leads to a reduced Gröbner basis of a special form called "triangular" form. This means that some of the polynomials of the Gröbner basis depend only on certain variables, which simplifies the process of finding all solutions of the original system.

Example 9 Let $G_1,\ G_2 \in \mathbb{Q}[x_1, x_2]$ be as in Example 6 but now endowed with the lexicographic order (see (a) of Example 5). Then, $\{x_2^3 + x_2^2 + 1,\ x_1 + x_2^2 + x_2\}$ is a Gröbner basis with respect to this monomial order. Note that this Gröbner basis has a triangular form: the first element depends only on x_2. The solutions of the polynomial system $G_1 = G_2 = 0$ can be obtained by first solving $x_2^3 + x_2^2 + 1 = 0$ and then plugging the solutions for x_2 into $x_1 = -(x_2^2 + x_2)$.

Gröbner basis techniques can also be used to develop an *elimination theory*. Let us state a standard problem for ideals: if $I \subseteq \Bbbk[\boldsymbol{x}]$ is an ideal and $\boldsymbol{y}$ is a subset of $\boldsymbol{x}$, then compute generators for the ideal $I \cap \Bbbk[\boldsymbol{y}]$. To do that, we use the monomial order defined in (c) of Example 5. As explained in Sect. 1.4, elimination techniques play an important role in the effective study of module theory and homological algebra.

Example 10 If we consider again Example 9, we can check that we have:

$$\langle G_1,\ G_2\rangle \cap \mathbb{Q}[x_2] = \langle x_2^3 + x_2^2 + 1\rangle.$$

The theory of Gröbner bases has been extended to noncommutative polynomial rings. See the work of Bergman [3] for a very general and theoretic approach. A more algorithmically oriented but less general approach was presented in [26]. It only considers the so-called *rings of solvable type* (see also [31]). However, for our purposes, the latter suffices as most of the Ore algebras of interest are of solvable type. In this setting, again Buchberger's algorithm can be used to compute Gröbner bases, with only slight modifications due to noncommutativity.

Theorem 2 ([11, 31]) *Let $\Bbbk$ be a field, $\mathbb{A} := \Bbbk[x_1, \ldots, x_n]$ the polynomial ring with coefficients in $\Bbbk$, and $\mathbb{O} := \mathbb{A}[\partial_1; \sigma_1, \delta_1] \cdots [\partial_m; \sigma_m, \delta_m]$ a polynomial Ore algebra satisfying the following conditions*

$$\sigma_i(x_j) = a_{ij}\, x_j + b_{ij}, \quad \delta_i(x_j) = c_{ij}, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n,$$

for certain $a_{ij} \in \mathbb{k} \setminus \{0\}$, $b_{ij} \in \mathbb{k}$, *and* $c_{ij} \in \mathbb{A}$. *Let* $\prec$ *be an admissible monomial order on the following set of monomials:*

$$\mathrm{Mon}(\mathbb{O}) := \{x_1^{\alpha_1} \cdots x_m^{\alpha_m}\, \partial_1^{\nu_1} \cdots \partial_n^{\nu_n} \mid (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}^m,\ (\nu_1, \ldots, \nu_n) \in \mathbb{N}^n\}.$$

If the $\prec$*-greatest term* u *in each non-zero* c_{ij} *satisfies* $u \prec x_j\, \partial_i$, *then given a set of noncommutative polynomials in* $\mathbb{O}$, *a noncommutative version of Buchberger's algorithm terminates for this admissible monomial order and its result is a Gröbner basis with respect to this order.*

For more general results, we refer the reader to [26, 31, 37]. In particular, for the Weyl algebra $A_n(\mathbb{Q})$ (see (c) of Example 3), the existence of Gröbner bases and the generalization of Buchberger's algorithm have been studied, e.g., in [36, 39, 53].

Example 11 Let us consider $\mathbb{O} := B_2(\mathbb{Q})$ and the following linear PD system:

$$\begin{cases} \partial_1^2\, y = 0, \\ x_1\, \partial_2\, y + x_2\, y = 0. \end{cases} \tag{1.13}$$

Applying ∂_1 to the second equation of (1.13), we get $x_1\, \partial_1\, \partial_2\, y + \partial_2\, y + x_2\, \partial_1\, y = 0$. Applying again ∂_1 to the equation then yields $x_1\, \partial_1^2\, \partial_2\, y + 2\, \partial_1\, \partial_2\, y + x_2\, \partial_1^2\, y = 0$ and using (1.13), we get $\partial_1\, \partial_2\, y = 0$, and thus $\partial_2\, y + x_2\, \partial_1\, y = 0$. Eliminating $\partial_2\, y$ from the last equation by means of the second equation of (1.13), we obtain $x_1\, \partial_1\, y - y = 0$. If we now apply ∂_2 to the latter equation and use $\partial_1\, \partial_2\, y = 0$, we obtain $\partial_2\, y = 0$, which by substitution in the second equation of (1.13) gives $y = 0$. The solution of (1.13) is then $y = 0$, a fact which is not obvious from (1.13). The computation of a Gröbner basis for the left $\mathbb{O}$-ideal $I := \mathbb{O}\, \partial_1^2 + \mathbb{O}\, (x_1\, \partial_2 + x_2)$ for the total degree order follows the same line and yields $I = \mathbb{O}$.

1.3.2 Gröbner Bases for Modules over Ore Algebras

We now explain how we can extend the concept of a Gröbner basis from finitely generated left ideals to finitely generated left modules over an Ore algebra $\mathbb{O}$. Let us first state again the definition of a module.

Definition 9 Let D be a noncommutative ring. A *left D-module* M is an abelian group $(M, +)$ equipped with a scalar multiplication

$$\begin{aligned} D \times M &\longrightarrow M \\ (d, m) &\longmapsto d\, m, \end{aligned}$$

which satisfies the following properties

(a) $d_1\,(m_1 + m_2) = d_1\,m_1 + d_1\,m_2$,
(b) $(d_1 + d_2)\,m_1 = d_1\,m_1 + d_2\,m_1$,
(c) $(d_2\,d_1)\,m_1 = d_2\,(d_1\,m_1)$,
(d) $1\,m_1 = m_1$,

for all $d_1,\ d_2 \in D$ and for all $m_1,\ m_2 \in M$.

Remark 3 The definition of a left D-module is similar to the one of a vector space but where the scalars belong to a noncommutative ring D and not to a (skew) field (e.g., $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$) as for vector spaces.

A left D-module M is said to be *finitely generated* if M admits a finite set of generators, namely there exists a finite set $S := \{m_i\}_{i=1,\ldots,r}$ of elements of M such that for every $m \in M$, there exist $d_i \in D$ for $i = 1, \ldots, r$ such that:

$$m = \sum_{i=1}^{r} d_i\,m_i.$$

S is called a *set of generators* of M. Similar definitions hold for right D-modules.

In what follows, we consider D to be a polynomial Ore algebra $\mathbb{O}$. Let $\mathrm{Mon}(\mathbb{O})$ be the set of monomials of $\mathbb{O}$ and $\{f_j\}_{j=1,\ldots,p}$ the *standard basis* of the free finitely generated left $\mathbb{O}$-module $\mathbb{O}^{1\times p} := \{(\lambda_1\ \ldots\ \lambda_p) \mid \lambda_i \in \mathbb{O},\ i = 1, \ldots, p\}$, namely the k^{th} component of f_j is 1 if $k = j$ and 0 otherwise. First, we extend the monomial order $\prec$ from $\mathrm{Mon}(\mathbb{O})$ to the set of monomials of the form $u\,f_j$, where $u \in \mathrm{Mon}(\mathbb{O})$ and $j = 1, \ldots, p$, i.e., to $\mathrm{Mon}(\mathbb{O}^{1\times p}) := \bigcup_{j=1}^{p} \mathrm{Mon}(\mathbb{O})\,f_j$. This extension is also denoted by $\prec$ and it has to satisfy the following two conditions:

(a) $\forall\ w \in \mathrm{Mon}(\mathbb{O}) :\ u\,f_i \prec v\,f_j \implies w\,u\,f_i \prec w\,v\,f_j$.
(b) $u \prec v \implies u\,f_j \prec v\,f_j$ for $j = 1, \ldots, p$.

Without loss of generality, we let $f_p \prec f_{p-1} \prec \cdots \prec f_1$. There are two natural extensions of a monomial order to $\mathrm{Mon}(\mathbb{O}^{1\times p})$.

Definition 10 Let $\prec$ be an admissible monomial order on $\mathrm{Mon}(\mathbb{O})$, $u,\ v \in \mathrm{Mon}(\mathbb{O})$, and $\{f_j\}_{j=1,\ldots,p}$ the standard basis of the left $\mathbb{O}$-module $\mathbb{O}^{1\times p}$.

(a) The *term over position order* on $\mathrm{Mon}(\mathbb{O}^{1\times p})$ induced by $\prec$ is defined by:

$$u\,f_i \prec v\,f_j \iff u \prec v \ \text{ or } \ u = v \ \text{ and } \ f_i \prec f_j.$$

(b) The *position over term order* on $\mathrm{Mon}(\mathbb{O}^{1\times p})$ induced by $\prec$ is defined by:

$$u\,f_i \prec v\,f_j \iff f_i \prec f_j \ \text{ or } \ f_i = f_j \ \text{ and } \ u \prec v.$$

Remark 4 The term over position order is of more computational value with regard to efficiency. The position over term order can be used to eliminate components.

If an admissible monomial order on $\mathrm{Mon}(\mathbb{O}^{1\times p})$ is fixed, then leading monomials and leading coefficients in $\mathbb{O}^{1\times p}$ are defined similarly as in the case of ideals. Let $R \in \mathbb{O}^{q\times p}$ and $L := \mathbb{O}^{1\times q}\, R$ be the left $\mathbb{O}$-submodule of $\mathbb{O}^{1\times p}$. Buchberger's algorithm carries over to L. For more details, we refer, e.g., to [18, 24].

Example 12 We consider again Example 4. Let us compute a Gröbner basis of the $\mathbb{O} := \mathbb{Q}(\nu, c)\left[\partial_x, \partial_y\right]$-submodule $L := \mathbb{O}^{1\times 3}\, R$ of $\mathbb{O}^{1\times 3}$, i.e.,

$$(-\nu\,\Delta + c)\,f_1 + \partial_x\,f_3, \quad (-\nu\,\Delta + c)\,f_2 + \partial_y\,f_3, \quad \partial_x\,f_1 + \partial_y\,f_2,$$

for the position over term order induced by the monomial order $\prec_{\mathrm{tdeg}}$ (see (b) of Example 5). The Gröbner basis of L is then given by:

$$\partial_x\,f_1 + \partial_y\,f_2,\ (-\nu\,\partial_y^2 + c)\,f_1 + \nu\,\partial_x\,\partial_y\,f_2 + \partial_x\,f_3,\ \Delta\,f_3,\ (-\nu\,\Delta + c)\,f_2 + \partial_y\,f_3.$$

We find again that the pressure p satisfies $\Delta\,p = 0$ as shown in Example 4.

Let us shortly explain how Gröbner basis techniques can be used to compute left kernels (syzygy module computation), left factorizations and left inverses, … of matrices with entries in $\mathbb{O}$. For more details, we refer to [9].

Algorithm 1 Computation of the left kernel of $R \in \mathbb{O}^{q\times p}$, i.e., find $S \in \mathbb{O}^{r\times q}$ such that $\ker_{\mathbb{O}}(.R) := \{\lambda \in \mathbb{O}^{1\times q} \mid \lambda\,R = 0\} = \mathbb{O}^{1\times r}\,S := \{\mu\,S \mid \mu \in \mathbb{O}^{1\times r}\}$.

- **Input:** An Ore algebra $\mathbb{O}$ satisfying the hypotheses of Theorem 2 and a finitely generated left $\mathbb{O}$-submodule $L := \mathbb{O}^{1\times q}\,R$ of $\mathbb{O}^{1\times p}$, where $R \in \mathbb{O}^{q\times p}$.
- **Output:** A matrix $S \in \mathbb{O}^{r\times q}$ such that $\ker_{\mathbb{O}}(.R) = \mathbb{O}^{1\times r}\,S$.

(a) Introduce the indeterminates $\eta_1, \ldots, \eta_p, \zeta_1, \ldots, \zeta_q$ over $\mathbb{O}$ and define the set:

$$\mathcal{P} := \left\{ \sum_{j=1}^{p} R_{ij}\,\eta_j - \zeta_i \mid i = 1, \ldots, q \right\}.$$

(b) Compute a Gröbner basis $\mathcal{G}$ of $\mathcal{P}$ in the free left $\mathbb{O}$-module generated by the η_j's and the ζ_i's for $j = 1, \ldots, p$ and $i = 1, \ldots, q$, namely, $\bigoplus_{j=1}^{p} \mathbb{O}\,\eta_j \bigoplus \bigoplus_{i=1}^{q} \mathbb{O}\,\zeta_i$, with respect to a term order which eliminates the η_j's (see (c) of Example 5).

(c) Compute $\mathcal{G} \cap \left(\bigoplus_{i=1}^{q} \mathbb{O}\,\zeta_i\right) = \left\{\sum_{i=1}^{q} S_{ki}\,\zeta_i \mid k = 1, \ldots, r\right\}$ by selecting the elements of $\mathcal{G}$ containing only the ζ_i's, and return $S := (S_{ij}) \in \mathbb{O}^{r\times q}$.

Algorithm 2 Computation of a left factorization: given two matrices $R \in \mathbb{O}^{q\times p}$ and $R' \in \mathbb{O}^{q'\times p}$, find a matrix $R'' \in \mathbb{O}^{q\times q'}$ (if it exists) satisfying $R = R''\,R'$.

- **Input:** An Ore algebra $\mathbb{O}$ satisfying the hypotheses of Theorem 2 and two matrices $R \in \mathbb{O}^{q\times p}$ and $R' \in \mathbb{O}^{q'\times p}$.
- **Output:** $R'' \in \mathbb{O}^{q\times q'}$ (if it exists) such that $R = R''\,R'$ and [ ] otherwise.

(a) Introduce the indeterminates $\eta_1, \ldots, \eta_p, \zeta_1, \ldots, \zeta_{q'}$ over $\mathbb{O}$ and define the set:

$$\mathcal{P} := \left\{ \sum_{j=1}^{p} R'_{ij}\, \eta_j - \zeta_i \mid i = 1, \ldots, q' \right\}.$$

(b) Compute a Gröbner basis $\mathcal{G}$ of $\mathcal{P}$ in the free left $\mathbb{O}$-module generated by the η_j's and the ζ_i's for $j = 1, \ldots, p$ and $i = 1, \ldots, q'$, namely, $\bigoplus_{j=1}^{p} \mathbb{O}\, \eta_j \bigoplus \bigoplus_{i=1}^{q'} \mathbb{O}\, \zeta_i$ with respect to a term order which eliminates the η_j's (see (c) of Example 5).

(c) Define the following set:

$$\mathcal{Q} := \left\{ \sum_{j=1}^{p} R_{kj}\, \eta_j \mid k = 1, \ldots, q \right\}.$$

(d) Compute the reduction H_i of each element Q_i of $\mathcal{Q}$ by $\mathcal{G}$.

(e) If one of the H_i's contains η_j, i.e., if the normal form of Q_i contains not only ζ_i's, then return [ ], else return $R'' := (R''_{ij}) \in \mathbb{O}^{q \times q'}$, where $H_i = \sum_{j=1}^{q'} R''_{ij}\, \zeta_j$ for $i = 1, \ldots, q$.

Algorithm 3 Computation of a left inverse: given a matrix $R \in \mathbb{O}^{q \times p}$, find (if it exists) a left inverse $S \in \mathbb{O}^{p \times q}$ of R over $\mathbb{O}$, namely $S\, R = I_p$.

- **Input:** An Ore algebra $\mathbb{O}$ satisfying the hypotheses of Theorem 2 and $R \in \mathbb{O}^{q \times p}$.
- **Output:** A matrix $S \in \mathbb{O}^{p \times q}$ such that $S\, R = I_p$ if S exists and [ ] otherwise.

(a) Introduce the indeterminates $\eta_1, \ldots, \eta_p, \zeta_1, \ldots, \zeta_q$ over $\mathbb{O}$ and define the set:

$$\mathcal{P} := \left\{ \sum_{j=1}^{p} R_{ij}\, \eta_j - \zeta_i \mid i = 1, \ldots, q \right\}.$$

(b) Compute a Gröbner basis $\mathcal{G}$ of $\mathcal{P}$ in the free left $\mathbb{O}$-module generated by the η_j's and the ζ_i's for $j = 1, \ldots, p$ and $i = 1, \ldots, q$, namely, $\bigoplus_{j=1}^{p} \mathbb{O}\, \eta_j \bigoplus \bigoplus_{i=1}^{q} \mathbb{O}\, \zeta_i$, with respect to a term order which eliminates the η_j's (see (c) of Example 5).

(c) Remove from $\mathcal{G}$ the elements which do not contain any η_i and call $\mathcal{H}$ this new set.

(d) Write $\mathcal{H}$ in the form $Q_1\, (\eta_1 \ \ldots \ \eta_p)^T - Q_2\, (\zeta_1 \ \ldots \ \zeta_q)^T$, where Q_1 and Q_2 are two matrices with entries in $\mathbb{O}$.

(e) If Q_1 is invertible over $\mathbb{O}$, then return $S := Q_1^{-1}\, Q_2 \in \mathbb{O}^{p \times q}$, else return [ ].

Right analogues of the above algorithms (i.e., computation of right kernels, right factorizations, and right inverses) can be obtained by considering an involution of the Ore algebra $\mathbb{O}$ (see Definition 3). For instance, the computation of a right inverse

of a matrix $R \in \mathbb{O}^{q\times p}$ over an Ore algebra can be done by applying Algorithm 3 to the matrix $\theta(R) := (\theta(R_{ij}))^T \in \mathbb{O}^{p\times q}$ (obtained by applying an involution θ of $\mathbb{O}$ to each entry R_{ij} of R and then transposing the result) and applying the involution to the left inverse $T \in \mathbb{O}^{q\times p}$ of $\theta(R)$ to get $S := \theta(T) \in \mathbb{O}^{p\times q}$ which then satisfies:

$$R\,S = \theta^2(R)\,\theta(T) = \theta(T\,\theta(R)) = \theta(I_q) = I_q.$$

For an implementation of these algorithms in a computer algebra system, see the OreModules package [10].

1.4 Algebraic Analysis Approach to Linear Systems Theory

1.4.1 Linear Functional Systems and Finitely Presented Left Modules

As explained in Sect. 1.1, we study linear functional systems of the form $R\,\eta = 0$, where $R \in D^{q\times p}$, D is a noetherian domain (e.g., a noetherian Ore algebra $\mathbb{O}$ of functional operators (see Sect. 1.2)), and η is a vector of unknown functions. More precisely, if $\mathcal{F}$ is a left D-module (see Definition 9), then we can consider the following linear system or *behavior*:

$$\ker_{\mathcal{F}}(R.) := \{\eta \in \mathcal{F}^p \mid R\,\eta = 0\}.$$

See Example 1 for the different models of the stirred tank considered in Sect. 1.1.

Remark 5 In this framework, we can consider the following classes of systems:

- State-space/input-output representation of 1-D linear systems. Considering, e.g.,

$$R := (\partial\, I_n - A \quad -B) \in \mathbb{O}^{n\times(n+m)},\ \eta := (x(t)^T \quad u(t)^T)^T \in \mathcal{F}^{n+m},$$
$$R := (P(\partial) \quad -Q(\partial)) \in \mathbb{O}^{q\times(q+r)},\ \eta := (y(t)^T \quad u(t)^T)^T \in \mathcal{F}^{q+r},$$

  where $\mathbb{O} := \mathbb{A}\left[\partial; \mathrm{id}_A, \frac{d}{dt}\right]$ is a ring of OD operators with coefficients in a differential ring $\mathbb{A}$ and P has full row rank (i.e., $\ker_{\mathbb{O}}(.P) = 0$), we obtain the linear systems $\dot{x}(t) = A(t)\,x(t) + B(t)\,u(t)$ and $P(\partial)\,y(t) = Q(\partial)\,u(t)$. Similarly, we can consider the Ore algebra $\mathbb{O} := \mathbb{A}[S; \sigma, 0]$ of shift operators with coefficients in the difference ring $\mathbb{A}$ and S instead of ∂ in the above matrices to get the linear systems $x_{k+1} = A_k\,x_k + B_k\,u_k$ and $P(S)\,y_k = Q(S)\,u_k$.
- In the first above example, if we consider the Ore algebra $\mathbb{O} := \mathbb{A}\left[\partial; \mathrm{id}_A, \frac{d}{dt}\right]$, where $\mathbb{A} := \mathbb{B}[S; \sigma, 0]$ and $\mathbb{B}$ is a difference ring, then we obtain the system $\dot{x}(t) = A(t, S)\,x(t) + B(t, S)\,u(t)$, called in the literature a system over ring. Note that a general linear differential constant time-delay system is defined by $R\,\eta = 0$, where $R \in \mathbb{O}^{q\times p}$, $\eta \in \mathcal{F}^p$ and, e.g., $\mathcal{F} = C^\infty(\mathbb{R}_{\geq 0})$.

- General linear nD systems can be defined by $R\,\eta = 0$, where $R \in \mathbb{O}^{q\times p}$ and $\mathbb{O}$ is, for instance, one of the Ore algebras considered in Example 2. For instance, a simple discrete Roesser model can be defined by $R\,\eta = 0$, where

$$R := \begin{pmatrix} S_1\, I_{r_h} - A_{11} & -A_{12} & -B_1 \\ -A_{21} & S_2\, I_{r_v} - A_{22} & -B_2 \end{pmatrix} \in \mathbb{O}^{(r_h+r_v)\times(r_h+r_v+m)},$$

$\eta := (x_h^T \;\; x_v^T \;\; u^T)^T$, $x_h \in \mathcal{F}^{r_h}$, $x_v \in \mathcal{F}^{r_v}$, $u \in \mathcal{F}^m$, and $\mathbb{O}$ is the Ore algebra defined by (c) in Example 2. Continuous or a mixed continuous and discrete Roesser model can be defined similarly using the other Ore algebras defined in Example 2.

Linear systems (e.g., a linearization of a nonlinear system around a given solution) can be studied within the algebraic analysis approach. The next example explains how the generic linearization of a nonlinear system can also be studied.

Example 13 We consider the nonlinear OD system defined by

$$\dot{x}(t) = f(x(t), u(t)), \tag{1.14}$$

where we first suppose that $f = (f_1 \;\cdots\; f_n)^T$, where f_i is a polynomial for $i = 1, \ldots, n$. Let us denote $X := X_1, \ldots, X_n$ and $U := U_1, \ldots, U_m$. Let k be a *differential field* (e.g., a field which is a differential ring), $k\{X, U\}$ the differential ring formed by polynomials in a finite number of the X_i's, U_j's, and of their derivatives with coefficients in k, and $\mathfrak{p}$ the *differential ideal* defined by the differential polynomials $\dot{X}_i - f_i(X, U)$ for $i = 1, \ldots, n$, and their derivatives. Then, we can define the ring $\mathbb{A} := k\{X, U\}/\mathfrak{p}$ formed by the differential polynomials modulo the ideal $\mathfrak{p}$. If we denote by x_i (resp., u_j) the residue class of X_i (resp., U_j) in $\mathbb{A}$, $x := x_1, \ldots, x_n$, $u := u_1, \ldots, u_m$, and $\mathbb{A} = k\{x, u\}$, then these polynomials can be rewritten as polynomials in x_i, u_j, and the derivatives of the u_j's. Clearly, $\mathbb{A}$ is a differential ring with the derivation $\delta := \frac{d}{dt}$. It can be proved that $\mathfrak{p}$ is a *prime ideal*, i.e., that $\mathbb{A}$ is an integral domain. Thus, we can define the quotient field $\mathbb{K} := Q(\mathbb{A})$ of $\mathbb{A}$, i.e., the ring of fractions of $\mathbb{A}$, which is a differential field for the derivation δ. Let $\mathbb{O} := \mathbb{B}[\partial; \mathrm{id}_{\mathbb{A}}, \delta]$ be the skew polynomial ring of OD operators with coefficients in $\mathbb{B} := \mathbb{A}$ or $\mathbb{K}$. The generic linearization of (1.14) is then defined by $R\,\eta = 0$, where $R := \left(\partial\, I_n - \frac{\partial f}{\partial x} \;\; -\frac{\partial f}{\partial u}\right) \in D^{n\times(n+m)}$ and $\eta := (dx^T \;\; du^T)^T$, and can be studied by means of the finitely presented left $\mathbb{O}$-module $M := \mathbb{O}^{1\times(n+m)}/(\mathbb{O}^{1\times n}\, R)$. The cases of a rational, analytic or meromorphic function f can be studied similarly by considering the differential ring or field $\mathbb{B}$ formed by the rational/analytic/meromorphic functions which satisfy (1.14).

Within the algebraic analysis approach to linear systems theory [9, 43, 47, 51, 57], the linear system or behaviour is studied by means of the *factor* left D-module

$$M := D^{1\times p}/(D^{1\times q}\, R)$$

formed by the set of the *residue classes* $\pi(\lambda)$ of $\lambda \in D^{1\times p}$ modulo the left D-submodule $L := D^{1\times q}\, R$ of $D^{1\times p}$ (i.e., $\pi(\lambda) = \pi(\lambda')$ if there exists $\mu \in D^{1\times q}$ such that $\lambda = \lambda' + \mu\, R$) and equipped with the following left D-module structure:

$$\forall\, \lambda,\ \lambda' \in D^{1\times p},\ \forall\, d \in D,\ \pi(\lambda) + \pi(\lambda') := \pi(\lambda + \lambda'),\ d\,\pi(\lambda) := \pi(d\,\lambda).$$

Remark 6 If $D := \mathbb{O}$ is an Ore algebra satisfying the hypotheses of Theorem 2, then we can check if $\pi(\lambda) = \pi(\lambda')$ for $\lambda,\ \lambda' \in \mathbb{O}^{1\times p}$ since $\lambda - \lambda' \in L := \mathbb{O}^{1\times q}\, R$ if and only if $\mathrm{red}(\lambda - \lambda',\ \mathcal{G}) = 0$, where $\mathcal{G}$ is a Gröbner basis of L (see Sect. 1.3).

The left D-module M is said to be *finitely presented* and R is called a *presentation matrix* [52]. If $\{f_j\}_{j=1,\ldots,p}$ is the standard basis of $D^{1\times p}$ and $y_j := \pi(f_j)$ for $j = 1, \ldots, p$, then $\{y_j\}_{j=1,\ldots,p}$ is a set of generators of the left D-module M (see Sect. 1.3.2). The generators y_j's of M satisfy non-trivial relations since we have:

$$\sum_{j=1}^{p} R_{ij}\, y_j = 0, \quad i = 1, \ldots, q.$$

For the details of these results, see Chap. 2 of this book. Note that the y_j's do not belong to $\mathcal{F}$ but are just elements of M. To speak about $\mathcal{F}$-solutions of $R\,\eta = 0$, we have to consider the *homomorphisms* from M to $\mathcal{F}$, namely the maps $f : M \longrightarrow \mathcal{F}$ satisfying the following (left D-linear) condition:

$$\forall\, d_1,\ d_2 \in D,\ \forall\, m_1,\ m_2 \in M,\ f(d_1\, m_1 + d_2\, m_2) = d_1\, f(m_1) + d_2\, f(m_2).$$

We recall that $f \in \hom_D(M, \mathcal{F})$ is said to be an *isomorphism* if f is both injective and surjective [52]. If an isomorphism exists between M and $\mathcal{F}$, then we say that M and $\mathcal{F}$ are *isomorphic*, which is denoted by $M \cong \mathcal{F}$.

A standard result of homological algebra concerning the *left exactness* of the contravariant functor $\hom_D(\cdot\,, \mathcal{F})$ [52] yields the following fundamental result for the algebraic analysis approach of linear systems theory (also called *Malgrange's isomorphism*).

Theorem 3 *We have the following isomorphism of abelian groups (i.e., $\mathbb{Z}$-modules):*

$$\ker_{\mathcal{F}}(R.) \cong \hom_D(M, \mathcal{F}). \tag{1.15}$$

For a direct proof of Theorem 3, see Chap. 2 of this book.

Remark 7 If D is not a commutative ring, then neither $\ker_{\mathcal{F}}(R.)$ nor $\hom_D(M, \mathcal{F})$ are left D-modules. For instance, if we consider $D := A_1(\mathbb{Q})$, $R := \partial + \frac{t-m}{\sigma^2} \in D$ where t, m and σ are constants parameters (e.g., transcendental elements over $\mathbb{Q}$), and $M := D/(D\, R)$, then $\eta := e^{-\frac{(t-m)^2}{2\sigma^2}} \in \ker_{\mathcal{F}}(R.)$, where $\mathcal{F} := C^\infty(\mathbb{R})$. But

$$R\,(\partial\,\eta) = \left(\partial^2 + \partial\,\frac{(t-m)}{\sigma^2} - \frac{1}{\sigma^2}\right)\eta = \partial\,\left(\partial + \frac{t-m}{\sigma^2}\right)\eta - \frac{1}{\sigma^2}\,\eta = -\frac{1}{\sigma^2}\,\eta,$$
$$R\,(t\,\eta) = (t\,\partial + 1)\,\eta + t\,\frac{(t-m)}{\sigma^2}\,\eta = t\,\left(\partial + \frac{t-m}{\sigma^2}\right)\eta + \eta = \eta,$$

which shows that neither $\dot{\eta}$ nor $t\,\eta$ belongs to $\ker_{\mathcal{F}}(R.)$, i.e., $\ker_{\mathcal{F}}(R.)$ has no left D-module structure. However they are abelian groups (i.e., the $\mathbb{Z}$-modules) and $\Bbbk$-vector spaces if D is a $\Bbbk$-algebra and $\Bbbk$ is a field included in the *center* of D:

$$Z(D) := \{d \in D \mid d\,D = D\,d\}.$$

If $\mathcal{F} := D$, then $\hom_D(M, D)$ inherits a right D-module structure [47, 52].

Using the isomorphism (1.15), the linear system $\ker_{\mathcal{F}}(R.)$ depends only on M and $\mathcal{F}$. Hence, we can study its built-in properties by means of those of the modules M and $\mathcal{F}$. Note that the functional space $\mathcal{F}$ where the solutions are sought can be altered and the behaviour of the solutions highly depend on it (in a similar way as for the $\mathcal{F}$-solutions of $x^2 + 1 = 0$ for $\mathcal{F} := \mathbb{R}$ or $\mathbb{C}$) [43]. In what follows, we will suppose that $\mathcal{F}$ is a rich enough functional space (i.e., is an *injective cogenerator* left D-module [52]) so that $\mathcal{F}$ plays a similar role as the algebraic closure in algebraic geometry. Hence, we can study the properties of $\ker_{\mathcal{F}}(R.)$ by means of those of M. For the study of the role of $\mathcal{F}$, we refer to [43, 57] and the references therein.

We also note that $\hom_D(M, \mathcal{F})$ depends only on the *isomorphism type* of M, i.e., if $M \cong M'$, then we have $\hom_D(M, \mathcal{F}) \cong \hom_D(M', \mathcal{F})$. If M (resp., M') is finitely presented by $R \in D^{q\times p}$ (resp., $R' \in D^{q'\times p'}$), then we get

$$\ker_{\mathcal{F}}(R.) \cong \ker_{\mathcal{F}}(R'.),$$

i.e., there is a 1–1 correspondence between the solutions of the first system and the solutions of the second one. For more details and applications of this result to Serre's reduction, Stafford's reduction, the decomposition problem, see [12, 14, 47] and the references therein. Two different representations $R \in D^{q\times p}$ and $R' \in D^{q'\times p'}$ of the *same linear system* define two isomorphic modules:

$$M := D^{1\times p}/(D^{1\times q}\,R) \cong M := D^{1\times p'}/(D^{1\times q'}\,R').$$

Homological algebra methods are developed to study modules up to isomorphism. In particular, even if a particular representation is used, fundamental theorems in homological algebra show that the results do not depend on it. For mathematical systems theory, it is a change of paradigm since systems are usually studied by means of their particular representations (e.g., state-space or polynomial representations). The equivalence between the different approaches is studied below. Within the algebraic analysis approach, we first define the equivalence of linear systems in terms of isomorphic left D-modules finitely presented by these representations, and then use mathematical methods which do only depend on the isomorphism type. For instance,

if the concept of controllability is a built-in property of the linear system and not of its representation, then it should be a module property. For standard classes of linear systems, it has been shown that certain definitions of controllability correspond to the concept of a *torsion-free module* (for more details, see Sect. 1.4.3). Let us introduce basic definitions of module theory [16, 34, 52].

Definition 11 Let D be a noetherian domain and M a finitely generated left D-module.

(a) M is *free* if there exists $r \in \mathbb{N}$ such that $M \cong D^{1\times r}$. Then, r is called the *rank* of the free left D-module M and is denoted by $\mathrm{rank}_D(M)$.
(b) M is *stably free* if there exist $r,\ s \in \mathbb{N}$ such that $M \oplus D^{1\times s} \cong D^{1\times r}$. Then, $r - s$ is called the *rank* of the stably free left D-module M.
(c) M is *projective* if there exist $r \in \mathbb{N}$ and a left D-module N such that

$$M \oplus N \cong D^{1\times r},$$

where $\oplus$ denotes the direct sum of left D-modules.
(d) M is *reflexive* if the canonical left D-homomorphism

$$\varepsilon : M \longrightarrow \hom_D(\hom_D(M, D), D), \quad \varepsilon(m)(f) = f(m),$$

for all $f \in \hom_D(M, D)$ and all $m \in M$, is an isomorphism.
(e) M is *torsion-free* if the *torsion left D-submodule* $t(M)$ of M is 0, where:

$$t(M) := \{m \in M \mid \exists\, d \in D \setminus \{0\} :\ d\, m = 0\}.$$

The elements of $t(M)$ are called the *torsion elements* of M.
(f) M is *torsion* if $t(M) = M$, i.e., if every element of M is a torsion element.

Considering $s = 0$ in (b) (resp., $N := D^{1\times s}$ in (c)) of Definition 11, a free (resp., stably free) module is stably free (resp., projective). A projective module is torsion-free since it can be embedded into a free, and thus into a torsion-free module. The converse of these results are not usually true. In some particular cases, they can hold.

Theorem 4 ([16, 34, 49, 52]) *We have the following results.*

(a) If D is a principal ideal domain*, i.e., every left/right ideal of the domain D is principal (e.g., $D := \mathbb{A}\left[\partial; \mathrm{id}_A, \frac{d}{dt}\right]$, where $\mathbb{A} := \mathbb{k}$, $\mathbb{k}(t)$, or $\mathbb{k}[\![t]\!][t^{-1}]$ is the field of formal Laurent power series, where $\mathbb{k}$ is a field of characteristic* 0*, or $\mathbb{A} := \mathbb{k}\{t\}[t^{-1}]$ is the field of Laurent power series, where $\mathbb{k} := \mathbb{R}$, $\mathbb{C}$), then every finitely generated torsion-free left/right D-module is free.*
*(b) If $D := \mathbb{k}[x_1, \ldots, x_n]$ is a commutative polynomial ring with coefficients in a field $\mathbb{k}$, then every finitely generated projective D-module is free (*Quillen–Suslin theorem*).*

(c) *If D is the Weyl algebra $A_n(\mathbb{k})$ or $B_n(\mathbb{k})$, where $\mathbb{k}$ is a field of characteristic 0, then every finitely generated projective left/right D-module is stably free and every finitely generated stably free left/right D-module of rank at least 2 is free* (Stafford's theorem).

(d) *If $D := \mathbb{A}\left[\partial; \mathrm{id}_A, \frac{d}{dt}\right]$ where $\mathbb{A} := \mathbb{k}[\![t]\!]$ is the ring of formal power series in t and $\mathbb{k}$ is a field of characteristic 0, or $\mathbb{A} := \mathbb{k}\{t\}$ is the ring of locally convergent power series in t, where $\mathbb{k} := \mathbb{R}$ or $\mathbb{C}$, then every finitely generated projective left/right D-module is stably free and every finitely generated stably free left/right D-module of rank at least 2 is free.*

In Sect. 1.4.3, we will give a dictionary between properties of a linear functional system and properties of the finitely presented left module associated with it.

1.4.2 Basic Results of Homological Algebra

In this section, we briefly review how to effectively check whether or not a finitely presented left D-module M has torsion elements, is torsion-free, reflexive, or projective (see Definition 11), when D is a noetherian domain with *finite global dimension* [52]. To do that, let us introduce a few concepts of homological algebra [52].

Let D be a noetherian domain, $R \in D^{q\times p}$, and $M := D^{1\times p}/(D^{1\times q}\, R)$ the left D-module finitely presented by R. If $.R \in \hom_D(D^{1\times q}, D^{1\times p})$ is defined by

$$\begin{aligned} .R\colon D^{1\times q} &\longrightarrow D^{1\times p} \\ \lambda &\longmapsto \lambda\, R, \end{aligned}$$

then we obtain $\mathrm{coker}_D(.R) = D^{1\times p}/\mathrm{im}_D(.R) = D^{1\times p}/(D^{1\times q}\, R) = M$. Since D is a left noetherian ring, $D^{1\times q}$ is a *noetherian* left D-module, i.e., every left D-submodule of $D^{1\times q}$ is finitely generated [52]. In particular, $\ker_D(.R)$ is a finitely generated left D-module, i.e., there exists a finite generator set $\{\lambda_i\}_{i=1,\ldots,r}$ of $\ker_D(.R)$. Then, we have $\ker_D(.R) = \mathrm{im}_D(.R_2) = D^{1\times r}\, R_2$, with the notation $R_2 := (\lambda_1^T \ \ldots \ \lambda_r^T)^T \in D^{r\times q}$. Let us introduce a few definitions.

Definition 12 (a) A *complex* of left D-modules is a sequence of left D-modules M_i and D-homomorphisms $d_i : M_i \longrightarrow M_{i-1}$ for $i \in \mathbb{Z}$ such that $d_i \circ d_{i+1} = 0$, i.e., $\mathrm{im}\, d_{i+1} \subseteq \ker d_i$ for all $i \in \mathbb{Z}$. Such a complex is denoted by:

$$\ldots \xrightarrow{d_{i+2}} M_{i+1} \xrightarrow{d_{i+1}} M_i \xrightarrow{d_i} M_{i-1} \xrightarrow{d_{i-1}} M_{i-2} \xrightarrow{d_{i-2}} \ldots \tag{1.16}$$

(b) The *defect of exactness* of (1.16) at M_i is the left D-module $H(M_i) := \ker d_i / \mathrm{im}\, d_{i+1}$.

(c) The complex (1.16) is said to be *exact at* M_i if $H(M_i) = 0$, i.e., $\ker d_i = \mathrm{im}\, d_{i+1}$, and *exact* if $H(M_i) = 0$ for all $i \in \mathbb{Z}$.

(d) An exact sequence of the form

$$\ldots \xrightarrow{.R_3} D^{1\times p_2} \xrightarrow{.R_2} D^{1\times p_1} \xrightarrow{.R_1} D^{1\times p_0} \xrightarrow{\pi} M \longrightarrow 0, \qquad (1.17)$$

where $R_i \in D^{p_i\times p_{i-1}}$ and $.R_i \in \hom_D(D^{1\times p_i}, D^{1\times p_{i-1}})$ is defined by $(.R_i)\lambda = \lambda R_i$ for all $\lambda \in D^{1\times p_i}$, is called a *free resolution* of M.

With $R_1 = R$, we can easily check that we have the following exact sequence

$$0 \longrightarrow \ker_D(.R_2) \xrightarrow{i} D^{1\times r} \xrightarrow{.R_2} D^{1\times q} \xrightarrow{.R_1} D^{1\times p} \xrightarrow{\pi} M \longrightarrow 0,$$

where i is the canonical injection and π the canonical projection. Repeating for R_2 what we did for R and so on, we get a free resolution (1.17) of M. If $D := \mathbb{O}$ is an Ore algebra satisfying the hypotheses of Theorem 2, then Algorithm 1 can be used to compute a free resolution of M.

Applying the left exact contravariant functor $\hom_D(\cdot, \mathcal{F})$ [52] to the complex

$$\ldots \xrightarrow{.R_3} D^{1\times p_2} \xrightarrow{.R_2} D^{1\times p_1} \xrightarrow{.R_1} D^{1\times p_0} \longrightarrow 0,$$

obtained by removing M from (1.17)—called a *truncated free resolution* of M—and using $\hom_D(D^{1\times p_i}, \mathcal{F}) \cong \mathcal{F}^{p_i}$, we then obtain the following complex

$$\ldots \xleftarrow{R_3.} \mathcal{F}^{p_2} \xleftarrow{R_2.} \mathcal{F}^{p_1} \xleftarrow{R_1.} \mathcal{F}^{p_0} \longleftarrow 0, \qquad (1.18)$$

where $R_i. : \mathcal{F}^{p_{i-1}} \longrightarrow \mathcal{F}^{p_i}$ is defined by $(R_i.)\,\eta = R_i\,\eta$ for all $\eta \in \mathcal{F}^{p_{i-1}}$. The *extension $\mathbb{Z}$-modules* $\mathrm{ext}^i_D(M, \mathcal{F})$ are then the defects of exactness of (1.18).

Theorem 5 ([52]) *The defects of exactness of (1.18) depend only on M and $\mathcal{F}$, i.e., they do not depend on the choice of the free resolution (1.17) of M. These abelian groups are denoted by:*

$$\begin{cases} \mathrm{ext}^0_D(M, \mathcal{F}) = \hom_D(M, \mathcal{F}) = \ker_{\mathcal{F}}(R_1.), \\ \mathrm{ext}^i_D(M, \mathcal{F}) = \ker_{\mathcal{F}}(R_{i+1}.)/\,\mathrm{im}_{\mathcal{F}}(R_i.), \quad i \geq 1. \end{cases}$$

Theorem 5 is a fundamental result of homological algebra. It shows that the $\mathrm{ext}^i_D(M, \mathcal{F})$'s do not depend on a particular representation of the linear system.

Remark 8 Let us give an interpretation of the $\mathrm{ext}^i_D(M, \mathcal{F})$'s. They define the obstructions of the *solvability problem* which aims at finding $\eta \in \mathcal{F}^{p_{i-1}}$ which satisfies the inhomogeneous linear system $R_i\,\eta = \zeta$ for a fixed $\zeta \in \mathcal{F}^{p_i}$. Indeed, if such an η exists, then we have $R_{i+1}\,\zeta = R_{i+1}\,(R_i\,\eta) = 0$ since $\ker_D(.R_i) = D^{1\times p_{i+1}}\,R_{i+1}$, i.e., $\zeta \in \ker_{\mathcal{F}}(R_{i+1}.)$. This condition is a necessary one for the solvability problem. This problem is solvable if and only if the residue class of ζ in $\mathrm{ext}^i_D(M, \mathcal{F})$ is 0, i.e., if and only if $\zeta \in \mathrm{im}_{\mathcal{F}}(R_i.)$, which means that $\eta \in \mathcal{F}^{p_{i-1}}$ exists such that $\zeta = R_i\,\eta$.

Remark 9 If $\mathcal{F} := D$, then the $\text{ext}^i_D(M, D)$'s inherit a right D-module structure.

The concept of a free resolution of a module can be extended to the concept of a *projective resolution* in which projective modules are used instead of (finitely generated) free left D-modules $D^{1\times p_i}$ [52]. The *length* of a projective resolution is the number of non-zero projective modules defining this resolution. The minimal length of the projective resolutions of a left D-module M is called the *left projective dimension* of M and it is denoted by $\text{lpd}_D(M)$. The *left global dimension* of a ring D is the supremum of $\text{lpd}_D(M)$ for all left D-modules M and it is denoted by $\text{lgld}_D(M)$. For more details, see [52]. Similar definitions can be given for right D-modules. If D is a noetherian ring, i.e., a left and a right noetherian ring, a result due to Kaplansky shows that the left and right global dimensions of D coincide [52] and it is then denoted by $\text{gld}(D)$.

Example 14 ([16]) We have the following examples.

(a) If $\mathbb{A}$ has finite left global dimension and σ is an automorphism of $\mathbb{A}$, then we have $\text{lgld}(\mathbb{A}) \leqslant \text{lgld}(\mathbb{A}[\partial; \sigma, \delta]) \leqslant \text{lgld}(\mathbb{A}) + 1$. Moreover, if $\delta = 0$, then we have $\text{lgld}(\mathbb{A}[\partial; \sigma, \delta]) = \text{lgld}(\mathbb{A}) + 1$.
(b) If $\Bbbk$ is a field, then $\text{gld}(\Bbbk[x_1, \ldots, x_n]) = n$.
(c) If $\Bbbk$ is a field of characteristic 0 (e.g., $\Bbbk := \mathbb{Q}, \mathbb{R}, \mathbb{C}$), then $\text{gld}(A_n(\Bbbk)) = n$ and $\text{gld}(B_n(\Bbbk)) = n$.

Theorem 6 ([9]) *Let D be a noetherian ring with finite global dimension* $\text{gld}(D) := n$, $M := D^{1\times p}/(D^{1\times q}\, R)$ *the left D-module finitely presented by the matrix $R \in D^{q\times p}$, and $N := D^q/(R\, D^p)$ the so-called* Auslander transpose *of M. Then, we have the following results:*

(a) M is a torsion left D-module if and only if $\hom_D(M, D) = 0$.
(b) $t(M) \cong \text{ext}^1_D(N, D)$.
(c) M is a torsion-free left D-module if and only if $\text{ext}^1_D(N, D) = 0$.
(d) M is a reflexive left D-module if and only if $\text{ext}^i_D(N, D) = 0$ *for* $i = 1, 2$.
(e) M is a projective left D-module if and only if $\text{ext}^i_D(N, D) = 0$ *for* $i = 1, \ldots, n$.
(f) If R is a full row rank matrix, i.e., $\ker_D(.R) = 0$, *then M is a projective left D-module if and only if $N \cong \text{ext}^1_D(M, D) = 0$, i.e., if and only if R admits a right inverse.*

Remark 10 If $D := \Bbbk[x_1, \ldots, x_n]$, then Theorem 6 and (a) of Example 14 show that the concepts of torsion-free, reflexive, and projective modules are instances of a sequence of n module properties characterized by the successive vanishing of the $\text{ext}^i_D(N, D)$'s for $i = 1, \ldots, n$. If R has full row rank and $\Bbbk := \mathbb{R}$ or $\mathbb{C}$, it can be proved that $\text{ext}^i_D(N, D) = 0$ for $i = 0, \ldots, r-1$ and $\text{ext}^r_D(N, D) \neq 0$, if and only if the algebraic variety defined by all the $q \times q$ minors of R has a strict complex dimension equal to $n - r$. This result is a generalization of the different concepts of coprimeness developed in the literature of multidimensional systems.

According to Theorem 6, certain module properties are characterized by the vanishing of some of the $\mathrm{ext}^i_D(N, D)$'s. We point out that $N := D^q/(R\,D^p)$ is a right D-module, and thus, it does not define a linear system. To compute the $\mathrm{ext}^i_D(N, D)$'s, we first have to compute a free resolution of right D-modules

$$0 \longleftarrow N \xleftarrow{\kappa} D^{q_0} \xleftarrow{Q_1.} D^{q_1} \xleftarrow{Q_2.} D^{q_2} \xleftarrow{Q_3.} \cdots \tag{1.19}$$

where $Q_1 := R$, $q_0 := q$, and $q_1 := p$, then dualize it to get the following complex of left D-modules:

$$0 \longrightarrow D^{1\times q_0} \xrightarrow{.Q_1} D^{1\times q_1} \xrightarrow{.Q_2} D^{1\times q_2} \xrightarrow{.Q_3} \cdots$$

Then, we have $\mathrm{ext}^i_D(N, D) = \ker_D(.Q_{i+1})/\mathrm{im}_D(.Q_i)$ for $i \geq 0$, where we set $\mathrm{im}_D(.Q_0) = 0$. Since D is a left noetherian ring, the left D-module $\ker_D(.Q_{i+1})$ is finitely generated, and thus there exists $Q'_i \in D^{q'_{i-1}\times q_i}$ such that $\ker_D(.Q_{i+1}) = \mathrm{im}_D(.Q'_i) = D^{1\times q'_{i-1}}\,Q'_i$, which yields $\mathrm{ext}^i_D(N, D) = (D^{1\times q'_{i-1}}\,Q'_i)/(D^{1\times q_{i-1}}\,Q_i)$. If $D := \mathbb{O}$ is an Ore algebra satisfying the hypotheses of Theorem 2, then we can use Algorithm 1 to compute the matrix Q'_i and then Algorithm 2 to check whether or not there exists a matrix $Q''_i \in D^{q'_{i-1}\times q_{i-1}}$ such that $Q'_i = Q''_i\,Q_i$, i.e., to check whether or not $D^{1\times q'_{i-1}}\,Q'_i = D^{1\times q_{i-1}}\,Q_i$, i.e., whether or not $\mathrm{ext}^i_D(N, D)$ is 0 for $i \geq 1$. The only point that does not seem to be constructive is the use of Algorithm 1 to compute the free resolution of N since we have to compute right kernel and not left kernel. Moreover, the computation of Gröbner bases is usually not available in computer algebra systems for right ideals or right modules. To do that, we have to use an involution θ of D (see Definition 3). Indeed, it can be used to turn the right D-module structure into a left D-module structure as explained in the next lemma.

Lemma 1 *Let N be a right D-module and θ an involution of D. Then, we can define the left D-module $\widetilde{N}$ which is equal to N as a set, endowed with the same addition as N, and the left D-action on $\widetilde{N}$ is defined by:*

$$\forall\, d \in D, \quad \forall\, n \in \widetilde{N}, \quad d\,n := n\,\theta(d).$$

Let $M := \mathbb{O}^{1\times p}/(\mathbb{O}^{1\times q}\,R)$ be a left $\mathbb{O}$-module finitely presented by the matrix $R \in \mathbb{O}^{q\times p}$ and let θ be an involution of $\mathbb{O}$. Then, we can define the matrix $\theta(R) := (\theta(R_{ij}))^T \in \mathbb{O}^{p\times q}$, i.e., the transpose of the matrix obtained by applying the involution θ to the matrix R component-wise. Note that we always have $R = \theta^2(R)$, i.e., a matrix S can always be written as $\theta(T)$ for a certain matrix $T := \theta(S)$. We now consider the left $\mathbb{O}$-module finitely presented by $\theta(R)$, namely,

$$\widetilde{N} := \mathbb{O}^{1\times q}/(\mathbb{O}^{1\times p}\,\theta(R)). \tag{1.20}$$

It is called the *adjoint module* of M. Then, one can prove that (b), (c), (d), and (e) of Theorem 6 hold where N is substituted by $\widetilde{N}$. Hence, we can use Algorithm 1 to

compute a free resolution of $\widetilde{N}$

$$0 \longleftarrow \widetilde{N} \xleftarrow{\sigma} D^{1\times q_0} \xleftarrow{.\theta(Q_1)} D^{1\times q_1} \xleftarrow{.\theta(Q_2)} D^{1\times q_2} \xleftarrow{.\theta(Q_3)} \cdots,$$

then dualizing it by applying the involution θ to get the complex of left D-modules:

$$0 \longrightarrow D^{1\times q_0} \xrightarrow{.Q_1} D^{1\times q_1} \xrightarrow{.Q_2} D^{1\times q_2} \xrightarrow{.Q_3} \cdots$$

Then, we have $\mathrm{ext}^i_D(N, D) = (D^{1\times q'_{i-1}} Q'_i)/D^{1\times q_{i-1}} Q_i)$, where $Q'_i \in D^{q'_{i-1}\times q_i}$ is a matrix defined by $\ker_D(.Q_{i+1}) = \mathrm{im}_D(.Q'_i)$. Finally, as above, using Algorithm 2, we can check whether or not $\mathrm{ext}^i_D(N, D) = 0$ for $i \geq 1$. In this way, we can effectively check the conditions of Theorem 6, and thus whether or not M has torsion elements, is torsion-free, reflexive, or projective.

Example 15 Let us consider again Stokes equations defined in Example 4. With the notations of Example 12, using Theorem 6, we can easily prove that the finitely presented $\mathbb{O}$-module $M := \mathbb{O}^{1\times 3}/L$ is torsion. Indeed, since $\det R \neq 0$, R has full row rank, i.e., $\ker_{\mathbb{O}}(R.) = 0$, and we have the following free resolution of N

$$0 \longleftarrow N \xleftarrow{\kappa} \mathbb{O}^3 \xleftarrow{R.} \mathbb{O}^3 \longleftarrow 0,$$

which, by duality, yields the following complex

$$0 \longrightarrow \mathbb{O}^{1\times 3} \xrightarrow{.R} \mathbb{O}^{1\times 3} \longrightarrow 0,$$

and thus we get $t(M) \cong \mathrm{ext}^1_{\mathbb{O}}(N, \mathbb{O}) \cong \ker 0/\mathrm{im}_{\mathbb{O}}(.R) = \mathbb{O}^{1\times 3}/\mathrm{im}_{\mathbb{O}}(.R) = M$. If u (resp., v, p) denotes the residue class of the first (resp., second, third) element of the standard basis of $\mathbb{O}^{1\times 3}$ in M, then eliminating v and p (resp., u and p, resp., u and v) from (1.12) as shown in Example 12, we obtain:

$$\begin{cases} \Delta\,(\nu\,\Delta - c)\,u = 0, \\ \Delta\,(\nu\,\Delta - c)\,v = 0, \\ \Delta\,p = 0. \end{cases}$$

Hence, each generator u, v, p of M satisfies a PDE, i.e., is a torsion element.

Example 16 Let us illustrate Theorem 6 on a simple linear DTD system defined by:

$$\begin{cases} \dot{x}_1(t) = x_1(t) + x_2(t-1) + u(t), \\ \dot{x}_2(t) = x_1(t-1) + x_2(t) + u(t). \end{cases} \tag{1.21}$$

Let $\mathbb{O} := \mathbb{Q}\left[\partial; \mathrm{id}_{\mathbb{Q}}, \frac{d}{dt}\right][\delta; \sigma, 0]$ be the commutative Ore algebra of DTD operators, where σ is defined by $\sigma(a(t)) = a(t-1)$ and $M := \mathbb{O}^{1\times 3}/(\mathbb{O}^{1\times 2}\,R)$ the $\mathbb{O}$-module

finitely presented by the following matrix:

$$R := \begin{pmatrix} \partial - 1 & -\delta & -1 \\ -\delta & \partial - 1 & -1 \end{pmatrix} \in \mathbb{D}^{2\times 3}.$$

Let us introduce the Auslander transpose $N := \mathbb{D}^2/(R\,\mathbb{D}^3)$ of M. We note that we have $N \cong \widetilde{N} := \mathbb{D}^{1\times 2}/(\mathbb{D}^{1\times 3}\,R^T)$ because $\mathbb{D}$ is a commutative ring and $\theta = \mathrm{id}_{\mathbb{D}}$. Let us explicitly compute the $\mathrm{ext}^i_{\mathbb{D}}(\widetilde{N}, \mathbb{D})$'s. Since $\mathrm{gld}(\mathbb{D}) = 2$ (see (b) of Example 14), one can prove that $\mathrm{ext}^i_{\mathbb{D}}(\widetilde{N}, \mathbb{D}) = 0$ for $i \geq 3$, a fact that we will check again. Using Algorithm 1, we can check that $\widetilde{N}$ admits the free resolution

$$0 \longleftarrow \widetilde{N} \xleftarrow{\sigma} \mathbb{D}^2 \xleftarrow{.R^T} \mathbb{D}^3 \xleftarrow{.Q_2^T} \mathbb{D} \longleftarrow 0,$$

where $Q_2^T := (1 \quad 1 \quad \partial - 1 - \delta)$. Dualizing this exact sequence, we get the complex:

$$0 \longrightarrow \mathbb{D}^{1\times 2} \xrightarrow{.R} \mathbb{D}^{1\times 3} \xrightarrow{.Q_2} \mathbb{D} \longrightarrow 0.$$

Then, we have

$$\begin{cases} \hom_{\mathbb{D}}(\widetilde{N}, \mathbb{D}) = \ker_{\mathbb{D}}(.R) = 0, \\ \mathrm{ext}^1_{\mathbb{D}}(\widetilde{N}, \mathbb{D}) = \ker_{\mathbb{D}}(.Q_2)/\mathrm{im}_{\mathbb{D}}(.R), \\ \mathrm{ext}^2_{\mathbb{D}}(\widetilde{N}, \mathbb{D}) = \mathbb{D}/(\mathbb{D}^{1\times 3}\,Q_2) = 0, \\ \mathrm{ext}^i_{\mathbb{D}}(\widetilde{N}, \mathbb{D}) = 0, \ i \geq 3, \end{cases}$$

since R has full row rank and $1 \in \mathbb{D}^{1\times 3}\,Q_2$. Using Algorithm 1 again, we get $\ker_{\mathbb{D}}(.Q_2) = \mathbb{D}^{1\times 2}\,R'$, where:

$$R' := \begin{pmatrix} 1 & -1 & 0 \\ 0 & \partial - 1 - \delta & -1 \end{pmatrix}.$$

By (b) of Theorem 6, we get $t(M) \cong (\mathbb{D}^{1\times 2}\,R')/(\mathbb{D}^{1\times 2}\,R)$. It means that the rows of R' modulo the system equations define a generating set of the torsion $\mathbb{D}$-submodule $t(M)$ of M. The first (resp., second) row of R' yields the torsion element $z_1 := x_1 - x_2$ (resp., $z_2 := (\partial - 1 - \delta)\,x_2 - u = \delta\,z_1$). Hence, $t(M)$ is generated by z_1. If we consider the following inhomogeneous linear system

$$\begin{cases} x_1 - x_2 = z_1, \\ (\partial - 1)\,x_1 - \delta\,x_2 - u = 0, \\ (\partial - 1)\,x_2 - \delta\,x_1 - u = 0, \end{cases}$$

then computing a Gröbner basis for a monomial order which eliminates x_1, x_2, and u, we obtain $(\partial + \delta - 1)\,z_1 = 0$. Let us now study the torsion-free $\mathbb{D}$-module

$M/t(M) := \mathbb{O}^{1\times 3}/(\mathbb{O}^{1\times 2}\, R')$. One can show that $M/t(M) \cong \mathbb{O}^{1\times 3}\, Q_2 = \mathbb{O}$ since $\text{ext}^2_{\mathbb{O}}(\widetilde{N}, \mathbb{O}) = 0$ (see, e.g., [9, 47, 51]). By (e) of Theorem 6, the $\mathbb{O}$-module $M/t(M)$ is projective. Using Algorithm 3, we can check that the following matrix

$$L := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

is a left inverse of R'^T, and thus $S := L^T$ is a right inverse of R', which shows again that $M/t(M)$ is a projective $\mathbb{O}$-module by (f) of Theorem 6. By the Quillen–Suslin theorem (see (b) of Theorem 4), $M/t(M)$ is then a free $\mathbb{O}$-module of rank 1. This result can be easily checked again by noticing that $M/t(M)$ is defined by

$$\begin{cases} y_1 - y_2 = 0, \\ (\partial - 1 - \delta)\, y_2 - v = 0, \end{cases} \Longleftrightarrow \quad \begin{cases} y_1 = y_2, \\ v = (\partial - 1 - \delta)\, y_2, \end{cases}$$

which shows that y_2 is a basis of $M/t(M)$. Finally, since $\mathbb{O}$ is a commutative polynomial ring, we can use Remark 10 to prove again the results obtained above. Indeed, the ideal $\text{Fitt}_0(N)$ defined by all the 2×2 minors of R is defined by the ideal $I := (\partial + \delta - 1)$. The algebraic variety formed by the zeros of I is $\partial + \delta - 1 = 0$, which is 1-dimensional. Using Remark 10, we then get $\text{ext}^1_{\mathbb{O}}(N, \mathbb{O}) \neq 0$, which proves again that $t(M) \neq 0$. Similarly, if $N' := \mathbb{O}^{2\times 1}/(R'\, \mathbb{O}^{3\times 1})$ is the Auslander transpose of $M/t(M)$, then we have $\text{Fitt}_0(N') = (\partial - 1 - \delta, 1) = \mathbb{O}$, which shows that $\text{ext}^i_{\mathbb{O}}(N', \mathbb{O}) = 0$ for $i = 1,\ 2$, which proves again that $M/t(M)$ is a projective and thus a free $\mathbb{O}$-module.

Finally, if R is a full row rank matrix, then (f) of Theorem 6 shows that M is a projective left D-module if and only if we have $N \cong \text{ext}^1_D(M, D) = 0$ or equivalently if and only if $\widetilde{N} = 0$. In this case, we do not have to test the vanishing of all the $\text{ext}^i_D(N, D)$'s for $i = 1, \ldots, n$ as shown in (e) of Theorem 6.

Example 17 Let $\mathbb{O} := \mathbb{A}\left[\partial; \text{id}_A, \frac{d}{dt}\right]$ be a ring of OD operators with coefficients in a noetherian differential ring $\mathbb{A}$, $A \in \mathbb{A}^{n\times n}$, $B \in \mathbb{A}^{n\times m}$, and the left $\mathbb{O}$-module $M := \mathbb{O}^{1\times(n+m)}/(\mathbb{O}^{1\times n}\, R)$ finitely presented by $R := (\partial\, I_n - A \quad -B) \in \mathbb{O}^{n\times(n+m)}$ which defines the linear system $\dot{x}(t) = A\, x(t) + B\, u(t)$. Since R has full row rank, (f) of Theorem 6 shows that M is a projective left $\mathbb{O}$-module if and only if R admits a right inverse, i.e., using the involution θ defined in (b) of Example 3, if and only if $\theta(R) := (-\partial\, I_n - A^T \quad -B^T)^T \in \mathbb{O}^{(n+m)\times n}$ admits a left inverse S. This is equivalent to say that the *adjoint system* $\theta(R)\, \lambda = 0$, i.e.,

$$\begin{cases} \dot{\lambda} + A^T\, \lambda = 0, \\ B^T\, \lambda = 0, \end{cases} \tag{1.22}$$

has only the trivial solution $\lambda = 0$ since $S\, \theta(R) = I_q$ yields $\lambda = S\, (\theta(R)\, \lambda) = 0$. The above system is not a Gröbner basis for the total degree order since if we

differentiate the zero-order equation, we get $\dot{B}^T \lambda + B^T \dot{\lambda} = 0$, i.e., using the first-order equation, we obtain the new zero-order equation $(\dot{B}^T - B^T A^T) \lambda = 0$. We can repeat the same procedure with this last equation. Hence, if we define the sequence of matrices B_i defined by $B_0 := B^T$ and $B_{i+1} := \dot{B}_i - B_i A^T$ for $i \geq 1$, we obtain that $(B_0^T \quad B_1^T \quad B_2^T \quad \ldots)^T \lambda = 0$. Since $\mathbb{A}$ is supposed to be noetherian, the increasing sequence of A-submodules $\mathcal{O}_k := \sum_{i=0}^{k} \mathbb{A}^{1\times m} B_i$ of $\mathbb{A}^{1\times n}$ stabilizes (see, e.g., [52]), i.e., there exists $r \in \mathbb{N}$ such that $\mathcal{O}_s = \mathcal{O}_r$ for all $s \geq r$. Then, we get:

$$(22) \quad \Longleftrightarrow \quad \left\{ \begin{array}{l} \dot{\lambda} + A^T \lambda = 0, \\ \begin{pmatrix} B_0 \\ \vdots \\ B_r \end{pmatrix} \lambda = 0. \end{array} \right.$$

Hence, the above system has the only solution $\lambda = 0$ if and only if the matrix $\mathcal{O} := (B_0^T \quad B_1^T \quad B_2^T \quad \ldots \quad B_r^T)^T$ admits a left inverse with entries in $\mathbb{A}$ or equivalently if and only if $\mathcal{O}^T$ admits a right inverse with entries in $\mathbb{A}$. If $\mathbb{A}$ is a field, then we can take $r = n - 1$ by the Cayley–Hamilton theorem and the condition on the existence of a right inverse for $\mathcal{O}^T$ then becomes that $\mathcal{O}^T$ has full row rank. Finally, if $\mathbb{A} := \Bbbk$ is a field of constants, i.e., $\dot{a} = 0$ for all $a \in A$, as, e.g., $\mathbb{A} = \mathbb{Q}$ or $\mathbb{R}$, then we get the standard controllability condition $\text{rank}_{\Bbbk}(B \quad A\,B \quad \ldots \quad A^{n-1}\,B) = n$ (see [27, 32]).

1.4.3 Dictionary Between System Properties and Module Properties

Let us introduce a few more definitions.

Definition 13 ([52]) We have the following definitions:

(a) A left D-module $\mathcal{F}$ is said to be *injective* if for every left D-module M, we have $\text{ext}_D^i(M, \mathcal{F}) = 0$ for $i \geq 1$.
(b) A left D-module $\mathcal{F}$ is said to be *cogenerator* if for every left D-module M and $m \in M \setminus \{0\}$, there exists $\varphi \in \text{hom}_D(M, \mathcal{F})$ such that $\varphi(m) \neq 0$.

It can be shown that a left D-module $\mathcal{F}$ is injective if and only if for every matrix $R \in D^{q\times p}$ and $\zeta \in \mathcal{F}^q$ satisfying $R_2\,\zeta = 0$, where $R_2 \in D^{r\times q}$ is any matrix such that $\ker_D(.R) = \text{im}_D(.R_2)$, there exists $\eta \in \mathcal{F}^p$ solving the inhomogeneous linear system $R\,\eta = \zeta$ [52]. A standard result in homological algebra shows that there always exists an injective cogenerator left module for a ring D [52].

Example 18 If Ω is an open convex subset of $\mathbb{R}^n$ and $\mathcal{F} := C^\infty(\Omega)$ or $\mathcal{D}'(\Omega)$ (i.e., the space of distributions with support in Ω), then $\mathcal{F}$ is an injective cogenerator $D := \Bbbk\left[\partial_1; \text{id}_{\Bbbk}, \frac{\partial}{\partial x_1}\right] \cdots \left[\partial_n; \text{id}_{\Bbbk}, \frac{\partial}{\partial x_n}\right]$-module, where $\Bbbk := \mathbb{R}$ or $\mathbb{C}$. For more details, see [43] and the references therein.

Example 19 If $\mathcal{F}$ is the set of real-valued functions on $\mathbb{R}$ which are smooth except for a finite number of points, then $\mathcal{F}$ is an injective cogenerator left $B_1(\mathbb{R})$-module [57].

Based on the results of [21–23, 45, 46, 56], we can give the following definitions.

Definition 14 ([9]) Let D be a noetherian domain, $R \in D^{q\times p}$, $\mathcal{F}$ an injective cogenerator left D-module, and the linear system (behaviour) defined by R and $\mathcal{F}$:

$$\ker_{\mathcal{F}}(R.) := \{\eta := (\eta_1 \ \ldots \ \eta_p)^T \in \mathcal{F}^p \mid R\,\eta = 0\}.$$

(a) An *observable* is a D-linear combination of the system variables η_i.
(b) An observable $\psi(\eta)$ is called *autonomous* if it satisfies a D-linear relation by itself, i.e., $d\,\psi(\eta) = 0$ for some $d \in D \setminus \{0\}$. An observable is said to be *free* if it is not autonomous.
(c) The linear system is said to be *controllable* if every observable is free.
(d) The linear system is said to be *parametrizable* if there exists a matrix $Q \in D^{p\times m}$ such that $\ker_{\mathcal{F}}(R.) = Q\,\mathcal{F}^m$, i.e., if for every $\eta \in \ker_{\mathcal{F}}(R.)$, there exists $\xi \in \mathcal{F}^m$ such that $\eta = Q\,\xi$. Then, Q is called a *parametrization* and ξ a *potential*.
(e) The linear system is said to be *flat* if there exists a parametrization $Q \in D^{p\times m}$ which admits a left inverse $T \in D^{m\times p}$, i.e, $T\,Q = I_p$. In other words, a flat system is a parametrizable system such that every component ξ_i of a potential ξ is an observable of the system. The potential ξ is then called a *flat output*.

We are now in position to state the correspondence between the properties of a linear system defined in Definition 14 and the properties of the associated finitely generated left module defined in Definition 11.

Theorem 7 ([9]) *With the hypotheses and notations of Definition 14, we have:*

(a) The observables of the linear system are in one-to-one correspondence with the elements of M.
(b) The autonomous observables of the linear system are in one-to-one correspondence with the torsion elements of M. Consequently, the linear system is controllable iff M is torsion-free.
(c) The linear system is parametrizable iff there exists a matrix $Q \in D^{p\times m}$ such that we have $M := D^{1\times p}/(D^{1\times q}\,R) \cong D^{1\times p}\,Q$, i.e., iff M is a torsion-free left D-module. Then, the matrix Q is a parametrization, i.e., $\ker_{\mathcal{F}}(R.) = Q\,\mathcal{F}^m$.
(d) The linear system is flat iff M is a free left D-module. Then, the bases of M are in one-to-one correspondence with the flat outputs of the linear system.

Example 20 Let us give the system interpretations of the results obtained in Example 16. First, we have the autonomous element $z_1(t) := x_1(t) - x_2(t)$ of (1.21) since it satisfies the autonomous DTD equation $\dot{z}_1(t) - z_1(t) - z_1(t-1) = 0$. It is a non controllable element of (1.21) since its trajectory cannot be changed by means of u. Moreover, the controllable system associated with (1.21) is defined by $M/t(M) := \mathbb{O}^{1\times 3}/(\mathbb{O}^{1\times 2}\,R')$, which is a free $\mathbb{O}$-module of rank 1. Thus, if

$\mathcal{F}$ is any $\mathbb{O}$-module (e.g., $\mathcal{F} := C^\infty(\mathbb{R}_{\geq 0})$), then the corresponding linear system $R'\,\eta = 0$, where $\eta := (y_1 \quad y_2 \quad v)^T$, is flat and y_2 is a flat output. Finally, the matrix Q_2 defined in Example 16 is an injective parametrization of $\ker_{\mathcal{F}}(R'.)$, i.e., we have $R'\,\eta = 0$ if and only if there exists $\xi \in \mathcal{F}$ such that $\eta = Q_2\,\xi$. Finally, we can check that Q_2 admits a left inverse $S_2 := (0 \quad 1 \quad 0)$ (see Algorithm 3), which shows that $\xi = (S_2\ Q_2)\,\xi = S_2\,\eta$ is uniquely defined by η.

Finally, we illustrate Theorems 6 and 7 with standard linear functional systems coming from control theory and mathematical physics.

Example 21 (a) Let us consider a wind tunnel model studied in [41] and defined by the following linear DTD system:

$$\begin{cases} \dot{x}_1(t) + a\,x_1(t) - k\,a\,x_2(t-h) = 0, \\ \dot{x}_2(t) - x_3(t) = 0, \\ \dot{x}_3(t) + \omega^2\,x_2(t) + 2\,\zeta\,\omega\,x_3(t) - \omega^2\,u(t) = 0, \end{cases} \tag{1.23}$$

where a, k, ω, and ζ are constant parameters. Checking that $\mathrm{ext}^1_{\mathbb{O}}(N, \mathbb{O}) = 0$, (c) of Theorem 6 shows that (1.23) defines a torsion-free module over the ring of DTD operators, and thus (1.23) is parametrizable by (c) of Theorem 7. The matrix Q_1 obtained during the computation of $\mathrm{ext}^1_{\mathbb{O}}(N, \mathbb{O})$ (see (1.19)) is then a parametrization of (1.23) and we have

$$(23) \iff \begin{cases} x_1(t) = \omega^2\,k\,a\,z(t-h), \\ x_2(t) = \omega^2\,\dot{z}(t) - a\,\omega^2\,z(t), \\ x_3(t) = \omega^2\,\ddot{z}(t) + \omega^2\,a\,\dot{z}(t), \\ u(t) = z^{(3)}(t) + (2\,\zeta\,\omega + a)\,\ddot{z}(t) + (\omega^2 + 2\,a\,\omega\,\zeta)\,\dot{z}(t) + a\,\omega\,z(t), \end{cases}$$

for all z which belongs to an injective module $\mathcal{F}$ over the ring of DTD operators.

(b) Let us consider the *first group of Maxwell equations* defined by

$$\begin{cases} \dfrac{\partial \mathbf{B}}{\partial t} + \boldsymbol{\nabla} \wedge \mathbf{E} = \mathbf{0}, \\ \boldsymbol{\nabla}.\,\mathbf{B} = 0, \end{cases} \tag{1.24}$$

where $\mathbf{B}$ (resp., $\mathbf{E}$) denotes the *magnetic* (resp., *electric*) *field*. We can prove that the differential module associated with (1.24) is reflexive (see [9]). In particular, (1.24) is parametrizable and using the matrix Q_1 obtained in the computation of $\mathrm{ext}^1_{\mathbb{O}}(N, \mathbb{O})$ (see (1.19)), we obtain

$$(24) \iff \begin{cases} \mathbf{E} = -\dfrac{\partial \mathbf{A}}{\partial t} - \boldsymbol{\nabla}\,V, \\ \mathbf{B} = \boldsymbol{\nabla} \wedge \mathbf{A}, \end{cases} \tag{1.25}$$

where $(\mathbf{A},\ V)$ is the so-called *quadri-potential* formed by smooth functions over $\mathbb{R}^3$. The second matrix Q_2 defining a free resolution of N (see (1.19)) then defines a parametrization of the inhomogeneous part of (1.25), i.e., we have

$$\left\{\begin{array}{l} -\dfrac{\partial \mathbf{A}}{\partial t} - \boldsymbol{\nabla} V = \mathbf{0}, \\ \boldsymbol{\nabla} \wedge \mathbf{A} = \mathbf{0}, \end{array}\right. \Longleftrightarrow \left\{\begin{array}{l} \mathbf{A} = \boldsymbol{\nabla} \xi, \\ V = -\dfrac{\partial \xi}{\partial t}, \end{array}\right.$$

where ξ is an arbitrary smooth function on $\mathbb{R}^3$ (used, e.g., for the *Lorenz gauge*).

(c) Similarly as for the first group of Maxwell equations, we can prove that *the equilibrium of the stress tensor* defined by

$$\left\{\begin{array}{l} \dfrac{\partial \sigma_x}{\partial x} + \dfrac{\partial \tau_{xy}}{\partial y} + \dfrac{\partial \tau_{zx}}{\partial z} = 0, \\ \dfrac{\partial \tau_{xy}}{\partial x} + \dfrac{\partial \sigma_y}{\partial y} + \dfrac{\partial \tau_{yz}}{\partial z} = 0, \\ \dfrac{\partial \tau_{zx}}{\partial x} + \dfrac{\partial \tau_{yz}}{\partial y} + \dfrac{\partial \sigma_z}{\partial z} = 0, \end{array}\right. \tag{1.26}$$

defines a reflexive differential module (see [47]) and we have the parametrization

$$(26) \Longleftrightarrow \left\{\begin{array}{l} \sigma_x = \dfrac{\partial^2 \chi_3}{\partial y^2} + \dfrac{\partial^2 \chi_2}{\partial z^2} + \dfrac{\partial^2 \psi_1}{\partial y\, \partial z}, \\ \tau_{yz} = -\dfrac{\partial^2 \chi_1}{\partial y\, \partial z} - \dfrac{1}{2}\dfrac{\partial}{\partial x}\left(-\dfrac{\partial \psi_1}{\partial x} + \dfrac{\partial \psi_2}{\partial y} + \dfrac{\partial \psi_3}{\partial z}\right), \\ \sigma_y = \dfrac{\partial^2 \chi_1}{\partial z^2} + \dfrac{\partial^2 \chi_3}{\partial x^2} + \dfrac{\partial^2 \psi_2}{\partial z\, \partial x}, \\ \tau_{zx} = -\dfrac{\partial^2 \chi_2}{\partial z\, \partial x} - \dfrac{1}{2}\dfrac{\partial}{\partial y}\left(\dfrac{\partial \psi_1}{\partial x} - \dfrac{\partial \psi_2}{\partial y} + \dfrac{\partial \psi_3}{\partial z}\right), \\ \sigma_z = \dfrac{\partial^2 \chi_2}{\partial x^2} + \dfrac{\partial^2 \chi_1}{\partial y^2} + \dfrac{\partial^2 \psi_3}{\partial x\, \partial y}, \\ \tau_{xy} = -\dfrac{\partial^2 \chi_3}{\partial x\, \partial y} - \dfrac{1}{2}\dfrac{\partial}{\partial z}\left(\dfrac{\partial \psi_1}{\partial x} + \dfrac{\partial \psi_2}{\partial y} - \dfrac{\partial \psi_3}{\partial z}\right), \end{array}\right.$$

where the ψ_i's and the χ_j's are smooth functions on $\mathbb{R}^3$. Finally, if we set $\psi_1 = \psi_2 = \psi_3 = 0$ (resp., $\chi_1 = \chi_2 = \chi_3 = 0$), then we obtain the so-called *Maxwell's parametrization* (resp., *Morera's parametrization*). For more details, see [47].

(d) Let us consider the following time-varying linear OD system:

$$\left\{\begin{array}{l} \dot{x}_1(t) - t\, u_1(t) = 0, \\ \dot{x}_2(t) - u_2(t) = 0. \end{array}\right.$$

Using Example 17, we can easily check that this system is controllable, i.e., defines a stably free left module over $\mathbb{O} := A_1(\mathbb{Q})$. By (c) of Theorem 4 (i.e., by Stafford's theorem), this module is then free, i.e., the time-varying linear system is flat by (d) of Theorem 7. The effective computation of an injective parametrization is usually a difficult task. To do that, following constructive versions of Stafford's theorems [48, 49] and their implementations in the STAFFORD package[48, 49], we obtain the following injective parametrization

$$\begin{cases} x_1(t) = t^2\,\xi_1(t) - t\,\dot{\xi}_2(t) + \xi_2(t), \\ x_2(t) = t\,(t+1)\,\xi_1(t) - (t+1)\,\dot{\xi}_2(t) + \xi_2(t), \\ u_1(t) = t\,\dot{\xi}_1(t) + 2\,\xi_1(t) - \ddot{\xi}_2(t), \\ u_2(t) = t\,(t+1)\,\dot{\xi}_1(t) + (2\,t+1)\,\xi_1(t) - (t+1)\,\ddot{\xi}_2(t), \end{cases}$$

where ξ_1 and ξ_2 are arbitrary functions in a left $\mathbb{O}$-module $\mathcal{F}$, and:

$$\begin{cases} \xi_1(t) = (t+1)\,u_1(t) - u_2(t), \\ \xi_2(t) = (t+1)\,x_1(t) - t\,x_2(t). \end{cases}$$

In the language of module theory, we obtain that $\{\xi_1 = (t+1)\,u_1 - u_2,\ \xi_2 = (t+1)\,x_1 - t\,x_2\}$ is a basis of the free left $A_1(\mathbb{Q})$-module M, where x_1, x_2, u_1 and u_2 denote here the generators of the module as explained in Sect. 1.4.1. Finally, we point out that the above injective parametrization does not contain singularities contrary to

$$\begin{cases} u_1(t) = t^{-1}\,\dot{x}_1(t), \\ u_2(t) = \dot{x}_2(t), \end{cases}$$

where x_1 and x_2 are arbitrary functions, which admits a singularity at $t = 0$.

(e) If we consider the following linear DTD system

$$\begin{cases} \dot{y}_1(t) - y_1(t-h) + 2\,y_1(t) + 2\,y_2(t) - 2\,u(t-h) = 0, \\ \dot{y}_1(t) + \dot{y}_2(t) - \dot{u}(t-h) - u(t) = 0, \end{cases}$$

using (e) or (f) of Theorem 6, then we can check that it defines a projective module over the commutative polynomial ring of DTD operators with constant coefficients. By the Quillen–Suslin theorem (see (b) of Theorem 4), this module is free. The computation of bases and injective parametrizations is usually difficult and requires an effective version of the Quillen–Suslin theorem [20]. Using the QUILLENSUSLIN package [20], we get the following injective parametrization

$$\begin{cases} y_1(t) = \xi(t), \\ y_2(t) = \frac{1}{2}\,(-\ddot{\xi}(t-h) + \dot{\xi}(t-2\,h) - \dot{\xi}(t) + \xi_1(t-h) - 2\,\xi(t)), \\ u(t) = \frac{1}{2}\,(\dot{\xi}(t-h) - \ddot{\xi}(t)), \end{cases}$$

for all ξ belonging to a module over the ring of DTD operators. Finally, note that y_1 defines a basis of the free module defined by the above system.

1.5 Mathematica Packages

1.5.1 The HolonomicFunctions *Package*

The Mathematica package named HolonomicFunctions has been developed by the second-named author in the frame of his Ph.D. thesis [29]. It can be downloaded for free from the website http://www.risc.jku.at/research/combinat/software/HolonomicFunctions/, and a complete documentation is given in the manual [30]. We start with a fresh Mathematica session and load the package with the following command:

```
In[1]:= << RISC`HolonomicFunctions`
```

```
HolonomicFunctions Package version 1.7.1 (09-Oct-2013)
written by Christoph Koutschan
Copyright 2007–2013, Research Institute for Symbolic Computation (RISC),
Johannes Kepler University, Linz, Austria
⟶ Type? HolonomicFunctions for help.
```

In its core, the package provides functionality to construct Ore algebras and to work with Ore polynomials. First, we demonstrate how this is done using a very standard application, namely the operator from (b) of Example 2. For this purpose, we define a multivariate Ore algebra with rational function coefficients, which is built up of the shift operator S_n and the ordinary differential operator D_t:

```
In[2]:= alg = OreAlgebra[S[n], Der[t]]
Out[2]= ℚ(t, n)[S_n; S_n, 0][D_t; 1, D_t]
```

The symbol $\mathbb{K}$ in Out[2] has no particular meaning, and just indicates that the constant field can be everything that covers the user's input; for example $\mathbb{K}$ could contain the rational numbers $\mathbb{Q}$ as a proper subfield.

We can now convert an input expression to an Ore polynomial that belongs to this Ore algebra and do some arithmetic (note the usage of the noncommutative multiplication ** in `Mathematica`):

```
In[3]:= op = ToOrePolynomial[S[n] + Der[t] - n/t, alg]
```

$$\text{Out[3]=}\ S_n + D_t - \frac{n}{t}$$

```
In[4]:= op ** (Der[t] + t n)
```

$$\text{Out[4]=}\ S_n D_t + D_t^2 + (n\,t + t)\,S_n + \left(n\,t - \frac{n}{t}\right) D_t + \left(n - n^2\right)$$

To construct an Ore algebra with polynomial coefficients, we just have to include the variables n and t in the command. Note that each monomial is displayed according to the order in which the generators of the ring are given:

```
In[5]:= alg1 = OreAlgebra[S[n], Der[t], n, t]
Out[5]= ℚ[n, t][Sn; Sn, 0][Dt; 1, Dt]
In[6]:= ChangeOreAlgebra[t ** op, alg1]
Out[6]= Sn t + Dt t − n − 1
```

The HOLONOMICFUNCTIONS package provides a rather general implementation of Ore algebras, which is advantageous for the applications in control theory discussed in Sect. 1.4. For instance, the coefficients of an Ore polynomial ring need not necessarily be polynomials or rational functions. The software also allows us to have, for example, elementary functions in the coefficients:

```
In[7]:= alg = OreAlgebra[Der[t]]
Out[7]= ℚ(t)[Dt; 1, Dt]
In[8]:= op = ToOrePolynomial[Cos[t] ** Der[t] ** Sin[t], alg]
Out[8]= sin(t) cos(t) Dt + cos²(t)
In[9]:= op + Sin[t]^2
Out[9]= sin(t) cos(t) Dt + (sin²(t) + cos²(t))
```

Note that the obvious simplification in the last step is not carried out. By default, HOLONOMICFUNCTIONS keeps the coefficients of Ore polynomials in expanded form, without further simplifications. But there are options to specify a normal form for the coefficients and how to add and multiply them:

```
In[10]:= alg1 = OreAlgebra[Der[t], CoefficientNormal → Simplify, CoefficientPlus
          → (Simplify[#1 + #2]&), CoefficientTimes → (Simplify[#1 * #2]&)]
Out[10]= ℚ(t)[Dt; 1, Dt]
In[11]:= op1 = ChangeOreAlgebra[op, alg1]
Out[11]= sin(t) cos(t) Dt + cos²(t)
In[12]:= op1 + Sin[t]^2
Out[12]= sin(t) cos(t) Dt + 1
```

Ideally, these options are chosen in a way that expressions identically zero are actually simplified to 0. This is, for instance, not the case when dealing with rational function coefficients in expanded form (as we did above).

Apart from the coefficient domain, HOLONOMICFUNCTIONS provides also a lot of flexibility concerning Ore extensions. As we have seen already, the most common operator symbols are predefined, but there is also a way for the user to define own operator symbols. As an example, we can construct an Ore algebra with a generic Ore extension:

```
In[13]:= OreSigma[d] := σ;
In[14]:= OreDelta[d] := δ;
In[15]:= alg = OreAlgebra[d]
```

Out[15]= $\mathbb{K}[d; \sigma, \delta]$

In[16]:= **ToOrePolynomial[d^2 ** t, alg]**

Out[16]= $\sigma(\sigma(t))d^2 + (\delta(\sigma(t)) + \sigma(\delta(t))d + \delta(\delta(t))$

Based on the arithmetic of Ore polynomials, an implementation of Buchberger's algorithm for computing Gröbner bases is part of the HOLONOMICFUNCTIONS package. In the following, we consider a family of orthogonal polynomials, namely the Legendre polynomials, which satisfy a second-order differential equation as well as a three-term recurrence. We represent these equations as operators in a suitable Ore algebra and show, by means of a Gröbner basis computation, that Buchberger's product criterion cannot be exploited in noncommutative domains (note that the two Ore polynomials have leading power products D_t^2 and S_n^2, whose gcd is 1):

In[17]:= **ode = (t^2 − 1) * D[f[n, t], t, t] + 2x * D[f[n, t], t] − n(n + 1) * f[n, t]**

Out[17]= $\left(t^2 - 1\right)\, f^{(0,2)}(n,t) + 2\,t\, f^{(0,1)}(n,t) - n\,(n+1) f(n,t)$

In[18]:= **rec = (n + 1) * f[n + 1, t] − t(2n + 1) * f[n, t] + n * f[n − 1, t]**

Out[18]= $n\, f(n-1,t) - (2\,n+1)\,t\, f(n,t) + (n+1)\, f(n+1,t)$

In[19]:= **ops = ToOrePolynomial[{ode, rec}, f[n, x]]**

Out[19]= $\{(t^2-1)\, D_t^2 + 2\,t\, D_t + (-n^2 - n),\ (n+2)\, S_n^2 + (-2\,n\,t - 3\,t)\, S_n + (n+1)\}$

In[20]:= **OreGroebnerBasis[ops]**

Out[20]= $\{(-n-1)\, S_n + (t^2-1)\, D_t + (nt + t),\ (t^2-1)\, D_t^2 + 2\,t\, D_t + (-n^2 - n)\}$

Although this paper is mostly about applications of the above-described methods in control theory, we want to mention briefly the main application for which the HOLONOMICFUNCTIONS package has been developed. That is: proving special function identities, involving integrals and symbolic sums, in the spirit of Zeilberger's holonomic systems approach [54]. Once the input functions are represented by their annihilators (together with initial conditions), one can use Gröbner basis techniques to compute the annihilator of an integral or sum, by employing the method of creative telescoping [55]. An identity then is established, for example, by observing that both sides satisfy the same differential equation or recurrence. As an example, consider the following identity involving the Laguerre polynomials $L_n^a(t)$ and the Bessel function $J_a(t)$:

$$e^{-t}\, t^{a/2}\, n!\, L_n^a(t) = \int_0^{+\infty} e^{-\tau}\, \tau^{\frac{a}{2}+n}\, J_a\left(2\sqrt{\tau\, t}\right)\, \mathrm{d}\tau. \tag{1.27}$$

By using closure properties of holonomic functions, the HOLONOMICFUNCTIONS package automatically computes the annihilator of the function on the left-hand side of (1.27). The result is given as a Gröbner basis:

In[21]:= **Annihilator[Exp[−t] * t^(a/2) * n! * LaguerreL[n, a, t], {S[a], S[n], Der[t]}]**

Out[21]= $\{2\, S_n - 2\,t\, D_t + (-a - 2\,n - 2),\ 4\,t^2\, D_t^2 + (4\,t^2 + 4\,t)\, D_t + (-a^2 + 2\,a\,t + 4\,n\,t + 4\,t),$
$2\,t\, S_a^2 + (2\,a\,t + 2\,t^2 + 2\,t)\, D_t + (-a^2 + a\,t - a + 2\,n\,t + 2\,t)\}$

For the right-hand side of (1.27), one computes the annihilator of the integrand, and then applies creative telescoping to it, in the form of Chyzak's algorithm [8]:

```
In[22]:= ann = Annihilator[Exp[-τ] * τ^(a/2 + n) * BesselJ[a, 2 Sqrt[τ t]],
           {S[a], S[n], Der[t], Der[τ]}]
Out[22]= {2 t D_t - 2 τ D_τ + (a + 2 n - 2 τ), S_n - τ, τ^2 D_τ^2 + (-a τ - 2 n τ + 2 τ^2 + τ) D_τ +
          (a n - a τ + n^2 - 2 n τ + τ^2 + τ t + τ), t S_a^2 + (a τ + τ) D_τ + (-a^2 - a n + a τ -
          a - n + τ t + τ)}
In[23]:= CreativeTelescoping[ann, Der[τ]]
Out[23]= {{-2 S_n + 2 t D_t + (a + 2 n + 2), 4 t^2 D_t^2 + (4 t^2 + 4 t) D_t + (-a^2 + 2 a t + 4 n t + 4 t),
          2 t S_a^2 + (2 a t + 2 t^2 + 2 t) D_t + (-a^2 + a t - a + 2 n t + 2 t)}, {-2 τ, -4 τ t, -2 τ t}}
```

Note that the first part of Out[23] agrees (up to sign) with Out[21], the annihilator of the left-hand side. In order to complete the proof of (1.27), one has to investigate whether the *certificate* (the second part of Out[23]) contributes an inhomogeneous part to the computed equations (this is not the case here), and one has to compare initial values. These steps are currently beyond the capabilities of the package and have to be done by hand; see the examples in [29] where this is demonstrated in detail.

1.5.2 The OREALGEBRAICANALYSIS *Package*

A `Mathematica` package, called OREALGEBRAICANALYSIS, has been recently developed by the first, third, and fourth-named authors.[1] It is freely available with a library of examples (see [15]).

The OREALGEBRAICANALYSIS package can be used to study (determined/overdetermined/underdetermined) linear functional systems appearing, e.g., in control theory and in mathematical physics. For instance, structural properties of linear functional systems can algorithmically be decided (e.g., existence and computation of autonomous elements, (injective, minimal, chain of) parametrizations, potentials, flat outputs, decide Willems' controllability and observability). We point out that the algorithms implemented in this package are generic in the sense that they do not depend on the Ore algebras.

To define, manipulate, and compute in Ore algebras of functional operators, we use the `Mathematica` package HOLONOMICFUNCTIONS described in the previous section. The package OREALGEBRAICANALYSIS extends these Gröbner basis techniques to finitely presented left modules over the same classes of Ore algebras. It also contains algorithms for module theory (e.g., test whether or not a module admits torsion elements, is torsion-free, reflexive, projective, stably free, free) and homological algebra (e.g., computation of free resolutions, projective dimension, extension modules with value in the underlying ring, homological invariants, …).

[1] This work was supported by the *PHC PARROT 29586NG* between France and Estonia.

The OREALGEBRAICANALYSIS package includes the main procedures implemented in the `Maple` packages OREMODULES [10] and OREMORPHISMS [13]. Since HOLONOMICFUNCTIONS can handle larger classes of Ore algebras than the `Maple` package ORE_ALGEBRA,[2] OREALGEBRAICANALYSIS can study larger classes of linear functional systems than the `Maple` packages OREMODULES and OREMORPHISMS. Moreover, the internal design of `Mathematica` can allow us to consider classes of systems which could not easily be considered in `Maple` such as generic linearizations of nonlinear functional systems defined by explicit equations and systems containing transcendental functions (e.g., trigonometric functions, special functions). See the following examples.

We will now shortly illustrate the main functions and applications of the OREALGEBRAICANALYSIS package with explicit examples. For more examples, see [15].

Example 22 Let us consider an example studied in [42]. We start the `Mathematica` session by loading the package

```
In[24]:= << OreAlgebraicAnalysis`
```

and then entering the system equations in the form:

```
In[25]:= eqs = {x1'[t] → x1[t]u[t] + u[t − 2],
                x2'[t] → u[t] + u[t − 1],
                x3'[t] → u[t − 1] − u[t − 2]};
         vars = {x1[t], x2[t], x3[t], u[t]};
```

Let us now introduce the following Ore algebra A of DTD operators:

```
In[26]:= replA = ModelToReplacementRules[ eqs, t ];
           A = OreAlgebraWithRelations[ Der[t], S[−1][t], replA ]
```

$$\text{Out[26]=}\ \mathbb{K}(t)[D_t; 1, D_t][\left(S_t^{-1}\right); \#1/.\, t \to t-1\&,\ 0\&]$$

The matrix R of DTD operators which defines the generic linearization of the above nonlinear system is then given by:

```
In[27]:= MatrixForm[R = ToOrePolynomialD[ eqs, vars, A ]]
```

$$\text{Out[27]=}\ \begin{pmatrix} D_t - u[t] & 0 & 0 & -\left(S_t^{-1}\right)^2 - x_1[t] \\ 0 & D_t & 0 & -\left(S_t^{-1}\right) - 1 \\ 0 & 0 & D_t & \left(S_t^{-1}\right)^2 - \left(S_t^{-1}\right) \end{pmatrix}$$

Let $M = A^{1\times 4}/(A^{1\times 3}\, R)$ be the left A-module finitely presented by the matrix R. The adjoint of R is then defined by:

```
In[28]:= MatrixForm[ Radj = Involution[ R, A ]]
```

$$\text{Out[28]=}\ \begin{pmatrix} D_t - u[-t] & 0 & 0 \\ 0 & D_t & 0 \\ 0 & 0 & D_t \\ -\left(S_t^{-1}\right)^2 - x_1[-t] & -\left(S_t^{-1}\right) - 1 & \left(S_t^{-1}\right)^2 - \left(S_t^{-1}\right) \end{pmatrix}$$

Let us check whether or not M is a torsion-free left A-module:

[2] http://algo.inria.fr/chyzak/Mgfun/Sessions/Ore_algebra.html.

```
In[29]:= {Ann, Rp, Q} = Exti[ Radj, A, 1 ];
         MatrixForm[ Ann ]
```

$$\text{Out[29]=} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & D_t \end{pmatrix}$$

```
In[30]:= MatrixForm[ Rp ]
```

$$\text{Out[30]=} \begin{pmatrix} 0 & -D_t & 0 & \left(S_t^{-1}\right)+1 \\ -D_t+u[t] & 0 & -D_t & \left(S_t^{-1}\right)+x_1[t] \\ 0 & 0 & D_t & \left(S_t^{-1}\right)^2-\left(S_t^{-1}\right) \\ 0 & -\left(S_t^{-1}\right)^2+\left(S_t^{-1}\right)-\left(S_t^{-1}\right)-1 & 0 & \end{pmatrix}$$

The matrix Q is a parametrization of the controllable part (it is too large to be printed here; see [15]). Since `Ann` is not the identity matrix, we deduce that M admits nontrivial torsion elements and thus the corresponding system admits autonomous elements $\tau_1, \ldots, \tau_4$, defined by:

```
In[31]:= {aut, eqs, rels} = AutonomousElements[ R,
           {dx1[t], dx2[t], dx3[t], du[t]}, τ, A, Relations → True];
         aut
```

$$\begin{aligned}\text{Out[31]=}\ & \{\tau[1][t] \to du[-1+t]+du[t]-dx_2'[t],\\ & \tau[2][t] \to du[-1+t]+u[t]dx_1[t]+du[t]x_1[t]-dx_1'[t]-dx_3'[t],\\ & \tau[3][t] \to du[-2+t]-du[-1+t]+dx_3'[t],\\ & \tau[4][t] \to -dx_2[-2+t]+dx_2[-1+t]-dx_3[-1+t]-dx_3[t]\,\}\end{aligned}$$

```
In[32]:= eqs
```

$$\text{Out[32]=}\ \{\,\tau[1][t] == 0,\ \tau[2][t] == 0,\ \tau[3][t] == 0,\ \tau[4]'[t] == 0\,\}$$

```
In[33]:= rels
```

$$\begin{aligned}\text{Out[33]=}\ & \{-\tau[2][t]-\tau[3][t] == 0,\ -\tau[1][t] == 0,\ \tau[3][t] == 0,\\ & -\tau[1][-2+t]+\tau[1][-1+t]+\tau[3][-1+t]+\tau[3][t]+\tau[4]'[t] == 0\,\}\end{aligned}$$

We note that the first three autonomous elements τ_1, τ_2, τ_3 are trivial. The only nontrivial autonomous element is $\tau_4 = -\mathrm{d}x_2(t-2) + \mathrm{d}x_2(t-1) - \mathrm{d}x_3(t-1) - \mathrm{d}x_3(t)$, which satisfies $\dot{\tau}_4 = 0$.

Example 23 Let us consider the following nonlinear DTD system considered in [6]:

```
In[34]:= eqs = {x1'[t] → x2[t − 1]u[t],
                x2'[t] → x3[t]u[t],
                x3'[t] → u[t]};
         vars = {x1[t], x2[t], x3[t], u[t]};
```

Let us introduce the following Ore algebra A of DTD operators

```
In[35]:= replA = ModelToReplacementRules[ eqs, t ];
         A = OreAlgebraWithRelations[ Der[t], S[−1][t], replA ]
```

$$\text{Out[35]=}\ \mathbb{K}(t)[D_t; 1, D_t][\left(S_t^{-1}\right);\ \#1/.\,t \to t-1\&,\ 0\&]$$

the matrix R of DTD operators which defines the generic linearization of the above nonlinear system

In[36]:= **MatrixForm[R = ToOrePolynomialD[eqs, vars, A]]**

Out[36]=
$$\begin{pmatrix} D_t & -u[t]\left(S_t^{-1}\right) & 0 & -x_2[t-1] \\ 0 & D_t & -u[t] & -x_3[t] \\ 0 & 0 & D_t & -1 \end{pmatrix}$$

and the left A-module $M = A^{1\times 4}/(A^{1\times 3}\,R)$ finitely presented by R. Let us first compute the adjoint of R:

In[37]:= **MatrixForm[Radj = Involution[R, A]]**

Out[37]=
$$\begin{pmatrix} D_t & 0 & 0 \\ -u[1-t]\left(S_t^{-1}\right) & D_t & 0 \\ 0 & -u[-t] & D_t \\ -x_2[-1-t] & -x_3[-t] & -1 \end{pmatrix}$$

Let us check whether or not M is a torsion-free left A-module:

In[38]:= **{Ann, Rp, Q} = Simplify[Exti[Radj, A, 1]];**
MatrixForm[Ann]

Out[38]=
$$\begin{pmatrix} D_t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & u[t]D_t - u'[t] \end{pmatrix}$$

In[39]:= **MatrixForm[Rp]**

Out[39]=
$$\begin{pmatrix} 0 & -1 & x_3[t] & 0 \\ 0 & 0 & -D_t & 1 \\ D_t & 0 & -u[t]\,x_3[t-1]\left(S_t^{-1}\right) & -x_3[t-1] \end{pmatrix}$$

The matrix Q is too large to be printed here. For more details, see [15].

In[40]:= **{aut, eqs, rels} = AutonomousElements[R, {dx$_1$[t], dx$_2$[t], dx$_3$[t], du[t]},**
τ, A, Relations → True]; aut

Out[40]= $\{\tau[1][t] \to -\mathrm{dx}_2[t] + \mathrm{dx}_3[t]x_3[t],$
$\tau[2][t] \to \mathrm{du}[t] - \mathrm{dx}_3'[t],$
$\tau[3][t] \to -\mathrm{du}[t]x_2[t-1] - u[t]\mathrm{dx}_3[t-1]x_3[t-1] + \mathrm{dx}_1'[t]\}$

The autonomous elements τ_1, τ_2, τ_3 satisfy the following equations:

In[41]:= **eqs**

Out[41]= $\{\tau[1]'[t] == 0,$
$\tau[2][t] == 0,$
$-\tau[3][t]u'[t] + u[t]\tau[3]'[t] == 0\}$

The A-linear relations among the autonomous elements are given by:

In[42]:= **rels**

Out[42]= $\{u[t]\tau[1][t-1] + \tau[3][t] == 0,$
$-x[3][t]\tau[2][t] - \tau[1]'[t] == 0,$
$\tau[2][t] == 0\}$

Let us now prove that the set of autonomous elements can be generated by τ_1. Let us introduce the matrix L defining rels:

In[43]:= **MatrixForm[L = ToOrePolynomialD[rels,**
{τ[1][t], τ[2][t], τ[3][t]}, A]]

Out[43]=
$$\begin{pmatrix} u[t]\left(S_t^{-1}\right) & 0 & 1 \\ -D_t & -x_3[t] & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

Let us consider the following matrix

In[44]:= **MatrixForm[γ = {{1, 0, 0}}]**

Out[44]=
$$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$$

which corresponds to the position of τ_1. To express τ_2 and τ_3 in terms of τ_1, we first check whether or not the matrix T, formed by stacking L with γ, admits a left inverse.

In[45]:= **U = LeftInverse[T = Join[L, γ], A]**

Out[45]= $\{\{0, 0, 0, 1\}, \{0, 0, -1, 0\}, \{1, 0, 0, -u[t]\left(S_t^{-1}\right)\}\}$

Hence, if we consider the last column of the left inverse U of T, i.e.

In[46]:= **MatrixForm[V = Take[U, All, −1]]**

Out[46]=
$$\begin{pmatrix} 1 \\ 0 \\ -u[t]\left(S_t^{-1}\right) \end{pmatrix}$$

then we obtain:

In[47]:= **Thread[Table[τ[i][t], {i, 3}] → ApplyMatrix[V, {τ[1][t]}]]**

Out[47]= $\{\tau[1][t] \to \tau[1][t], \tau[2][t] \to 0, \tau[3][t] \to -u[t]\tau[1][t-1]\}$

From this point, we will use some procedures which are not freely available (see [2]). Finally, let us integrate the one-form defined by τ_1:

In[48]:= **BookForm[sp = SpanK[{ApplyMatrixD[Rp[[1]], vars]}, t]]**

Out[48]= SpanK$[-\mathrm{d}x_2[t] + x_3[t]\mathrm{d}x_3[t]]$

In[49]:= **IntegrateOneForms[sp]**

Out[49]= $\{x_2[t] - \frac{1}{2}x_3[t]^2\}$

Thus, x_1 is an autonomous element of the nonlinear DTD system.

References

1. Becker, T., Kredel, H., Weispfenning, V.: Gröbner Bases: A Computational Approach to Commutative Algebra. Springer, London (1993)
2. Belikov, J., Kaparin, V., Kotta, Ü., Tõnso, M.: NLControl: a software project addressing nonlinear control systems. http://www.nlcontrol.ioc.ee
3. Bergman, G.M.: The diamond lemma for ring theory. Adv. Math. **29**, 178–218 (1978)
4. Bronstein, M., Petkovšek, M.: An introduction to pseudo-linear algebra. Theor. Comput. Sci. **157**(1), 3–33 (1996)
5. Buchberger, B.: Ein Algorithmus zum Auffinden der Basiselemente des Restklassenrings nach einem nulldimensionalen Polynomideal. Ph.D. thesis, University of Innsbruck (1965); English translation: J. Symb. Comput. **41**(3–4), 475–511 (2006)

6. Califano, C., Li, S., Moog, C.: Controllability of driftless nonlinear time-delay systems. J. Symb. Comput. **62**, 294–301 (2013)
7. Chakhar, A., Cluzeau, T., Quadrat, A.: An algebraic analysis approach to certain classes of nonlinear partial differential systems. In: Proceedings of nDS'11, Poitiers, France, 05–07 Sept 2011
8. Chyzak, F.: An extension of Zeilberger's fast algorithm to general holonomic functions. Discret. Math. **217**(1–3), 115–134 (2000)
9. Chyzak, F., Quadrat, A., Robertz, D.: Effective algorithms for parametrizing linear control systems over Ore algebras. Appl. Algebr. Eng., Commun. Comput. **16**, 319–376 (2005)
10. Chyzak, F., Quadrat, A., Robertz, D.: OreModules: A Symbolic Package for the Study of Multidimensional Linear Systems. Lecture Notes in Control and Information Sciences, vol. 352, pp. 233–264. Springer, Berlin (2007). https://who.rocq.inria.fr/Alban.Quadrat/OreModules/index.html
11. Chyzak, F., Salvy, B.: Non-commutative elimination in Ore algebras proves multivariate identities. J. Symb. Comput. **26**, 187–227 (1998)
12. Cluzeau, T., Quadrat, A.: Factoring and decomposing a class of linear functional systems. Linear Algebr. Its Appl. **428**, 324–381 (2008)
13. Cluzeau, T., Quadrat, A.: OreMorphisms: A Homological Algebraic Package for Factoring, Reducing and Decomposing Linear Functional Systems. Lecture Notes in Control and Information Sciences, vol. 388, pp. 179–194. Springer, Berlin (2009). https://who.rocq.inria.fr/Alban.Quadrat/OreMorphisms/index.html
14. Cluzeau, T., Quadrat, A.: Equivalences of linear functional systems. In this book
15. Cluzeau, T., Quadrat, A., Tõnso, M.: OreAlgebraicAnalysis: A `Mathematica` *package for the algorithmic study of linear functional systems.* OreAlgebraicAnalysis project (2015). https://who.rocq.inria.fr/Alban.Quadrat/OreAlgebraicAnalysis/index.html
16. McConnell, J.C., Robson, J.C.: Noncommutative Noetherian Rings. American Mathematical Society, Providence (2000)
17. Cox, D., Little, J., O'Shea, D.: Ideals, Varieties, and Algorithms. Springer, New York (1992)
18. Cox, D., Little, J., O'Shea, D.: Using Algebraic Geometry. Graduate Texts in Mathematics, vol. 185. Springer, Berlin (2005)
19. Eder, C., Faugère, J.-C.: A survey on signature-based Gröbner basis computations (2014). arXiv:1404.1774
20. Fabiańska, A., Quadrat, A.: Applications of the Quillen-Suslin theorem to multidimensional systems theory. Gröbner Bases in Control Theory and Signal Processing. Radon Series on Computation and Applied Mathematics, vol. 3, pp. 23–106. de Gruyter Publisher, Berlin (2007). https://who.rocq.inria.fr/Alban.Quadrat/QuillenSuslin/index.html
21. Fliess, M.: Some basic structural properties of generalized linear systems. Syst. Control Lett. **5**, 391–396 (1990)
22. Fliess, M., Lévine, J., Martin, P., Rouchon, P.: Flatness and defect of nonlinear systems: introductory theory and examples. Int. J. Control **61**, 1327–1361 (1995)
23. Fliess, M., Mounier, H.: Controllability and observability of linear delay systems: an algebraic approach. ESAIM: Control., Optim. Calc. Var. **3**, 301–314 (1998)
24. Greuel, G.-M., Pfister, G.: A Singular Introduction to Commutative Algebra. Springer, Berlin (2008)
25. Hotta, R., Takeuchi, K., Tanisaki, T.: D-Modules, Perverse Sheaves, and Representation Theory. Progress in Mathematics, vol. 236. Birkhäuser, Basel (2008)
26. Kandri-Rody, A., Weispfenning, V.: Non-commutative Gröbner bases in algebras of solvable type. J. Symb. Comput. **9**(1), 1–26 (1990)
27. Kailath, T.: Linear Systems. Prentice-Hall, Upper Saddle River (1980)
28. Kashiwara, M.: Algebraic Study of Systems of Partial Differential Equations. Mémoires de la Société Mathématique de France, vol. 63 (1995); English translation (Kyoto 1970)
29. Koutschan, C.: Advanced applications of the holonomic systems approach. Ph.D. thesis, University of Linz, Austria (2009)

30. Koutschan, C.: HolonomicFunctions (user's guide). RISC Report Series, Johannes Kepler University, 10 Jan 2010. http://www.risc.jku.at/research/combinat/software/HolonomicFunctions/
31. Kredel, H.: Solvable Polynomial Rings. Shaker (1993)
32. Kwakernaak, H., Sivan, R.: Linear Optimal Control Systems. Wiley, Hoboken (1972)
33. Janet, M.: Leçons sur les systèmes d'équations aux dérivées partielles. Gauthier-Villars (1929)
34. Lam, T.Y.: Lectures on Modules and Rings. Graduate Texts in Mathematics, vol. 189. Springer, Berlin (1999)
35. Lange-Hegermann, M., Robertz, D.: Thomas Decomposition and Nonlinear Control Systems. In this volume
36. Levandovskyy, V., Schönemann, H.: Plural — A computer algebra system for non-commutative polynomial algebras. In: Proceedings of ISSAC'03, pp. 176–183. ACM Press (2003). https://www.singular.uni-kl.de/Manual/4-0-2/sing_469.htm
37. Levandovskyy, V.: Non-commutative computer algebra for polynomial algebras: Gröbner bases, applications and implementation. Ph.D. thesis, University of Kaiserslautern (2005)
38. Levandovskyy, V.: Computing diagonal form and Jacobson normal form of a matrix using Gröbner bases. J. Symb. Comput. **46**, 595–608 (2011)
39. Macaulay 2 project. http://www.math.uiuc.edu/Macaulay2/
40. Malgrange, B.: Systèmes différentiels à coefficients constants. Séminaire Bourbaki **1962**(63), 1–11 (1962)
41. Manitius, A.: Feedback controllers for a wind tunnel model involving a delay: analytical design and numerical simulations. IEEE Trans. Autom. Control **29**, 1058–1068 (1984)
42. Marquez-Martinez, L.A.: A note on the accessibility for nonlinear time-delay systems. Comptes Rendus de l'Académie des Sciences, Paris, t. 329, série 1, pp. 545–550 (1999)
43. Oberst, U.: Multidimensional constant linear systems. Acta Applicandae Mathematicae **20**, 1–175 (1990)
44. Ore, Ø.: Theory of non-commutative polynomials. Ann. Math., Second. Ser. **34**(3), 480–508 (1933)
45. Pommaret, J.-F.: Partial Differential Control Theory. Kluwer, Netherlands (2001)
46. Pommaret, J.-F., Quadrat, A.: Algebraic analysis of linear multidimensional control systems. IMA J. Math. Control Inf. **16**, 275–297 (1999)
47. Quadrat, A.: An introduction to constructive algebraic analysis and its applications. Les cours du CIRM, Journées Nationales de Calcul Formel **1**(2), 281–471 (2010)
48. Quadrat, A., Robertz, D.: Computation of bases of free modules over the Weyl algebras. J. Symb. Comput. **42**, 1113–1141 (2007). https://who.rocq.inria.fr/Alban.Quadrat/OreModules/stafford.html
49. Quadrat, A., Robertz, D.: A constructive study of the module structure of rings of partial differential operators. Acta Applicandæ Mathematicæ **133**, 187–234 (2014)
50. Quadrat, A., Ushirobira, R.: Algebraic analysis for the Ore extension ring of differential time-varying delay operators. In: Proceedings of MTNS 2016
51. Robertz, D.: Recent progress in an algebraic analysis approach to linear systems. Multidimens. Syst. Signal Process. **26**, 349–388 (2015)
52. Rotman, J.J.: An Introduction to Homological Algebra. Springer, Berlin (2009)
53. Takayama, N.: Kan: a system for computation in algebraic analysis. http://www.math.kobe-u.ac.jp/KAN/index.html
54. Zeilberger, D.: A holonomic systems approach to special functions identities. J. Comput. Appl. Math. **32**(3), 321–368 (1990)
55. Zeilberger, D.: The method of creative telescoping. J. Symb. Comput. **11**, 195–204 (1991)
56. Zerz, E.: Topics in Multidimensional Linear Systems Theory. Lecture Notes in Control and Information Sciences, vol. 256. Springer, Berlin (2000)
57. Zerz, E.: An algebraic analysis approach to linear time-varying systems. IMA J. Math. Control Inf. **23**, 113–126 (2006)
58. Zerz, E., Levandowskyy, V.: Algebraic systems theory and computer algebraic methods for some classes of linear control systems. In: Proceedings of MTNS 2006. CLIPS PROJECT: http://www.math.rwth-aachen.de/~Eva.Zerz/CLIPS/

Chapter 2
Equivalences of Linear Functional Systems

Thomas Cluzeau and Alban Quadrat

Abstract Within the algebraic analysis approach to linear systems theory, we investigate the *equivalence problem* of linear functional systems, i.e., the problem of characterizing when all the solutions of two linear functional systems are in a one-to-one correspondence. To do that, we first provide a new characterization of isomorphic finitely presented modules in terms of inflation of their presentation matrices. We then prove several isomorphisms which are consequences of the *unimodular completion problem*. We then use these isomorphisms to complete and refine existing results concerning *Serre's reduction problem*. Finally, different consequences of these results are given. All the results obtained here are algorithmic for rings for which Gröbner basis techniques exist and the computations can be performed by the `Maple` packages OreModules and OreMorphisms or the `Mathematica` package `OreAlgebraicAnalysis`.

Keywords Linear systems theory · Equivalence problem · Control theory · Algebraic analysis · Computer algebra

2.1 Introduction

Mathematical systems which are studied in control theory, mathematical physics, and engineering sciences can usually be modeled by systems of *functional equations*, namely, equations whose unknowns are functions. These functions can depend on one or more continuous or discrete variables. Standard examples of functional

T. Cluzeau
Université de Limoges, CNRS, XLIM UMR 7252, DMI, 123 avenue Albert Thomas, 87060 Limoges Cedex, France
e-mail: thomas.cluzeau@unilim.fr

A. Quadrat (✉)
Inria Paris, Ouragan Project-Team, Institut de Mathématiques de Jussieu-Paris Rive Gauche, Sorbonne University, 4 place Jussieu, 75252 Paris Cedex 05, France
e-mail: alban.quadrat@inria.fr

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_2

equations are ordinary differential (OD) or partial differential (PD) equations, (partial) difference equations, differential time-delay equations, … Functional systems can be studied by a large amount of mathematical theories as functional analysis, numerical analysis, differential geometry, … In this paper, we focus on linear functional systems, i.e., on the case where the functional equations are linear. In particular, we use the *algebraic analysis approach* to linear systems theory to study built-in properties of linear functional systems. Algebraic analysis has been developed by Malgrange, Bernstein, Sato, Kashiwara, … For more details, see [14, 15, 17, 19, 21] and the references therein.

We shall study here linear functional systems which can be written as $R\,\eta = 0$, where R is a $q \times p$ matrix with entries in a (noncommutative) polynomial ring D of functional operators (e.g., OD or PD operators, shift operators, difference operators, OD time-delay operators) and η is a vector of unknown functions which belong to a functional space (e.g., smooth functions, distributions, hyperfunctions). More precisely, if $\mathcal{F}$ is a left D-module (see, e.g., [16, 24]), then we can consider the following *linear system*

$$\ker_{\mathcal{F}}(R.) := \{\eta \in \mathcal{F}^p \mid R\,\eta = 0\},$$

also called a *behavior* in control theory (see [19] and the references therein). The algebraic analysis approach to linear systems theory (see [3, 13, 19, 21, 23] and the references therein) is based on the fact that the linear system $\ker_{\mathcal{F}}(R.)$ can be studied by means of the *factor left D-module* $M := D^{1\times p}/(D^{1\times q}\,R)$ *finitely presented by the matrix* R. Indeed, Malgrange's isomorphism [17] states that we have

$$\ker_{\mathcal{F}}(R.) \cong \hom_D(M, \mathcal{F}),$$

where $\hom_D(M, \mathcal{F})$ denotes the abelian group (i.e., $\mathbb{Z}$-module) of all the left D-homomorphisms (i.e., left D-linear maps) from M to $\mathcal{F}$ (see Sect. 2.2 for more details). Hence, module properties of M and $\mathcal{F}$ are connected to system-theoretical properties of $\ker_{\mathcal{F}}(R.)$. Using constructive methods of *homological algebra* [24] for *Gröbner rings* D (namely, (noncommutative) polynomial rings for which *Gröbner bases* can be computed for all *admissible term orders* by means of *Buchberger's algorithm* [5]) [3, 6, 21], we can effectively characterize module properties of M which are important in control theory (see [3, 13, 19, 21, 23] and references therein). For more details, see Chap. 1. The corresponding algorithms are implemented in dedicated packages of computer algebra systems (e.g., OREMODULES [4] and OREMORPHISMS [7] developed in `Maple`, and OREALGEBRAICANALYSIS [12] developed in `Mathematica`).

The purpose of the paper is to use the algebraic analysis framework to consider the following three important issues in mathematical systems (resp., module) theory:

(a) *Equivalence problem*: Recognize whether or not two linear systems (resp., finitely presented modules) are isomorphic.
(b) *Unimodular completion problem*: Inflate (if possible) a given (rectangular) matrix into a unimodular, namely, an invertible (square) matrix.

(c) *Serre's reduction problem*: Find an equivalent system defined by fewer equations and fewer unknowns.

The first contribution of the chapter (see Theorem 1) provides an explicit characterization of isomorphic finitely presented modules in terms of inflations of their presentation matrices. This characterization yields a general characterization of equivalent linear systems which do not necessarily have the same number of unknowns and equations. A constructive version of the classical *Schanuel's lemma* (see, e.g., [24]) on the *syzygy modules* of these modules can then be found again as a direct application of Theorem 1. If D is a *stably finite ring* (e.g., a *noetherian ring*) (see, e.g., [16]) and one of the presentation matrices has full row rank, then this result yields a characterization of isomorphic modules in terms of the unimodular completion problem (which also characterizes Serre's reduction problem [1]). The second contribution (see Theorem 2) is to show how the completion problem induces isomorphisms between the different modules finitely presented by the matrices appearing in the inflations. This result can be seen as an extension of a result obtained for Serre's reduction problem in [1] (extension for non necessarily full row rank matrices). The results are illustrated by explicit examples where all the computations can be performed using the packages OREMODULES [4] and OREMORPHISMS [7].

The paper is organized as follows. In Sect. 2.2, we briefly review the algebraic analysis approach to linear systems theory. In Sect. 2.3, we recall useful results of [6] on homomorphisms and isomorphisms of finitely presented left D-modules. In Sect. 2.4, we give an explicit characterization of the inverse of an isomorphism and a characterization of isomorphic finitely presented modules in terms of inflations of their presentation matrices. Interesting consequences of this result in linear systems theory are then given. In Sect. 2.5, we give our second main result on the different isomorphisms induced by a solution to the unimodular completion problem. Finally, this result is applied to Serre's reduction problem to refine a result obtained in [1].

2.2 Linear Functional Systems and Finitely Presented Left Modules

In this section, we show how a linear system defines a *finitely presented left D-module* and conversely. This correspondence plays a fundamental role in what follows as linear systems will be studied by means of the corresponding modules.

Let D be a noetherian ring and $R \in D^{q\times p}$ a matrix defining the linear system $\ker_{\mathcal{F}}(R.) := \{\eta \in \mathcal{F}^p \mid R\,\eta = 0\}$ for a certain left D-module $\mathcal{F}$ (see Sect. 2.1). Using the matrix $R \in D^{q\times p}$, we can define the following multiplication map:

$$\begin{aligned} .R\colon D^{1\times q} &\longrightarrow D^{1\times p} \\ \lambda &\longmapsto \lambda\, R. \end{aligned}$$

Since D is a (noncommutative) ring and not a (skew) field, $D^{1\times q}$ and $D^{1\times p}$ are *(left/right) D-modules*. We recall that a module is an algebraic structure defined by the same conditions as those for a vector space but where the scalars belong to a ring and not a (skew) field (see, e.g., [16, 24]). If M_1 and M_2 are two left D-modules, then a homomorphism f from M_1 to M_2, which is denoted by $f \in \hom_D(M_1, M_2)$, is a map $f : M_1 \longrightarrow M_2$ satisfying the following condition:

$$\forall\, d_1,\, d_2 \in D,\ \forall\, m_1,\, m_2 \in M_1 :\ f(d_1\, m_1 + d_2\, m_2) = d_1\, f(m_1) + d_2\, f(m_2).$$

For all $\lambda_1,\ \lambda_2 \in D^{1\times q}$ and for all $d_1,\ d_2 \in D$, we have

$$(.R)(d_1\, \lambda_1 + d_2\, \lambda_2) = d_1\, (\lambda_1\, R) + d_2\, (\lambda_2\, R) = d_1\, ((.R_1)(\lambda_1)) + d_2\, ((.R_2)(\lambda_2)),$$

i.e., $.R \in \hom_D(D^{1\times q}, D^{1\times p})$. Similarly, we can define homomorphisms for right D-modules. The image $\mathrm{im}_D(.R) := \{\mu\, R \mid \mu \in D^{1\times q}\}$ of $.R$, also simply denoted by $D^{1\times q}\, R$, is the left D-module formed by all the left D-linear combinations of the rows of the matrix R. The *cokernel* of $.R$ is defined by the following *factor* left D-module:

$$M := D^{1\times p}/(D^{1\times q}\, R).$$

Two vectors $\lambda_1,\ \lambda_2 \in D^{1\times p}$ are said to belong to the same *residue class*, which is denoted by $\pi(\lambda_1) = \pi(\lambda_2)$, if we have $\lambda_1 - \lambda_2 \in D^{1\times q}\, R$, i.e., if there exists $\mu \in D^{1\times q}$ such that $\lambda_1 = \lambda_2 + \mu\, R$. The left D-module M is then defined by all the $\pi(\lambda)$'s for $\lambda \in D^{1\times p}$ with the following two binary operations:

$$\forall\, \lambda_1,\ \lambda_2 \in D^{1\times p},\ d \in D :\ \pi(\lambda_1 + \lambda_2) := \pi(\lambda_1) + \pi(\lambda_2),\ \pi(d\, \lambda_1) := d\, \pi(\lambda_1).$$

We can check that $\pi(\lambda_1) + \pi(\lambda_2)$ and $d\, \pi(\lambda_1)$ do not depend on the choice of the *representatives* $\lambda_1,\ \lambda_2$ of the residues classes $\pi(\lambda_1)$ and $\pi(\lambda_2)$, which shows that the two above binary operations are well-defined on M and $\pi \in \hom_D(D^{1\times p}, M)$ is called the canonical projection onto M.

The left D-module M is said to be *finitely presented* and R is called a *presentation matrix* [16, 24]. Let us explicitly describe M by means of *generators* and *relations*. If $\{f_j\}_{j=1,\ldots,p}$ denotes the *standard basis* of $D^{1\times p}$, namely, f_j is the row vector of length p formed by 1 at the jth position and 0 elsewhere, and $y_j := \pi(f_j)$ for $j = 1, \ldots, p$, then we claim that $\{y_j\}_{j=1,\ldots,p}$ is a generator set for M. Indeed, an element $m \in M$ is of the form $m = \pi(\lambda)$ for a certain $\lambda := (\lambda_1\ \ldots\ \lambda_p) = \sum_{j=1}^{p} \lambda_j\, f_j \in D^{1\times p}$, which yields $m = \sum_{j=1}^{p} \lambda_j\, y_j$ since $\pi \in \hom_D(D^{1\times p}, M)$. A left/right D-module which admits a finite set of generators is said to be *finitely generated*. The y_j's are not left D-linearly independent since, if $R_{i\bullet}$ denotes the ith row of R, using the fact that $R_{i\bullet} \in D^{1\times q}\, R$ and $\pi \in \hom_D(D^{1\times p}, M)$, we then obtain:

$$\forall\, i = 1, \ldots, q, \quad \sum_{j=1}^{p} R_{ij}\, y_j = \sum_{j=1}^{p} R_{ij}\, \pi(f_j) = \pi(R_{i\bullet}) = 0. \tag{2.1}$$

Hence, the set of generators $\{y_j\}_{j=1,\ldots,p}$ satisfies the *left D-linear relations* (2.1). If we note $y := (y_1 \ \ldots \ y_p) \in M^p$, then (2.1) can be rewritten as $R\,y = 0$.

If $\mathcal{F}$ is a left D-module, then we can define the following *behavior*

$$\ker_{\mathcal{F}}(R.) := \{\eta \in \mathcal{F}^p \mid R\,\eta = 0\},$$

i.e., the space of $\mathcal{F}$-solutions of $R\,\eta = 0$. We claim that there is an *isomorphism* (namely, an injective and a surjective homomorphism) between $\ker_{\mathcal{F}}(R.)$ and $\hom_D(M, \mathcal{F})$, which is denoted by $\ker_{\mathcal{F}}(R.) \cong \hom_D(M, \mathcal{F})$. Let us describe this isomorphism. If $\phi \in \hom_D(M, \mathcal{F})$, $\{y_j\}_{j=1,\ldots,p}$ the set of generators of M defined above, and $\eta_j := \phi(y_j)$ for $j = 1, \ldots, p$, then, using (2.1), we get

$$\forall\, i = 1, \ldots, q, \quad \sum_{j=1}^{p} R_{ij}\,\eta_j = \sum_{j=1}^{p} R_{ij}\,\phi(y_j) = \phi\left(\sum_{j=1}^{p} R_{ij}\,y_j\right) = \phi(0) = 0,$$

i.e., $\eta := (\eta_1 \ \ldots \ \eta_p)^T \in \ker_{\mathcal{F}}(R.)$. Conversely, if $\eta = (\eta_1 \ \ldots \ \eta_p)^T \in \ker_{\mathcal{F}}(R.)$, then we can define $\phi_\eta : M \longrightarrow \mathcal{F}$ by $\phi_\eta(\pi(\lambda)) := \lambda\,\eta$ for all $\lambda \in D^{1\times p}$. If $\pi(\lambda) = \pi(\lambda')$, then there exists $\mu \in D^{1\times q}$ such that $\lambda = \lambda' + \mu\,R$, which yields $\lambda\,\eta = \lambda'\,\eta$ since $R\,\eta = 0$, which shows that $\phi_\eta(\pi(\lambda)) = \phi_\eta(\pi(\lambda'))$, i.e., ϕ_η does not depend on the representative λ of $\pi(\lambda)$. Clearly, we have $\phi_\eta \in \hom_D(M, \mathcal{F})$. Now, if $\eta \in \ker_{\mathcal{F}}(R.)$, then we get $\phi_\eta(y_j) = \phi_\eta(\pi(f_j)) = f_j\,\eta = \eta_j$, which shows that the additive map

$$\begin{aligned} \chi : \ker_{\mathcal{F}}(R.) &\longrightarrow \hom_D(M, \mathcal{F}) \\ \eta &\longmapsto \phi_\eta, \end{aligned} \tag{2.2}$$

is injective. It is also surjective since, for every $\phi \in \hom_D(M, \mathcal{F})$, we can define $\eta := (\phi(y_1) \ \ldots \ \phi(y_p))^T \in \ker_{\mathcal{F}}(R.)$ and we have

$$\forall\, \lambda \in D^{1\times p}, \quad \phi_\eta(\pi(\lambda)) := \lambda\,\eta = \sum_{j=1}^{p} \lambda_j\,\eta_j = \phi\left(\sum_{j=1}^{p} \lambda_j\,y_j\right) = \phi(\pi(\lambda)),$$

which shows that $\phi = \phi_\eta = \chi(\eta)$ and finally proves that we have the isomorphism:

$$\ker_{\mathcal{F}}(R.) \cong \hom_D(M, \mathcal{F}).$$

Remark 1 We note that $\phi_{d\,\eta}(\pi(\lambda)) = \lambda\,d\,\eta$ is usually different from $d\,\lambda\,\eta = d\,\phi_\eta(\pi(\lambda))$ when D is a noncommutative ring, i.e., χ is not a left D-homomorphism. It is only an abelian group (i.e., a $\mathbb{Z}$-module) homomorphism between abelian groups (i.e., $\mathbb{Z}$-modules). If D is a k-algebra, where k is a field, then $\hom_D(M, \mathcal{F})$ inherits a k-vector space structure and χ is then an isomorphism of k-vector spaces.

Hence, the behavior $\ker_{\mathcal{F}}(R.)$ is the "dual" of the finitely presented left D-module $M := D^{1\times p}/(D^{1\times q}\,R)$ [14, 15]. We pass from a finitely presented left D-module M

(the algebraic side of a linear system) to a behavior $\ker_{\mathcal{F}}(R.)$ (the analytical side of a linear system) by applying the *contravariant left exact functor* $\hom_D(\cdot, \mathcal{F})$ (see, e.g., [24]). In particular, the algebraic study of M yields information on the behavior $\ker_{\mathcal{F}}(R.)$. For more details, see Chap. 1 [3, 6, 13, 19, 21] and the references therein.

In mathematical systems theory and control theory, we usually focus on particular classes of linear functional systems such as linear OD systems or DTD systems. In this case, we consider an algebra D of functional operators such as *skew polynomial rings*, *Ore algebras*, *Ore extensions*, … For more details, see Chap. 1 [3, 5, 12, 18] and the references therein. Let us give an explicit example.

Example 1 Let us consider the following linear DTD system

$$\begin{cases} \dot{x}_1(t) = x_2(t) + u(t), \\ \dot{x}_2(t) = x_1(t - 3\,h) + x_1(t - 2\,h) + u(t), \end{cases} \tag{2.3}$$

where h is a non-negative real, i.e., $h \in \mathbb{R}_{\geqslant 0}$. Let us consider the differential operator $\partial\, z(t) := \dot{z}(t)$ and the time-delay operator $\delta\, z(t) := z(t - h)$ which satisfy

$$(\partial\,\delta)\, z(t) = \partial\, z(t - h) = \dot{z}(t - h) = (\delta\,\partial)\, z(t),$$

i.e., on the level of operators, we have $\partial\,\delta = \delta\,\partial$, where the product stands for the composition of operators. Let $D := \mathbb{Q}[\partial, \delta]$ be the commutative polynomial algebra formed by the operators in ∂ and δ with coefficients in $\mathbb{Q}$. An element $d \in D$ is of the form $d = \sum_{0 \leqslant i,j \leqslant r} a_{ij}\, \partial^i\, \delta^j$, where $r \in \mathbb{N}$, $a_{ij} \in \mathbb{Q}$, and $\partial^i\, z(t) = z^{(i)}(t)$ (resp., $\delta^i\, z(t) = z(t - i\,h)$) is the ith composition of ∂ (resp., of δ). Then, (2.3) can be rewritten as $R\,\eta = 0$, where $\eta := (x_1 \quad x_2 \quad u)^T$ and:

$$R := \begin{pmatrix} \partial & -1 & -1 \\ -\delta^2\,(\delta + 1) & \partial & -1 \end{pmatrix} \in D^{2\times 3}.$$

We consider the finitely presented D-module $M := D^{1\times 3}/(D^{1\times 2}\, R)$, $\{f_j\}_{j=1,2,3}$ is the standard basis of $D^{1\times 3}$, $x_1 := \pi(f_1)$, $x_2 := \pi(f_2)$, and $u := \pi(f_3)$, where $\pi : D^{1\times 3} \longrightarrow M$ is the canonical projection. Then, as previously shown, $\{x_1,\ x_2,\ u\}$ is a set of generators of M which satisfies the following D-linear relations:

$$\begin{cases} \partial x_1 - x_2 - u = 0, \\ \partial x_2 - \delta^2\,(\delta + 1)\, x_1 - u = 0. \end{cases}$$

It is important to note that x_1, x_2, and u are not functions but only the "abstract" generators of M. To get functions, i.e., elements of a functional space $\mathcal{F}$ having a D-module structure (e.g., $\mathcal{F} := C^\infty(\mathbb{R})$), we have to consider $\hom_D(M, \mathcal{F}) \cong \ker_{\mathcal{F}}(R.) = \{\eta = (x_1\ x_2\ u)^T \in \mathcal{F}^3 \mid R\,\eta = 0\}$. Dualizing M with coefficients in $\mathcal{F}$, the generators of M are then mapped to $\mathcal{F}$ functions, i.e., $x_1 \longmapsto x_1(\cdot) \in \mathcal{F}$, $x_2 \longmapsto x_2(\cdot) \in \mathcal{F}$ and $u \longmapsto u(\cdot) \in \mathcal{F}$, satisfying (2.3).

For more examples, see Chap. 1 [3, 6, 21] and the references therein.

Finally, let us shortly introduce a few basic concepts of *homological algebra* (see, e.g., [24]) which will be used thereafter. A sequence of left/right D-modules $\{M_i\}_{i\in\mathbb{Z}}$ and of left/right D-homomorphisms $\{f_i \in \hom_D(M_i, M_{i-1})\}_{i\in\mathbb{Z}}$ are called a *complex* of left/right D-modules if we have $f_i \circ f_{i+1} = 0$ for all $i \in \mathbb{Z}$, i.e., if we have $\text{im}\, f_{i+1} \subseteq \ker f_i$ for $i \in \mathbb{Z}$. The complex is then denoted by:

$$\ldots \xrightarrow{f_{i+2}} M_{i+1} \xrightarrow{f_{i+1}} M_i \xrightarrow{f_i} M_{i-1} \xrightarrow{f_{i-1}} \ldots$$

The above complex is said to be an *exact sequence* if $\ker f_i = \text{im}\, f_{i+1}$ for all $i \in \mathbb{Z}$. For instance, using the fact that $\text{coker}_D(.R) := D^{1\times p}/\text{im}_D(.R) = M$, we get the following exact sequence

$$0 \longrightarrow \ker_D(.R) \xrightarrow{i} D^{1\times q} \xrightarrow{.R} D^{1\times p} \xrightarrow{\pi} M \longrightarrow 0,$$

where i is the standard injection and $\ker_D(.R) := \{\mu \in D^{1\times q} \mid \mu R = 0\}$ is the left D-module, called the *second syzygy module* of M, generated by all the D-linear combinations among the rows of R. If the rows of R are D-linearly independent, i.e., $\ker_D(.R) = 0$, then we say that R has *full row rank*.

An exact sequence of the form $0 \longrightarrow M' \xrightarrow{f} M \xrightarrow{g} M'' \longrightarrow 0$, i.e., where g is surjective ($\text{im}\, g = \ker 0 = M''$), $\ker g = \text{im}\, f$, and f injective ($\ker f = \text{im}\, 0 = 0$), is called a *short exact sequence*. For instance, if R has full row rank, then we have the following short exact sequence of left D-modules:

$$0 \longrightarrow D^{1\times q} \xrightarrow{.R} D^{1\times p} \xrightarrow{\pi} M \longrightarrow 0. \tag{2.4}$$

Example 2 We consider again Example 1. Let us check that R has full row rank. We have $\mu := (\mu_1 \quad \mu_2) \in \ker_D(.R)$ if and only if

$$\begin{cases} \mu_1\,\partial - \mu_2\,\delta^2\,(\delta+1) = 0, \\ -\mu_1 + \mu_2\,\partial = 0, \\ \mu_1 + \mu_2 = 0, \end{cases} \quad \Rightarrow \quad \begin{cases} \mu_1 = -\mu_2, \\ \mu_2\,(\partial+1) = 0, \end{cases}$$

which yields $\mu_2 = 0$ since $D := \mathbb{Q}[\partial, \delta]$ is an integral domain (i.e., D does not contain nonzero zero-divisors), and thus we get $\mu = 0$. Hence, we have the short exact sequence (2.4) with $p = 3$ and $q = 2$.

If D is a Gröbner ring, then *elimination techniques* (e.g., *Gröbner bases*, *Janet bases*, ...) can be used to compute $\ker_D(.R)$ (see Chap. 1 and [3, 21] and the references therein). Indeed, a set of generators of $\ker_D(.R)$ corresponds to a set of generators of the *compatibility conditions* $\mu\,\zeta = 0$ of the inhomogeneous linear system $R\,\eta = \zeta$. Thus, we have to eliminate η from $R\,\eta = \zeta$ to get a set of generators for $\ker_D(.R)$. For more details, see, e.g., [3, 21] and the OREMODULES package [4].

Example 3 We consider again (2.3) with the output $y(t) := x_1(t) + x_2(t)$, i.e.:

$$\begin{pmatrix} \partial & -1 \\ -\delta^2\,(\delta+1) & \partial \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} u(t) \\ u(t) \\ y(t) \end{pmatrix}.$$

To simplify (for instance, for an observability test), let us suppose that we have $u = 0$ and $y = 0$ so that we get the following linear DTD system:

$$\begin{pmatrix} \partial & -1 \\ -\delta^2\,(\delta+1) & \partial \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = 0. \tag{2.5}$$

Let $R \in D^{3\times 2}$ be the above matrix of DTD operators and $M := D^{1\times 2}/(D^{1\times 3}\,R)$ the D-module associated with (2.5). Using elimination techniques (see, e.g., [3, 21]) and their implementations in the OREMODULES package [4], we can check that we have $\ker_D(.R) = \mathrm{im}_D(.R_2)$, where:

$$R_2 := (\partial + \delta^2\,(\delta+1) \quad \partial + 1 \quad -\partial^2 + \delta^2\,(\delta+1)) \in D^{1\times 3}.$$

The row vector R_2 generates the D-module $\ker_D(.R)$ formed by the D-linear relations among the rows of R. We can check again that $R_2\,\zeta = 0$ generates the compatibility conditions of $R\,\eta = \zeta$. We note that $.R_2 \in \hom_D(D, D^{1\times 3})$ is injective since $\nu\,R_2 = 0$ yields $\nu\,(\partial + 1) = 0$, and thus we get $\nu = 0$ since D is an integral domain. Then, we obtain the following long exact sequence of D-modules

$$0 \longrightarrow D \xrightarrow{.R_2} D^{1\times 3} \xrightarrow{.R} D^{1\times 2} \xrightarrow{\pi} M \longrightarrow 0,$$

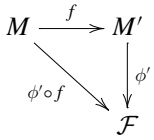
called a *finite free resolution* of the D-module M (see, e.g., [3, 21, 24]).

In Example 3, the D-module $\ker_D(.R)$ is a finitely generated D-module because $\ker_D(.R)$ is a D-submodule of the *noetherian* D-module $D^{1\times 3}$ (which is a direct sum of the *noetherian ring* $D := \mathbb{Q}[\partial, \delta]$). For more details, see, e.g., [16, 24]. In what follows, we shall assume that D is a *noetherian ring*, namely, every left/right ideal of D is finitely generated as a left/right D-module (see, e.g., [16, 24]). Then, for every matrix $R \in D^{q\times p}$, there always exists $R_2 \in D^{r\times q}$ (possibly reduced to 0) such that $\ker_D(.R) = \mathrm{im}_D(.R_2)$.

2.3 Homomorphisms of Behaviors/Finitely Presented Left Modules

Let $R \in D^{q\times p}$ and $R' \in D^{q'\times p'}$ be two matrices respectively defining the linear systems $R\,\eta = 0$ and $R'\,\eta' = 0$. In this section, we review results on transformations which map the $\mathcal{F}$-solutions of the first system to $\mathcal{F}$-solutions of the second one.

As learned in Sect. 2.2, we can define the two finitely presented left D-modules $M := D^{1\times p}/(D^{1\times q}\,R)$ and $M' := D^{1\times p'}/(D^{1\times q'}\,R')$ which are associated with the above linear systems. Now, composing $\phi' \in \hom_D(M', \mathcal{F}) \cong \ker_{\mathcal{F}}(R'.)$ with $f \in \hom_D(M, M')$, we obtain the following *commutative diagram*



and we get $f^\star(\phi') := \phi' \circ f \in \hom_D(M, \mathcal{F}) \cong \ker_{\mathcal{F}}(R.)$. If $\{y_j := \pi(f_j)\}_{j=1,\ldots,p}$ (resp., $\{y'_k := \pi(f'_k)\}_{k=1,\ldots,p'}$) is the set of generators of M (resp., M') defined as in Sect. 2.2, a solution $\eta' := (\phi'(y'_1)\ \ldots\ \phi'(y'_{p'}))^T$ of $R'\,\eta' = 0$ is sent to the solution $\eta := (\phi'(f(y_1))\ \ldots\ \phi'(f(y_p)))^T$ of $R\,\eta = 0$. To get an explicit description of η in terms of η', we have to explicitly know $f \in \hom_D(M, M')$, i.e., how f sends the y_j's to the y'_k's, i.e., to know the elements P_{jk} of D such that:

$$\forall\ j = 1, \ldots, p, \quad f(y_j) = \sum_{k=1}^{p'} P_{jk}\, y'_k. \tag{2.6}$$

Since f is a homomorphism, we have $f(0) = 0$. Using (2.1), for $i = 1, \ldots, q$, we get:

$$\begin{aligned} f\left(\sum_{j=1}^{p} R_{ij}\, y_j\right) = \sum_{j=1}^{p} R_{ij}\, f(y_j) &= \sum_{j=1}^{p} R_{ij}\left(\sum_{k=1}^{p'} P_{jk}\, y'_k\right) \\ &= \sum_{k=1}^{p'}\left(\sum_{j=1}^{p} R_{ij}\, P_{jk}\right)\, y'_k = 0. \end{aligned}$$

Using the fact that $y'_k := \pi'(f'_k)$, where $\pi' : D^{1\times p'} \longrightarrow M'$ is the canonical projection and $\{f'_k\}_{k=1,\ldots,p'}$ is the standard basis of $D^{1\times p'}$, we get

$$\pi'\left(\left(\sum_{j=1}^{p} R_{ij}\ P_{j1}\ \ \ldots\ \ \sum_{j=1}^{p} R_{ij}\ P_{jp'}\right)\right) = \sum_{k=1}^{p'}\left(\sum_{j=1}^{p} R_{ij}\,P_{jk}\right)\ y'_k = 0,$$

which shows the existence of row vectors $Q_i \in D^{1\times q'}$, $i = 1, \ldots, q$, such that:

$$\forall\ i = 1, \ldots, q, \quad \left(\sum_{j=1}^{p} R_{ij}\ P_{j1}\ \ \ldots\ \ \sum_{j=1}^{p} R_{ij}\ P_{jp'}\right) = Q_i\ R'.$$

If we note $P := (P_{jk})_{1\leqslant j\leqslant p, 1\leqslant k\leqslant p'} \in D^{p\times p'}$ and $Q := (Q_1^T\ \ldots\ Q_q^T)^T \in D^{q\times q'}$, then we obtain the following identity:

$$R\ P = Q\ R'. \tag{2.7}$$

Hence, we get that $f \in \hom_D(M, M')$ is defined by (2.6) where the P_{jk}'s satisfy (2.7).

Lemma 1 ([6]) *Let $M := D^{1\times p}/(D^{1\times q}\ R)$ (resp., $M' := D^{1\times p'}/(D^{1\times q'}\ R')$) be the left D-module finitely presented by $R \in D^{q\times p}$ (resp., $R' \in D^{q'\times p'}$) and $\pi : D^{1\times p} \longrightarrow M$ (resp., $\pi' : D^{1\times p'} \longrightarrow M'$) the canonical projection.*

(a) The existence of $f \in \hom_D(M, M')$ is equivalent to the existence of two matrices $P \in D^{p\times p'}$ and $Q \in D^{q\times q'}$ satisfying the following identity:

$$R\ P = Q\ R'.$$

Then, $f \in \hom_D(M, M')$ is defined by $f(\pi(\lambda)) = \pi'(\lambda\ P)$ for all $\lambda \in D^{1\times p}$, and we have the following commutative exact diagram

$$\begin{array}{ccccccc}
D^{1\times q} & \xrightarrow{.R} & D^{1\times p} & \xrightarrow{\pi} & M & \longrightarrow & 0 \\
\downarrow{\scriptstyle .Q} & & \downarrow{\scriptstyle .P} & & \downarrow{\scriptstyle f} & & \\
D^{1\times q'} & \xrightarrow{.R'} & D^{1\times p'} & \xrightarrow{\pi'} & M' & \longrightarrow & 0,
\end{array}$$

namely, every square commutes, i.e., $.P \circ .R = .R' \circ .Q$ and $f \circ \pi = \pi' \circ .P$.

(b) Let $R'_2 \in D^{q'_2\times q'}$ be such that $\ker_D(.R') = \mathrm{im}_D(.R'_2)$ and let $P \in D^{p\times p'}$ and $Q \in D^{q\times q'}$ be two matrices satisfying $R\ P = Q\ R'$. Then, the following matrices

$$\overline{P} := P + Z\ R', \quad \overline{Q} := Q + R\ Z + Z_2\ R'_2,$$

where $Z \in D^{p \times q'}$ *and* $Z_2 \in D^{q \times q'_2}$ *are two arbitrary matrices, satisfy the relation* $R\,\overline{P} = \overline{Q}\,R'$ *and* $f(\pi(\lambda)) = \pi'(\lambda\,P) = \pi'(\lambda\,\overline{P})$ *for all* $\lambda \in D^{1\times p}$.

For algorithms to compute the matrices P and Q for different classes of linear functional systems, we refer to [6] and the OREMORPHISMS and OREALGEBRAIC-ANALYSIS packages [7, 12].

Example 4 We consider again Example 3. Let $M' := D/(D^{1\times 2}\,R')$ be the $D := \mathbb{Q}[\partial, \delta]$-module finitely presented by the matrix $R' := (\partial^2 - \delta^2\ (\delta + 1)\quad \partial + 1)^T$, which corresponds to the following linear DTD system:

$$\begin{cases} \ddot{z}(t) - z(t - 3\,h) - z(t - 2\,h) = 0, \\ \dot{z}(t) + z(t) = 0. \end{cases} \tag{2.8}$$

Let $\pi' : D \longrightarrow M'$ be the canonical projection. A homomorphism $f : M \longrightarrow M'$ is defined by $f(\pi(\lambda)) = \pi'(\lambda\,P)$ for all $\lambda \in D^{1\times 2}$, where

$$P := \begin{pmatrix} 1 \\ \partial \end{pmatrix}, \quad Q := \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix},$$

since we can easily check that we have $R\,P = Q\,R'$.

Coming back to $\eta := (\phi'(f(y_1))\ \ldots\ \phi'(f(y_p)))^T$, using (2.6), we get

$$\eta_j := \phi'(f(y_j)) = \phi'\left(\sum_{k=1}^{p'} P_{jk}\,y'_k\right) = \sum_{k=1}^{p'} P_{jk}\,\phi'(y'_k) = \sum_{k=1}^{p'} P_{jk}\,\eta'_k,$$

which shows that $\eta := P\,\eta' \in \ker_{\mathcal{F}}(R.)$ for all $\eta' \in \ker_{\mathcal{F}}(R'.)$.

Corollary 1 *With the notations of Lemma 1, if* $\mathcal{F}$ *is a left* D*-module, then we have:*

$$\begin{aligned} P. : \ker_{\mathcal{F}}(R'.) &\longrightarrow \ker_{\mathcal{F}}(R.) \\ \eta' &\longmapsto \eta := P\,\eta'. \end{aligned} \tag{2.9}$$

The *contravariant functor* $\hom_D(\cdot, \mathcal{F})$ (see, e.g., [24]) transforms finitely presented left D-modules (resp., homomorphisms of finitely presented left D-modules) into $\mathcal{F}$-behaviors (resp., homomorphisms between $\mathcal{F}$-behaviors in the reverse direction).

Example 5 We consider again Examples 3 and 4. Using $f \in \hom_D(M, M')$, we have (2.9), where $P := (1\quad \partial)^T$, i.e., the additive mapping

$$z(t) \longmapsto \begin{cases} x_1(t) = z(t), \\ x_2(t) = \dot{z}(t), \end{cases} \tag{2.10}$$

sends $\mathcal{F}$-solutions of (2.8) to $\mathcal{F}$-solutions of (2.5), where $\mathcal{F}$ is a $D := \mathbb{Q}[\partial, \delta]$-module.

Let $f : M \longrightarrow M'$ be a homomorphism of left/right D-modules. Then, we can define the kernel, image, coimage, and cokernel of f as the following left/right D-modules:

$$\ker f := \{m \in M \mid f(m) = 0\}, \ \text{im}\, f := \{m' \in M' \mid \exists\, m \in M : \ m' = f(m)\},$$
$$\text{coim}\, f := M/\ker f, \qquad \text{coker}\, f := M'/\text{im}\, f.$$

Finally, let us explicitly characterize the latter modules.

Lemma 2 ([6]) *Let $M := D^{1\times p}/(D^{1\times q}\, R)$ (resp., $M' := D^{1\times p'}/(D^{1\times q'}\, R')$) be the left D-module finitely presented by $R \in D^{q\times p}$ (resp., $R' \in D^{q'\times p'}$). Moreover, let $f \in \hom_D(M, M')$ be defined by $P \in D^{p\times p'}$ and $Q \in D^{q\times q'}$ satisfying (2.7).*

(a) Let $S \in D^{r\times p}$ and $T \in D^{r\times q'}$ be two matrices such that

$$\ker_D(.(P^T \quad R'^T)^T) = \text{im}_D(.(S \quad -T)),$$

$L \in D^{q\times r}$ a matrix satisfying $R = L\, S$ and a matrix $S_2 \in D^{r_2\times r}$ such that $\ker_D(.S) = \text{im}_D(.S_2)$. Then, we have:

$$\ker f = (D^{1\times r}\, S)/(D^{1\times q}\, R) \cong D^{1\times r}/\left(D^{1\times(q+r_2)} \begin{pmatrix} L \\ S_2 \end{pmatrix}\right).$$

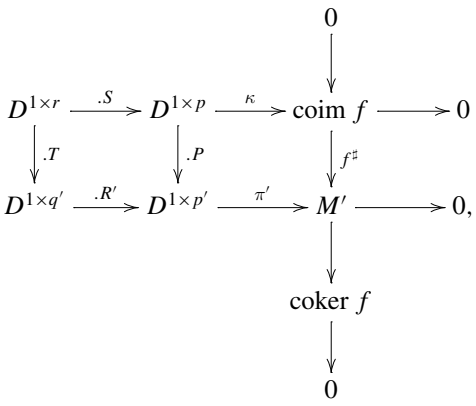
(b) With the above notations, we have:

$$\text{coim}\, f = D^{1\times p}/(D^{1\times r}\, S) \cong \text{im}\, f = \left(D^{1\times(p+q')} \begin{pmatrix} P \\ R' \end{pmatrix}\right)/(D^{1\times q'}\, R').$$

(c) We have $\text{coker}\, f = D^{1\times p'}/\left(D^{1\times(p+q')} \begin{pmatrix} P^T & R'^T \end{pmatrix}^T\right)$. *Thus,* coker f *admits the following beginning of a finite free resolution:*

$$D^{1\times r} \xrightarrow{.(S \quad -T)} D^{1\times(p+q')} \xrightarrow{. \begin{pmatrix} P \\ R' \end{pmatrix}} D^{1\times p'} \xrightarrow{\epsilon} \text{coker}\, f \longrightarrow 0. \tag{2.11}$$

(d) We have the following commutative exact diagram



where $f^\sharp : \operatorname{coim} f \longrightarrow M'$ *is defined by* $f^\sharp(\kappa(\lambda)) = \pi'(\lambda\, P)$ *for all* $\lambda \in D^{1\times p}$.

We note that $M := D^{1\times p}/(D^{1\times q}\, R)$ is the zero module if and only if we have $D^{1\times q}\, R = D^{1\times p}$, i.e., if and only if there exists a matrix $T \in D^{p\times q}$ such that $T\, R = I_p$, i.e., if and only if the presentation matrix R of M admits a *left inverse*. Using this result and Lemma 2, we can now characterize when $f \in \hom_D(M, M')$ is the zero homomorphism, injective, surjective or defines an isomorphism.

Lemma 3 ([6]) *With the notations of Lemma 2,* $f \in \hom_D(M, M')$ *is:*

(a) The zero homomorphism, i.e., $f = 0$*, if and only if one of the following equivalent conditions holds:*

a. There exists $Z \in D^{p\times q'}$ *such that* $P = Z\, R'$*. If* $R'_2 \in D^{q'_2\times q'}$ *is a matrix satisfying* $\ker_D(.R') = \operatorname{im}_D(.R'_2)$*, then there exists* $Z_2 \in D^{q\times q'_2}$ *such that:*

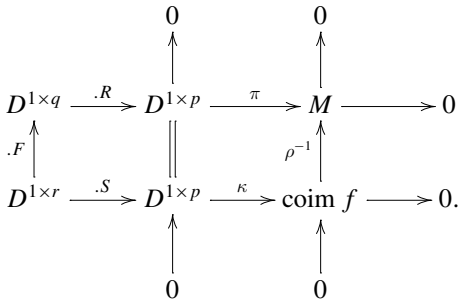
$$Q = R\, Z + Z_2\, R'_2.$$

b. The matrix S *admits a left inverse, i.e., there exists* $X \in D^{p\times r}$ *such that:*

$$X\, S = I_p.$$

(b) Injective, i.e., $\ker f = 0$*, if and only if one of the following equivalent conditions holds:*

a. There exists $F \in D^{r\times q}$ *such that* $S = F\, R$*. Then, if* $\rho : M \longrightarrow \operatorname{coim} f$ *is the canonical projection onto* $\operatorname{coim} f$*, then we have* $f = f^\sharp \circ \rho$*, where* $f^\sharp \in \hom_D(\operatorname{coim} f, M')$ *is defined in 4 of Lemma 2, and the following commutative exact diagram shows that* ρ *is an isomorphism:*

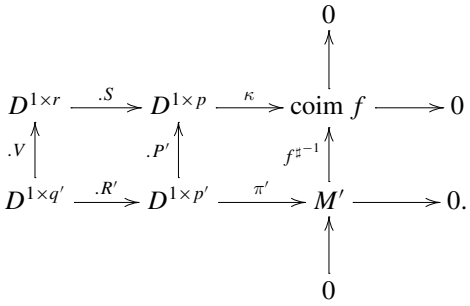


b. The matrix $(L^T \quad S_2^T)^T$ *admits a left inverse.*

(c) Surjective, i.e., $\mathrm{im}\, f = M'$*, if and only if* $(P^T \quad R'^T)^T$ *admits a left inverse. Then, the long exact sequence (2.11)* splits *(see, e.g., [24]), i.e., there exist four matrices* $P' \in D^{p' \times p}$, $Z' \in D^{p' \times q'}$, $U \in D^{p \times r}$, *and* $V \in D^{q' \times r}$ *such that:*

$$\begin{cases} P'\,P + Z'\,R' = I_{p'}, \\ P\,P' + U\,S = I_p, \\ P\,Z' - U\,T = 0, \\ R'\,P' - V\,S = 0, \\ R'\,Z' + V\,T = I_{q'}. \end{cases}$$

In this case, we have the following commutative exact diagram:



(d) An isomorphism, i.e., $M \cong M'$*, if and only if both matrices* $(L^T \quad S_2^T)^T$ *and* $(P^T \quad R'^T)^T$ *admit a left inverse. The inverse* f^{-1} *of* f *is then defined by*

$$\forall\, \lambda' \in D^{1 \times p'}, \quad f^{-1}(\pi'(\lambda')) := \pi(\lambda'\,P'),$$

where $P' \in D^{p' \times p}$ *is a matrix as defined in 3. Moreover, we have the following commutative exact diagram:*

$$\begin{array}{ccccccc} D^{1\times q} & \xrightarrow{.R} & D^{1\times p} & \xrightarrow{\pi} & M & \longrightarrow & 0 \\ {\scriptstyle .V\,F}\uparrow & & {\scriptstyle .P'}\uparrow & & {\scriptstyle f^{-1}}\uparrow & & \\ D^{1\times q'} & \xrightarrow{.R'} & D^{1\times p'} & \xrightarrow{\pi'} & M' & \longrightarrow & 0. \end{array}$$

Algorithms for checking whether or not a homomorphism of finitely presented left D-modules is injective, surjective, or defines an isomorphism (and if so, compute its inverse) are implemented in the OREMORPHISMS package [7].

Example 6 Let us check that the homomorphism f defined in Example 4 is an isomorphism by characterizing $\ker f$ and $\operatorname{coker} f$, and then let us explicitly compute its inverse f^{-1}. Using elimination techniques, we can first check that f is surjective, i.e., $\operatorname{coker} f = 0$, since $(P' \quad Z') := (0 \quad -1 \quad 0 \quad 1)$ is a left inverse of the matrix $(P^T \quad R'^T)^T$. We also have $\ker_D(.(P^T \quad R'^T)^T) = \operatorname{im}_D(.(S \quad -T))$, where:

$$S := \begin{pmatrix} 1 & 1 \\ 0 & \partial+1 \\ 0 & \delta^2\,(\delta+1)-1 \\ 0 & 0 \end{pmatrix}, \quad T := \begin{pmatrix} 0 & 1 \\ 0 & \partial \\ -\partial & \partial\,(\partial-1) \\ -\partial-1 & \partial^2-\delta^2\,(\delta+1) \end{pmatrix}.$$

Moreover, the identities of 3 of Lemma 3 are satisfied with:

$$U := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad V := \begin{pmatrix} 0 & -\partial+1 & 1 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}.$$

Let us now check that f is injective. We have $R = L\,S$ and $\ker_D(.S) = \operatorname{im}_D(.S_2)$, where:

$$L := \begin{pmatrix} \partial & -1 & 0 & 0 \\ -\delta^2\,(\delta+1) & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad S_2 := \begin{pmatrix} 0 & \delta^2\,(\delta+1)-1 & -\partial-1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The matrix $(L^T \quad S_2^T)^T$ admits the following left inverse defined by

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & \partial & 0 & 0 \\ 1 & 1 & \delta^2\,(\delta+1)-\partial & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

which shows that $\ker f = 0$ and proves that f is an isomorphism, i.e., $M \cong M'$. Hence, for every D-module $\mathcal{F}$, we get $\ker_{\mathcal{F}}(R.) \cong \ker_{\mathcal{F}}(R'.)$, i.e., there exists a 1-1 correspondence between the $\mathcal{F}$-solutions of (2.5) and the $\mathcal{F}$-solutions of (2.8) or, in other words, the linear DTD systems (2.5) and (2.8) are equivalent. More precisely, using 4 of Lemma 3, we obtain that $f^{-1} : M' \longrightarrow M$ is defined by

$f^{-1}(\pi'(\lambda')) := \pi(\lambda'\, P')$, where $P' := (0 \quad -1)$. In terms of behaviors, the following homomorphism

$$\begin{aligned} P'. : \ker_{\mathcal{F}}(R.) &\longrightarrow \ker_{\mathcal{F}}(R'.) \\ \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} &\longmapsto z(t) := -x_2(t), \end{aligned}$$

is the inverse of the homomorphism of behaviors $P.$ defined by (2.10).

2.4 Characterization of Isomorphic Modules

We characterize the existence of a left/right/two sided inverse of a homomorphism.

Lemma 4 *With the notations of Lemma 2, we have:*

(a) *f admits a right inverse $g \in \hom_D(M', M)$, i.e., $f \circ g = \mathrm{id}_{M'}$, or equivalently we have $M \cong \ker f \oplus M'$, if and only if there exist three matrices $P' \in D^{p' \times p}$, $Q' \in D^{q' \times q}$, and $Z' \in D^{p' \times q'}$ satisfying:*

$$R'\, P' = Q'\, R, \quad P'\, P + Z'\, R' = I_{p'}.$$

Then, for any matrix $R'_2 \in D^{r' \times q'}$ such that $\ker_D(.R') = \mathrm{im}_D(.R'_2)$, there exists $Z'_2 \in D^{q' \times r'}$ satisfying $Q'\, Q + R'\, Z' + Z'_2\, R'_2 = I_{q'}$.

(b) *f admits a left inverse $g \in \hom_D(M', M)$, i.e., $g \circ f = \mathrm{id}_M$, or equivalently we have $M' \cong M \oplus \mathrm{coker}\, f$, if and only if there exist three matrices $P' \in D^{p' \times p}$, $Q' \in D^{q' \times q}$ and $Z \in D^{p \times q}$ satisfying:*

$$R'\, P' = Q'\, R, \quad P\, P' + Z\, R = I_p.$$

Then, for any matrix $R_2 \in D^{r \times q}$ such that $\ker_D(.R) = \mathrm{im}_D(.R_2)$, there exists $Z_2 \in D^{q \times r}$ satisfying $Q\, Q' + R\, Z + Z_2\, R_2 = I_q$.

(c) *f is an isomorphism, and thus $M \cong M'$, if and only if there exist 4 matrices $P' \in D^{p' \times p}$, $Q' \in D^{q' \times q}$, $Z \in D^{p \times q}$, and $Z' \in D^{p' \times q'}$ satisfying:*

$$R'\, P' = Q'\, R, \quad P\, P' + Z\, R = I_p, \quad P'\, P + Z'\, R' = I_{p'}. \tag{2.12}$$

Then, for $R_2 \in D^{r \times q}$ (resp., $R'_2 \in D^{r' \times q'}$) such that $\ker_D(.R) = \mathrm{im}_D(.R_2)$ (resp., $\ker_D(.R') = \mathrm{im}_D(.R'_2)$), there exist matrices $Z_2 \in D^{q \times r}$, $Z'_2 \in D^{q' \times r'}$, $Y_2 \in D^{p' \times r}$, $Y'_2 \in D^{p \times r'}$ such that:

$$\begin{aligned} &Q\, Q' + R\, Z + Z_2\, R_2 = I_q, \quad && Q'\, Q + R'\, Z' + Z'_2\, R'_2 = I_{q'}, \\ &Z'\, Q' - P'\, Z = Y_2\, R_2, && P\, Z' - Z\, Q = Y'_2\, R'_2. \end{aligned} \tag{2.13}$$

Proof 1. The existence of $g \in \hom_D(M', M)$ is equivalent to the existence of two matrices $P' \in D^{p' \times p}$ and $Q' \in D^{q' \times q}$ such that $R'\,P' = Q'\,R$ (see 1 of Lemma 1). Composing the following two commutative exact diagrams

$$\begin{array}{ccccccccc}
D^{1\times q} & \xrightarrow{.R} & D^{1\times p} & \xrightarrow{\pi} & M & \longrightarrow & 0 \\
\downarrow{\scriptstyle .Q} & & \downarrow{\scriptstyle .P} & & \downarrow{\scriptstyle f} & & \\
D^{1\times q'} & \xrightarrow{.R'} & D^{1\times p'} & \xrightarrow{\pi'} & M' & \longrightarrow & 0,
\end{array}
\qquad
\begin{array}{ccccccccc}
D^{1\times q} & \xrightarrow{.R} & D^{1\times p} & \xrightarrow{\pi} & M & \longrightarrow & 0 \\
\uparrow{\scriptstyle .Q'} & & \uparrow{\scriptstyle .P'} & & \uparrow{\scriptstyle g} & & \\
D^{1\times q'} & \xrightarrow{.R'} & D^{1\times p'} & \xrightarrow{\pi'} & M' & \longrightarrow & 0,
\end{array}$$

and noting $\chi := \mathrm{id}_{M'} - f \circ g$, we get the following commutative exact diagram:

$$\begin{array}{ccccccc}
D^{1\times q'} & \xrightarrow{.R'} & D^{1\times p'} & \xrightarrow{\pi'} & M' & \longrightarrow & 0 \\
\uparrow{\scriptstyle .(I_{q'} - Q'\,Q)} & & \uparrow{\scriptstyle .(I_{p'} - P'\,P)} & & \uparrow{\scriptstyle \chi} & & \\
D^{1\times q'} & \xrightarrow{.R'} & D^{1\times p'} & \xrightarrow{\pi'} & M' & \longrightarrow & 0.
\end{array}$$

By 1.a of Lemma 3, $\chi = 0$ if and only if there exists a matrix $Z' \in D^{p' \times q'}$ such that $I_{p'} - P'\,P = Z'\,R'$, i.e., $P'\,P + Z'\,R' = I_{p'}$. According to 1.a of Lemma 3, there then exists a matrix $Z'_2 \in D^{q' \times r'}$ satisfying the relation $I_{q'} - Q'\,Q = R'\,Z' + Z'_2\,R'_2$, i.e., $Q'\,Q + R'\,Z' + Z'_2\,R'_2 = I_{q'}$, where $R'_2 \in D^{r' \times q'}$ is such that $\ker_D(.R') = \mathrm{im}_D(.R'_2)$. Finally, $M \cong \ker f \oplus M'$ is well-known to be equivalent to the *splitting* of the following short exact sequence

$$0 \longrightarrow \ker f \longrightarrow M \underset{g}{\overset{f}{\rightleftarrows}} M' \longrightarrow 0,$$

(see, e.g., [24]), i.e., it is equivalent to the existence of a left inverse g of f.

2 can be proved similarly as 1. The first points of 3 are direct consequences of 1 and 2. Finally, let us prove the third and fourth identities of (2.13). Using the identity $Q'\,R = R'\,P'$ and (2.12), we have

$$(Z'\,Q' - P'\,Z)\,R = (Z'\,R')\,P' - P'\,(Z\,R) = (I_{p'} - P'\,P)\,P' - P'\,(I_p - P\,P') = 0,$$

which yields $\mathrm{im}_D(.(Z'\,Q' - P'\,Z)) \subseteq \ker_D(.R) = \mathrm{im}_D(.R_2)$ and shows that there exists $Y_2 \in D^{p' \times r}$ such that $Z'\,Q' - P'\,Z = Y_2\,R_2$. Similarly, using $Q\,R' = R\,P$ and (2.12), we have

$$(P\,Z' - Z\,Q)\,R' = P\,(Z'\,R') - (Z\,R)\,P = P\,(I_{p'} - P'\,P) - (I_p - P\,P')\,P = 0,$$

which yields $\mathrm{im}_D(.(P\,Z' - Z\,Q)) \subseteq \ker_D(.R') = \mathrm{im}_D(.R'_2)$ and shows that there exists $Y'_2 \in D^{p \times r'}$ such that $P\,Z' - Z\,Q = Y'_2\,R'_2$.

Remark 2 We note that the existence of a right (resp., left) inverse g of $f \in \hom_D(M, M')$ implies that f is surjective (resp., injective) since we then have $m' = f(g(m'))$ (resp., $g(f(m)) = m$) for all $m' \in M'$ (resp., $m \in M$).

Example 7 We can check again that the D-module M and M' defined in Examples 3 and 4 are isomorphic by considering the matrices P and Q defined in Example 4 and the matrix P' defined in Example 6. Then, we have:

$$Q' := \begin{pmatrix} \partial & 1 & -\partial^2 + \delta^2 (\delta + 1) \\ 1 & 0 & -\partial \end{pmatrix}, \quad Z := \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & \partial \end{pmatrix}, \quad Z' := \begin{pmatrix} 0 & 1 \end{pmatrix}.$$

We can check that we have $Q\,Q' + R\,Z = I_3$, $Q'\,Q + R'\,Z' = I_2$, $Z'\,Q' - P'\,Z = 0$, and $P\,Z' - Z\,Q = 0$, i.e., $Z_2 = 0$, $Z'_2 = 0$, $Y_2 = 0$, and $Y'_2 = 0$.

Let us introduce a few definitions.

Definition 1 (a) We denote the *general linear group of degree* r over D by:

$$\mathrm{GL}_r(D) := \{U \in D^{r \times r} \mid \exists\, V \in D^{r \times r} : U\,V = V\,U = I_r\}.$$

(b) Two matrices $R,\ R' \in D^{q \times p}$ are said to be *equivalent* if there exist $U \in \mathrm{GL}_q(D)$ and $V \in \mathrm{GL}_p(D)$ such that:

$$R' = U\,R\,V.$$

In module theory, *Fitting's theorem* states that two finitely presented modules are isomorphic if and only if their presentation matrices R and R' can be inflated by zero and identity matrices in a way that the new matrices are equivalent. More precisely, Fitting's theorem states that $M := D^{1 \times p}/(D^{1 \times q}\,R) \cong M' := D^{1 \times p'}/(D^{1 \times q'}\,R')$ if and only if the following two matrices

$$L := \begin{pmatrix} R & 0 \\ 0 & I_{p'} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad L' := \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ I_p & 0 \\ 0 & R' \end{pmatrix} \in D^{(q+p'+p+q') \times (p+p')},$$

are equivalent. For a constructive version of Fitting's theorem, see [8].

In this paper, we give another characterization of isomorphic finitely presented modules in terms of inflations of their presentation matrices.

Theorem 1 *Let $R \in D^{q \times p}$ and $R' \in D^{q' \times p'}$ be two matrices with entries in a noetherian ring D. Then, the following assertions are equivalent:*

(a) $M := D^{1 \times p}/(D^{1 \times q}\,R) \cong M' := D^{1 \times p'}/(D^{1 \times q'}\,R')$.
(b) There exist 12 matrices

$$P \in D^{p\times p'},\ Q \in D^{q\times q'},\ P' \in D^{p'\times p},\ Q' \in D^{q'\times q},\ Z \in D^{p\times q},\ Z' \in D^{p'\times q'},$$
$$Z_2 \in D^{q\times r},\ Y_2 \in D^{p'\times r},\ Y_2' \in D^{p\times r'},\ Z_2' \in D^{q'\times r'},\ R_2 \in D^{r\times q},\ R_2' \in D^{r'\times q'}$$

satisfying the following two identities

$$\begin{pmatrix} R & -Q \\ P' & Z' \end{pmatrix} \begin{pmatrix} Z & P \\ -Q' & R' \end{pmatrix} + \begin{pmatrix} Z_2 \\ Y_2 \end{pmatrix} (R_2 \quad 0) = I_{q+p'}, \tag{2.14}$$

$$\begin{pmatrix} Z & P \\ -Q' & R' \end{pmatrix} \begin{pmatrix} R & -Q \\ P' & Z' \end{pmatrix} + \begin{pmatrix} Y_2' \\ Z_2' \end{pmatrix} \begin{pmatrix} 0 & R_2' \end{pmatrix} = I_{p+q'}, \tag{2.15}$$

where the matrices $R_2 \in D^{r\times q}$ and $R_2' \in D^{r'\times q'}$ are such that:

$$\ker_D(.R) = \mathrm{im}_D(.R_2), \quad \ker_D(.R') = \mathrm{im}_D(.R_2').$$

Proof By 3 of Lemma 4, $M \cong M'$ if and only if there exist $P' \in D^{p'\times p}$, $Q' \in D^{q'\times q}$, $Z \in D^{p\times q}$, and $Z' \in D^{p'\times q'}$ satisfying (2.12). Using (2.12) and (2.13), we get

$$\begin{aligned}
\begin{pmatrix} R & -Q \\ P' & Z' \end{pmatrix} \begin{pmatrix} Z & P \\ -Q' & R' \end{pmatrix} &= \begin{pmatrix} I_q - Z_2\, R_2 & 0 \\ -Y_2\, R_2 & I_{p'} \end{pmatrix} \\
&= I_{q+p'} - \begin{pmatrix} Z_2 \\ Y_2 \end{pmatrix} (R_2 \quad 0), \\
\begin{pmatrix} Z & P \\ -Q' & R' \end{pmatrix} \begin{pmatrix} R & -Q \\ P' & Z' \end{pmatrix} &= \begin{pmatrix} I_p & -Y_2'\, R_2' \\ 0 & I_{q'} - Z_2'\, R_2' \end{pmatrix} \\
&= I_{p+q'} - \begin{pmatrix} Y_2' \\ Z_2' \end{pmatrix} (0 \quad R_2'),
\end{aligned}$$

i.e., (2.14) and (2.15) hold. Conversely, if (2.14) and (2.15) hold, then we have $R\,P = Q\,R'$, $R'\,P' = Q'\,R$, $P\,P' + Z\,R = I_p$, and $P'\,P + Z'\,R' = I_{p'}$, which shows that $M \cong M'$ by 3 of Lemma 4.

Example 8 We consider again Examples 3 and 4. We first can check that we have $\ker_D(.R) = \mathrm{im}_D(.R_2)$ and $\ker_D(.R') = \mathrm{im}_D(.R_2')$, where:

$$\begin{cases} R_2 := \begin{pmatrix} \partial + \delta^2\,(\delta+1) & \partial + 1 & -\partial^2 + \delta^2\,(\delta+1) \end{pmatrix}, \\ R_2' := \begin{pmatrix} \partial + 1 & -\partial^2 + \delta^2\,(\delta+1) \end{pmatrix}. \end{cases}$$

Hence, R and R' are not full row rank matrices. Using Example 7, Theorem 1 yields:

$$\begin{pmatrix} \partial & -1 & 0 & 0 \\ -\delta^2(\delta+1) & \partial & -1 & 0 \\ 1 & 1 & 0 & -1 \\ 0 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 1 \\ -1 & 0 & \partial & \partial \\ -\partial & -1 & \partial^2 - \delta^2(\delta+1) & \partial^2 - \delta^2(\delta+1) \\ -1 & 0 & \partial & \partial+1 \end{pmatrix} = I_4.$$

Note that (2.15) is a consequence of the above identity since D is a commutative ring.

Example 9 We consider the following linear system of PDEs

$$\begin{cases} \dfrac{\partial^2 y(x_1,x_2)}{\partial x_1^2} - x_2 \dfrac{\partial^2 y(x_1,x_2)}{\partial x_2^2} - \dfrac{\beta}{2} \dfrac{\partial y(x_1,x_2)}{\partial x_2} = 0, \\ 2 \dfrac{\partial^2 y(x_1,x_2)}{\partial x_1 \partial x_2} + x_1 \dfrac{\partial^2 y(x_1,x_2)}{\partial x_2^2} = 0, \end{cases} \tag{2.16}$$

which is studied in probability theory [2]. Let $D := \mathbb{Q}(\beta)(x_1, x_2)\langle \partial_1, \partial_2 \rangle$ be the noncommutative ring of PD operators in $\partial_1 := \frac{\partial}{\partial x_1}$ and $\partial_2 := \frac{\partial}{\partial x_2}$ with coefficients in the field $\mathbb{Q}(\beta, x_1, x_2)$ of rational functions in x_1, x_2, and β. The ring D is called the *Weyl algebra* in two variables and it is usually denoted by $B_2(\mathbb{Q}(\beta))$. Let us consider the matrix of PD operators associated with (2.16)

$$R := \begin{pmatrix} \partial_1^2 - x_2 \partial_2^2 - \frac{\beta}{2} \partial_2 \\ 2 \partial_1 \partial_2 + x_1 \partial_2^2 \end{pmatrix} \in D^{2\times 1},$$

and the left D-module $M := D/(D^{1\times 2} R)$ finitely presented by R. It can be shown that M is *D-finite*, namely, M has a $\mathbb{Q}(\beta, x_1, x_2)$-finite dimensional vector space structure (see, e.g., [5]), and thus it can be written as an *integrable connection*, i.e., we can find a first-order realization of (2.16) (see, e.g., [6]). We can show that (2.16), i.e., $R\,\eta = 0$, is equivalent to $R'\,\eta' = 0$, where

$$R' := \begin{pmatrix} \partial_1 & 0 & -1 & 0 \\ 0 & \partial_1 & 0 & \frac{1}{2} x_1 \\ 0 & -\frac{\beta}{2} & \partial_1 & -x_2 \\ 0 & 0 & 0 & \partial_1 + \frac{(\beta+3)x_1}{{x_1}^2 - 4x_2} \\ \partial_2 & -1 & 0 & 0 \\ 0 & \partial_2 & 0 & -1 \\ 0 & 0 & \partial_2 & \frac{1}{2} x_1 \\ 0 & 0 & 0 & \partial_2 - \frac{2\beta+6}{{x_1}^2 - 4x_2} \end{pmatrix} \in D^{8\times 4},$$

i.e., we have $M \cong M' := D^{1\times 4}/(D^{1\times 8} R')$. This first-order realization can be computed by means of the OREMODULES package [4]. Let us compute the matrices appearing in (2.14) and (2.15). By construction, the isomorphism $f \in \hom_D(M, M')$

is defined by $f(\pi(\lambda)) := \pi'(\lambda\, P)$, where $\pi : D \longrightarrow M$ (resp., $\pi' : D^{1\times 4} \longrightarrow M'$) is the canonical projection, $\lambda \in D$, and $P := (1 \;\; 0 \;\; 0 \;\; 0)$ satisfies (2.7), where the matrix Q is defined by:

$$Q := \begin{pmatrix} \partial_1 & 0 & 1 & 0 & -x_2\,\partial_2 - \frac{\beta}{2} & -x_2 & 0 & 0 \\ 0 & 2 & 0 & 0 & x_1\,\partial_2 + 2\,\partial_1 & x_1 & 0 & 0 \end{pmatrix} \in D^{2\times 8}.$$

Since f is surjective, the matrix $(P^T \;\; R'^T)^T$ admits a left inverse $(P' \;\; Z')$, where $P' := \left(1 \;\; \partial_2 \;\; \partial_1 \;\; \partial_2^2\right)^T$ and:

$$Z' := \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\partial_2 & -1 & 0 & 0 \end{pmatrix} \in D^{4\times 8}.$$

Moreover, we have $R'\,P' = Q'\,R$, where the matrix Q' is defined by

$$Q' := \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{2} \\ 1 & 0 \\ -\frac{2\,x_1\,\partial_2}{{x_1}^2 - 4\,x_2} & \frac{-2\,x_2\,\partial_2 + x_1\,\partial_1}{{x_1}^2 - 4\,x_2} \\ 0 & 0 \\ 0 & 0 \\ 0 & \frac{1}{2} \\ \frac{4\,\partial_2}{{x_1}^2 - 4\,x_2} & \frac{-2\,\partial_1 + x_1\,\partial_2}{{x_1}^2 - 4\,x_2} \end{pmatrix} \in D^{8\times 2},$$

and $f^{-1}(\pi'(\lambda')) := \pi(\lambda'\,P')$ for all $\lambda' \in D^{1\times 4}$.

Using (2.12), we can check that $P\,P' = 1$, i.e., $Z = 0$, and $Q\,Q' = I_2$, i.e., $Z_2 = 0$ (see (2.13)). We also have $\ker_D(.R) = \mathrm{im}_D(.R_2)$ and $\ker_D(.R') = \mathrm{im}_D(.R'_2)$, where

$$R_2 := (2\,x_1\,\partial_2^2 + 4\,\partial_1\,\partial_2 \;\; -2\,\partial_1^2 + 2\,x_2\,\partial_2^2 + (4+\beta)\,\partial_2),$$

$$R'_2 := \begin{pmatrix} -\partial_2 & 1 & 0 & 0 & \partial_1 & 0 & -1 & 0 \\ 0 & -4\,x_2\,\partial_2 & -2\,x_1\,\partial_2 & 4\,x_2 - {x_1}^2 & 0 & -x_1\,\beta + 4\,x_2\,\partial_1 & 2\,\partial_1\,x_1 & 0 \\ 0 & -2\,x_1\,\partial_2 & -4\,\partial_2 & 0 & 0 & 2\,x_1\,\partial_1 - 2\,\beta & 4\,\partial_1 & -4\,x_2 + {x_1}^2 \\ 0 & 2\,(\beta+1)\,\partial_2 & 0 & (4\,x_2 - {x_1}^2)\,\partial_2 + 4 & 0 & -2\,(\beta+1)\,\partial_1 & 0 & ({x_1}^2 - 4\,x_2)\,\partial_1 + 2\,x_1 \end{pmatrix},$$

and, using (2.13), we get:

$$Z_2' = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\frac{2x_1\partial_1}{{x_1}^2-4x_2} & -\frac{1}{{x_1}^2-4x_2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ \frac{4\partial_1}{{x_1}^2-4x_2} & 0 & \frac{1}{x_1^2-4x_2} & 0 \end{pmatrix} \in D^{8\times 4}.$$

Using (2.13) again, we get $Y_2 = 0$ and $Y_2' = 0$, and thus we finally obtain:

$$\begin{pmatrix} R & -Q \\ P' & Z' \end{pmatrix} \begin{pmatrix} Z & P \\ -Q' & R' \end{pmatrix} = I_6, \begin{pmatrix} Z & P \\ -Q' & R' \end{pmatrix} \begin{pmatrix} R & -Q \\ P' & Z' \end{pmatrix} + \begin{pmatrix} 0 \\ Z_2' \end{pmatrix} \begin{pmatrix} 0 & R_2' \end{pmatrix} = I_9.$$

For applications of D-finite multidimensional systems in control theory, see [20].

A consequence of Theorem 1 connects isomorphisms of finitely presented modules to the *unimodular completion problem*, and therefore to the so-called *Serre's reduction problem* studied in [1, 9, 11] (see Sects. 2.1 and 2.5).

Corollary 2 *With the notations and the assumptions of Theorem 1, let us assume that we have* $q + p' = p + q'$.

(a) Then, we have:

$$\begin{pmatrix} R & -Q \\ P' & Z' \end{pmatrix} \begin{pmatrix} Z & P \\ -Q' & R' \end{pmatrix} = I_{q+p'} \iff \begin{pmatrix} Z & P \\ -Q' & R' \end{pmatrix} \begin{pmatrix} R & -Q \\ P' & Z' \end{pmatrix} = I_{p+q'}.$$

(b) If R or R' have full row rank, then the fact that $M \cong M'$ is equivalent to the existence of matrices $P \in D^{p\times p'}$, $Q \in D^{q\times q'}$, $P' \in D^{p'\times p}$, $Q' \in D^{q'\times q}$, $Z \in D^{p\times q}$, and $Z' \in D^{p'\times q'}$ such that:

$$\begin{pmatrix} R & -Q \\ P' & Z' \end{pmatrix} \begin{pmatrix} Z & P \\ -Q' & R' \end{pmatrix} = I_{q+p'}.$$

Proof 1 is a consequence of $q + p' = p + q'$ and the fact that D is a noetherian ring, and thus a *stably finite ring*, namely, a ring for which $U\,V = I_r$ for two matrices U, $V \in D^{r\times r}$ yields $V\,U = I_r$ (see, e.g., [16, 24]). Note that a commutative ring is stably finite since $U\,V = I_r$ implies that $\det U$ is a unit of D.

2 is a direct consequence of 1 and Theorem 1 with $R_2 = 0$ or $R_2' = 0$.

Example 10 Let R, $R' \in D^{q\times p}$ be two equivalent matrices, i.e., they satisfy

$$R' = Q^{-1}\, R\, P,$$

for certain $P \in \mathrm{GL}_p(D)$ and $Q \in \mathrm{GL}_q(D)$. If we note $M := D^{1\times p}/(D^{1\times q}\, R)$ and $M' := D^{1\times p}/(D^{1\times q}\, R')$, then $f \in \hom_D(M, M')$, defined by $f(\pi(\lambda)) := \pi'(\lambda\, P)$ for all $\lambda \in D^{1\times p}$, is an isomorphism and $f^{-1} \in \hom_D(M', M)$ is defined by $f^{-1}(\pi'(\lambda')) := \pi(\lambda'\, P^{-1})$ for all $\lambda' \in D^{1\times p}$, where $\pi : D^{1\times p} \longrightarrow M$ (resp., $\pi' : D^{1\times p'} \longrightarrow M'$) is the canonical projection. This result can be proved again since

$$\begin{pmatrix} R & -Q \\ P^{-1} & 0 \end{pmatrix} \begin{pmatrix} 0 & P \\ -Q^{-1} & R' \end{pmatrix} = I_{q+p}, \quad \begin{pmatrix} 0 & P \\ -Q^{-1} & R' \end{pmatrix} \begin{pmatrix} R & -Q \\ P^{-1} & 0 \end{pmatrix} = I_{p+q}, \tag{2.17}$$

and Theorem 1 then yields again the isomorphism $M \cong M'$.

Remark 3 If R is a full row rank matrix, then it is known that M is a *free* left D-module of rank $p - q$, i.e., $M \cong D^{1\times(p-q)}$, if and only if there exist $P' \in D^{(p-q)\times p}$, $P \in D^{p\times(p-q)}$, and $Z \in D^{p\times q}$ such that

$$\begin{pmatrix} R \\ P' \end{pmatrix} (Z \quad P) = I_p,$$

i.e., if and only if there exists $P' \in D^{(p-q)\times p}$ such that $(R^T \quad P'^T)^T \in \mathrm{GL}_p(D)$. For more details, see [22]. This result corresponds to the extreme case of Corollary 2 where $q' = 0$ (and thus, $p' = p - q$) and $M' = D^{1\times(p-q)}$, i.e., to the case of the following commutative exact diagram:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & D^{1\times q} & \xrightarrow{.R} & D^{1\times p} & \xrightarrow{\pi} & M & \longrightarrow & 0 \\ & & \downarrow & & \downarrow{\scriptstyle .P} & & \downarrow{\scriptstyle f} & & \\ & & 0 & \longrightarrow & D^{1\times(p-q)} & \xrightarrow{\pi'} & M' & \longrightarrow & 0. \end{array}$$

In particular, we have $P'\, P = I_{p-q}$, $P\, P' + Z\, R = I_p$, and $R\, P = 0$. We then get $R - R\, Z\, R = (R\, P)\, P' = 0$, i.e., $(I_q - R\, Z)\, R = 0$ which yields $R\, Z = I_q$ since R has full row rank. Then, we have $P\, P'\, Z = Z - Z\, (R\, Z) = 0$, and thus $(P'\, P)\, (P'\, Z) = 0$, i.e., $P'\, Z = 0$, which shows again that we have the following split exact sequence (see, e.g., [22, 24]):

$$0 \longrightarrow D^{1\times q} \underset{.Z}{\overset{.R}{\rightleftarrows}} D^{1\times p} \underset{.P'}{\overset{.P}{\rightleftarrows}} D^{1\times(p-q)} \longrightarrow 0.$$

Let us consider again Theorem 1 and the following two short exact sequences

$$0 \longrightarrow \mathrm{im}_D(.R) \xrightarrow{i} D^{1\times p} \xrightarrow{\pi} M \longrightarrow 0,$$
$$0 \longrightarrow \mathrm{im}_D(.R') \xrightarrow{i'} D^{1\times p'} \xrightarrow{\pi'} M' \longrightarrow 0,$$

where i (resp., i') denotes the canonical injection into $D^{1\times p}$ (resp., $D^{1\times p'}$).

In module theory, *Schanuel's lemma* (see, e.g., [24]) asserts that $M \cong M'$ yields:

$$\mathrm{im}_D(.R) \oplus D^{1\times p'} \cong \mathrm{im}_D(.R') \oplus D^{1\times p}. \tag{2.18}$$

As a consequence of Theorem 1, we obtain a constructive proof of Schanuel's lemma in which the isomorphism (2.18) and its inverse are explicitly described.

Corollary 3 *With the notations and the assumptions of Theorem 1, if we note*

$$U := \begin{pmatrix} I_p & -P \\ P' & I_{p'} - P'\,P \end{pmatrix} \in \mathrm{GL}_{p+p'}(D), \quad U^{-1} = \begin{pmatrix} I_p - P\,P' & P \\ -P' & I_{p'} \end{pmatrix},$$

then the following homomorphism of left D-modules

$$\begin{aligned} u : D^{1\times q}\,R \oplus D^{1\times p'} &\longrightarrow D^{1\times p} \oplus D^{1\times q'}\,R' \\ (\mu\,R \quad \lambda') &\longmapsto (\mu\,R \quad \lambda')\,U, \end{aligned} \tag{2.19}$$

is an isomorphism and its inverse u^{-1} is defined by:

$$\begin{aligned} u^{-1} : D^{1\times p} \oplus D^{1\times q'}\,R' &\longrightarrow D^{1\times q}\,R \oplus D^{1\times p'} \\ (\lambda \quad \mu'\,R') &\longmapsto (\lambda \quad \mu'\,R')\,U^{-1}. \end{aligned} \tag{2.20}$$

Proof Let $f \in \hom_D(M, M')$ be an isomorphism. With the notations of Theorem 1 and $P_2 := Q$, we then have the following commutative exact diagram:

$$\begin{array}{ccccccccc} D^{1\times r} & \xrightarrow{.R_2} & D^{1\times q} & \xrightarrow{.R} & D^{1\times p} & \xrightarrow{\pi} & M & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle .P_2} & & \downarrow{\scriptstyle .P} & & \downarrow{\scriptstyle f} & & \\ D^{1\times r'} & \xrightarrow{.R'_2} & D^{1\times q'} & \xrightarrow{.R'} & D^{1\times p'} & \xrightarrow{\pi'} & M' & \longrightarrow & 0. \end{array}$$

Using $R_2\,R = 0$, $R\,P = P_2\,R'$ yields $(R_2\,P_2)\,R' = (R_2\,R)\,P = 0$, i.e., $\mathrm{im}_D(.(R_2\,P_2)) \subseteq \ker_D(.R') = \mathrm{im}_D(.R'_2)$, and thus there exists $P_3 \in D^{r\times r'}$ such that $R_2\,P_2 = P_3\,R'_2$. Similarly, with the notation $P'_2 := Q'$, there exists $P'_3 \in D^{r'\times r}$ such that $R'_2\,P'_2 = P'_3\,R_2$ and we get the following commutative exact diagram:

$$\begin{array}{ccccccccc}
D^{1\times r} & \xrightarrow{.R_2} & D^{1\times q} & \xrightarrow{.R} & D^{1\times p} & \xrightarrow{\pi} & M & \longrightarrow & 0 \\
\uparrow{\scriptstyle .P_3'} & & \uparrow{\scriptstyle .P_2'} & & \uparrow{\scriptstyle .P'} & & \uparrow{\scriptstyle f^{-1}} & & \\
D^{1\times r'} & \xrightarrow{.R_2'} & D^{1\times q'} & \xrightarrow{.R'} & D^{1\times p'} & \xrightarrow{\pi'} & M' & \longrightarrow & 0.
\end{array}$$

Now, if we note

$$V := \begin{pmatrix} R & -P_2 \\ P' & Z' \end{pmatrix}, \quad V' := \begin{pmatrix} Z & P \\ -P_2' & R' \end{pmatrix},$$

then we have $(R_2 \quad 0)\, V = -P_3\, (0 \quad R_2')$ and $(0 \quad R_2')\, V' = -P_3'\, (R_2 \quad 0)$. Hence, if we note $L := D^{1\times(q+p')}/(D^{1\times r}\ (R_2 \quad 0))$ and $L' := D^{1\times(p+q')}/(D^{1\times r'}\ (0 \quad R_2'))$, then we have the following two commutative exact diagrams

$$\begin{array}{ccccccc}
D^{1\times r} & \xrightarrow{.(R_2\ 0)} & D^{1\times(q+p')} & \xrightarrow{\kappa} & L & \longrightarrow & 0 \\
\downarrow{\scriptstyle .-P_3} & & \downarrow{\scriptstyle .V} & & \downarrow{\scriptstyle g} & & \\
D^{1\times r'} & \xrightarrow{.(0\ R_2')} & D^{1\times(p+q')} & \xrightarrow{\kappa'} & L' & \longrightarrow & 0,
\end{array}
\qquad
\begin{array}{ccccccc}
D^{1\times r} & \xrightarrow{.(R_2\ 0)} & D^{1\times(q+p')} & \xrightarrow{\kappa} & L & \longrightarrow & 0 \\
\uparrow{\scriptstyle .-P_3'} & & \uparrow{\scriptstyle .V'} & & \uparrow{\scriptstyle h} & & \\
D^{1\times r'} & \xrightarrow{.(0\ R_2')} & D^{1\times(p+q')} & \xrightarrow{\kappa'} & L' & \longrightarrow & 0,
\end{array}$$

where $g \in \hom_D(L, L')$ and $h \in \hom_D(L', L)$ are respectively defined by:

$$\begin{aligned}
g : L &\longrightarrow L' \\
\kappa((\mu \quad \lambda')) &\longmapsto \kappa'((\mu\, R + \lambda'\, P' \quad -\mu\, P_2 + \lambda'\, Z')), \\
h : L' &\longrightarrow L \\
\kappa'((\lambda \quad \mu')) &\longmapsto \kappa((\lambda\, Z - \mu'\, P_2' \quad \lambda\, P + \mu'\, R')).
\end{aligned}$$

Then, (2.14) and (2.15) show that $h \circ g = \mathrm{id}_L$ and $g \circ h = \mathrm{id}_{L'}$, i.e., g is an isomorphism, $h = g^{-1}$, and $L' \cong L$. Now, note that we have $\mathrm{coker}_D(.R_2) \cong \mathrm{im}_D(.R)$, $\mathrm{coker}_D(.R_2') \cong \mathrm{im}_D(.R')$, $L \cong \mathrm{im}_D(.R) \oplus D^{1\times p'}$ and $L' \cong D^{1\times p} \oplus \mathrm{im}_D(.R')$, where the last two isomorphisms are defined by:

$$\begin{aligned}
L &\xrightarrow{\alpha} \mathrm{im}_D(.R) \oplus D^{1\times p'} & \qquad L' &\xrightarrow{\beta} D^{1\times p} \oplus \mathrm{im}_D(.R') \\
\kappa((\mu \quad \lambda')) &\longmapsto (\mu\, R \quad \lambda'), & \kappa'((\lambda \quad \mu')) &\longmapsto (\lambda \quad \mu'\, R').
\end{aligned}$$

The isomorphisms $u := \beta \circ g \circ \alpha^{-1}$ and $u^{-1} = \alpha \circ h \circ \beta^{-1}$ are then defined by:

$$\begin{aligned}
\mathrm{im}_D(.R) \oplus D^{1\times p'} &\xrightarrow{u} D^{1\times p} \oplus \mathrm{im}_D(.R') \\
(\mu\, R \quad \lambda') &\longmapsto (\mu\, R + \lambda'\, P' \quad (-\mu\, P_2 + \lambda'\, Z')\, R'), \\
D^{1\times p} \oplus \mathrm{im}_D(.R') &\xrightarrow{u^{-1}} \mathrm{im}_D(.R) \oplus D^{1\times p'} \\
(\lambda \quad \mu'\, R') &\longmapsto ((\lambda\, Z - \mu'\, P_2')\, R \quad \lambda\, P + \mu'\, R').
\end{aligned}$$

Using $P_2\,R' = R\,P$, (2.12), and $P_2'\,R = R'\,P'$, we obtain

$$(-\mu\,P_2 + \lambda'\,Z')\,R' = -(\mu\,R)\,P + \lambda'\,(I_{p'} - P'\,P),$$
$$(\lambda\,Z - \mu'\,P_2')\,R = \lambda\,(I_p - P\,P') - (\mu'\,R')\,P',$$

which finally yields (2.19) and (2.20).

2.5 The Unimodular Completion Problem

The *unimodular completion problem* consists in studying the possibility to inflate a matrix $R_1 \in D^{q\times p}$ into a unimodular $V \in \mathrm{GL}_{q+t}(D)$, where $q + t \geqslant p$. The next theorem shows that a solution to this problem induces different isomorphisms between the modules finitely presented by the matrices appearing in the inflations.

Theorem 2 *Let p, q, s, $t \in \mathbb{N}$ satisfy $q + t = p + s$ and $R_1 \in D^{q\times p}$, $R_2 \in D^{q\times s}$, $Q_1 \in D^{p\times t}$, $Q_2 \in D^{s\times t}$, $S_1 \in D^{p\times q}$, $S_2 \in D^{s\times q}$, $T_1 \in D^{t\times p}$, and $T_2 \in D^{t\times s}$ matrices such that:*

$$\begin{pmatrix} R_1 & R_2 \\ T_1 & T_2 \end{pmatrix} \begin{pmatrix} S_1 & Q_1 \\ S_2 & Q_2 \end{pmatrix} = I_{q+t}. \tag{2.21}$$

Then, we have:

$$\begin{cases} \mathrm{coker}_D(.R_1) \cong \mathrm{coker}_D(.Q_2), \\ \mathrm{coker}_D(.S_1) \cong \mathrm{coker}_D(.T_2), \\ \mathrm{coker}_D(.Q_1) \cong \mathrm{coker}_D(.R_2), \\ \mathrm{coker}_D(.T_1) \cong \mathrm{coker}_D(.S_2), \end{cases} \quad \begin{cases} \ker_D(.R_1) \cong \ker_D(.Q_2), \\ \ker_D(.S_1) \cong \ker_D(.T_2), \\ \ker_D(.Q_1) \cong \ker_D(.R_2), \\ \ker_D(.T_1) \cong \ker_D(.S_2). \end{cases} \tag{2.22}$$

Right D-module analogs of (2.22) hold, i.e., we have:

$$\mathrm{coker}_D(R_1.) \cong \mathrm{coker}_D(Q_2.), \quad \ker_D(R_1.) \cong \ker_D(Q_2.), \ldots$$

Proof By 1 of Corollary 2, the identity (2.21) yields the following identity:

$$\begin{pmatrix} S_1 & Q_1 \\ S_2 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 & R_2 \\ T_1 & T_2 \end{pmatrix} = I_{p+s}. \tag{2.23}$$

From (2.21), we get $R_1\,Q_1 = -R_2\,Q_2$ which yields the commutative exact diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \ker_D(.R_1) & \longrightarrow & D^{1\times q} & \xrightarrow{.R_1} & D^{1\times p} & \xrightarrow{\pi_1} & \operatorname{coker}_D(.R_1) \longrightarrow 0 \\
 & & \downarrow{\scriptstyle \alpha'_1} & & \downarrow{\scriptstyle .-R_2} & & \downarrow{\scriptstyle .Q_1} & & \downarrow{\scriptstyle \alpha_1} \\
0 & \longrightarrow & \ker_D(.Q_2) & \longrightarrow & D^{1\times s} & \xrightarrow{.Q_2} & D^{1\times t} & \xrightarrow{\kappa_2} & \operatorname{coker}_D(.Q_2) \longrightarrow 0,
\end{array}$$

where α_1 and α'_1 are respectively defined by:

$$\begin{array}{rlrl}
\alpha_1 : \operatorname{coker}_D(.R_1) & \longrightarrow \operatorname{coker}_D(.Q_2) & \alpha'_1 : \ker_D(.R_1) & \longrightarrow \ker_D(.Q_2) \\
\pi_1(\lambda_1) & \longmapsto \kappa_2(\lambda_1\, Q_1), & \mu_1 & \longmapsto -\mu_1\, R_2.
\end{array} \tag{2.24}$$

Similarly, from (2.23), we get $Q_2\, T_1 = -S_2\, R_1$ which yields the following commutative exact diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \ker_D(.Q_2) & \longrightarrow & D^{1\times s} & \xrightarrow{.Q_2} & D^{1\times t} & \xrightarrow{\kappa_2} & \operatorname{coker}_D(.Q_2) \longrightarrow 0 \\
 & & \downarrow{\scriptstyle \alpha'_2} & & \downarrow{\scriptstyle .-S_2} & & \downarrow{\scriptstyle .T_1} & & \downarrow{\scriptstyle \alpha_2} \\
0 & \longrightarrow & \ker_D(.R_1) & \longrightarrow & D^{1\times q} & \xrightarrow{.R_1} & D^{1\times p} & \xrightarrow{\pi_1} & \operatorname{coker}_D(.R_1) \longrightarrow 0,
\end{array}$$

where α_2 and α'_2 are respectively defined by:

$$\begin{array}{rlrl}
\alpha_2 : \operatorname{coker}_D(.Q_2) & \longrightarrow \operatorname{coker}_D(.R_1) & \alpha'_2 : \ker_D(.Q_2) & \longrightarrow \ker_D(.R_1) \\
\kappa_2(\nu_2) & \longmapsto \pi_1(\nu_2\, T_1), & \theta_2 & \longmapsto -\theta_2\, S_2.
\end{array} \tag{2.25}$$

Using (2.21) and (2.23), we get $T_1\, Q_1 = I_t - T_2\, Q_2$ and $Q_1\, T_1 = I_p - S_1\, R_1$, which yields

$$\left\{
\begin{array}{l}
(\alpha_1 \circ \alpha_2)(\kappa_2(\nu_2)) = \kappa_2(\nu_2\, T_1\, Q_1) = \kappa_2(\nu_2) - \kappa_2((\nu_2\, T_2)\, Q_2) = \kappa_2(\nu_2), \\
(\alpha_2 \circ \alpha_1)(\pi_1(\lambda_1)) = \pi_1(\lambda_1\, Q_1\, T_1) = \pi_1(\lambda_1) - \pi_1((\lambda_1\, S_1)\, R_1)) = \pi_1(\lambda_1),
\end{array}
\right.$$

and shows that α_1 is an isomorphism, $\operatorname{coker}_D(.Q_2) \cong \operatorname{coker}_D(.R_1)$, and $\alpha_2 = \alpha_1^{-1}$.

Now, using (2.23) and (2.21), we get $S_2\, R_2 = I_s - Q_2\, T_2$ and $R_2\, S_2 = I_q - R_1\, S_1$, which yields

$$\left\{
\begin{array}{l}
(\alpha'_1 \circ \alpha'_2)(\theta_2) = \theta_2\, (S_2\, R_2) = \theta_2 - (\theta_2\, Q_2)\, T_2 = \theta_2, \\
(\alpha'_2 \circ \alpha'_1)(\mu_1) = \mu_1\, (R_2\, S_2) = \mu_1 - (\mu_1\, R_1)\, S_1 = \mu_1,
\end{array}
\right.$$

for all $\theta_2 \in \ker_D(.Q_2)$ and for all $\mu_1 \in \ker_D(.R_1)$, which shows that α'_1 is an isomorphism, i.e., $\ker_D(.Q_2) \cong \ker_D(.R_1)$, and $\alpha'_2 = {\alpha'_1}^{-1}$.

In the above arguments, we can exchange the role played by R_1 (resp., Q_2) by that of S_1 (resp., T_2) in the identities (2.21) and (2.23) to get the following isomorphisms

$$\begin{array}{rl} \beta_1 : \mathrm{coker}_D(.S_1) \longrightarrow \mathrm{coker}_D(.T_2) & \beta_1^{-1} : \mathrm{coker}_D(.T_2) \longrightarrow \mathrm{coker}_D(.S_1) \\ \sigma_1(\zeta_1) \longmapsto \varepsilon_2(\zeta_1\, R_2), & \varepsilon_2(\xi_2) \longmapsto \sigma_1(\xi_2\, S_2), \end{array} \tag{2.26}$$

$$\begin{array}{rl} \beta_1' : \ker_D(.S_1) \longrightarrow \ker_D(.T_2) & \beta_1'^{-1} : \ker_D(.T_2) \longrightarrow \ker_D(.S_1) \\ \vartheta_1 \longmapsto -\vartheta_1\, Q_1, & \varpi_2 \longmapsto -\varpi_2\, T_1, \end{array} \tag{2.27}$$

where $\sigma_1 : D^{1\times q} \longrightarrow \mathrm{coker}_D(.S_1)$ (resp., $\varepsilon_2 : D^{1\times s} \longrightarrow \mathrm{coker}_D(.T_2)$) is the canonical projection, i.e., we have:

$$\mathrm{coker}_D(.S_1) \cong \mathrm{coker}_D(.T_2), \quad \ker_D(.S_1) \cong \ker_D(.T_2).$$

Using (2.23), we get $Q_1\, T_2 = -S_1\, R_2$, which yields the following commutative exact diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \ker_D(.Q_1) & \longrightarrow & D^{1\times p} & \xrightarrow{.Q_1} & D^{1\times t} & \xrightarrow{\kappa_1} & \mathrm{coker}_D(.Q_1) & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \gamma_1'} & & \downarrow{\scriptstyle .-S_1} & & \downarrow{\scriptstyle .T_2} & & \downarrow{\scriptstyle \gamma_1} & & \\ 0 & \longrightarrow & \ker_D(.R_2) & \longrightarrow & D^{1\times q} & \xrightarrow{.R_2} & D^{1\times s} & \xrightarrow{\pi_2} & \mathrm{coker}_D(.R_2) & \longrightarrow & 0, \end{array}$$

where γ_1 and γ_1' are respectively defined by:

$$\begin{array}{rl} \gamma_1 : \mathrm{coker}_D(.Q_1) \longrightarrow \mathrm{coker}_D(.R_2) & \gamma_1' : \ker_D(.Q_1) \longrightarrow \ker_D(.R_2) \\ \kappa_1(\nu_1) \longmapsto \pi_2(\nu_1\, T_2), & \theta_1 \longmapsto -\theta_1\, S_1. \end{array} \tag{2.28}$$

Using (2.21), we get $R_2\, Q_2 = -R_1\, Q_1$, which yields the following commutative exact diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \ker_D(.R_2) & \longrightarrow & D^{1\times q} & \xrightarrow{.R_2} & D^{1\times s} & \xrightarrow{\pi_2} & \mathrm{coker}_D(.R_2) & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \gamma_2'} & & \downarrow{\scriptstyle .-R_1} & & \downarrow{\scriptstyle .Q_2} & & \downarrow{\scriptstyle \gamma_2} & & \\ 0 & \longrightarrow & \ker_D(.Q_1) & \longrightarrow & D^{1\times p} & \xrightarrow{.Q_1} & D^{1\times t} & \xrightarrow{\kappa_1} & \mathrm{coker}_D(.Q_1) & \longrightarrow & 0, \end{array}$$

where γ_2 and γ_2' are respectively defined by:

$$\begin{array}{rl} \gamma_2 : \mathrm{coker}_D(.R_2) \longrightarrow \mathrm{coker}_D(.Q_1) & \gamma_2' : \ker_D(.R_2) \longrightarrow \ker_D(.Q_1) \\ \pi_2(\lambda_2) \longmapsto \kappa_1(\lambda_2\, Q_2), & \mu_2 \longmapsto -\mu_2\, R_1. \end{array} \tag{2.29}$$

Using (2.21) and (2.23), we get $T_2\, Q_2 = I_t - T_1\, Q_1$ and $Q_2\, T_2 = I_s - S_2\, R_2$, which yields

$$\begin{cases} (\gamma_2 \circ \gamma_1)(\kappa_1(\nu_1)) = \kappa_1(\nu_1\, (T_2\, Q_2)) = \kappa_1(\nu_1) - \kappa_1((\nu_1\, T_1)\, Q_1) = \kappa_1(\nu_1), \\ (\gamma_1 \circ \gamma_2)(\pi_2(\lambda_2)) = \pi_2(\lambda_2\, (Q_2\, T_2)) = \pi_2(\lambda_2) - \pi_2((\lambda_2\, S_2)\, R_2) = \pi_2(\lambda_2), \end{cases}$$

and shows that γ_1 is an isomorphism, i.e., $\mathrm{coker}_D(.Q_1) \cong \mathrm{coker}_D(.R_2)$, and $\gamma_2' = \gamma_2^{-1}$.

Using (2.23) and (2.21), we get $S_1\,R_1 = I_p - Q_1\,T_1$ and $R_1\,S_1 = I_q - R_2\,S_2$, which yields

$$\begin{cases} (\gamma_2' \circ \gamma_1')(\theta_1) = \theta_1\,(S_1\,R_1) = \theta_1 - (\theta_1\,Q_1)\,T_1 = \theta_1, \\ (\gamma_1' \circ \gamma_2')(\mu_2) = \mu_2\,(R_1\,S_1) = \mu_2 - (\mu_2\,R_2)\,S_2 = \mu_2, \end{cases}$$

for all $\theta_1 \in \ker_D(.Q_1)$ and for all $\mu_2 \in \ker_D(.R_2)$, which shows that γ_1' is an isomorphism, i.e., $\ker_D(.Q_1) \cong \ker_D(.R_2)$, and $\gamma_2' = \gamma_1'^{-1}$.

Finally, we can similarly show that we have the following isomorphisms

$$\begin{array}{rlrl} \delta_1 : \mathrm{coker}_D(.T_1) & \longrightarrow \mathrm{coker}_D(.S_2) & \delta_1^{-1} : \mathrm{coker}_D(.S_2) & \longrightarrow \mathrm{coker}_D(.T_1) \\ \varepsilon_1(\xi_1) & \longmapsto \sigma_2(\xi_1\,S_1), & \sigma_2(\zeta_2) & \longmapsto \varepsilon_1(\zeta_2\,R_1), \end{array} \tag{2.30}$$

$$\begin{array}{rlrl} \delta_1' : \ker_D(.T_1) & \longrightarrow \ker_D(.S_2) & \delta_1'^{-1} : \ker_D(.S_2) & \longrightarrow \ker_D(.T_1) \\ \varpi_1 & \longmapsto -\varpi_1\,T_2, & \theta_2 & \longmapsto -\theta_2\,Q_2, \end{array} \tag{2.31}$$

where $\varepsilon_1 : D^{1\times p} \longrightarrow \mathrm{coker}_D(.T_1)$ (resp., $\sigma_2 : D^{1\times q} \longrightarrow \mathrm{coker}_D(.S_2)$) is the canonical projection, i.e., we have:

$$\mathrm{coker}_D(.T_1) \cong \mathrm{coker}_D(.S_2), \quad \ker_D(.T_1) \cong \ker_D(.S_2).$$

Right D-module analogs of (2.22) can be proved similarly.

Remark 4 When $s \leqslant q$ and $t = p - (q - s) > 0$, Theorem 2 shows that we have $M := \mathrm{coker}_D(.R_1) \cong \mathrm{coker}_D(.Q_2)$, where $Q_2 \in D^{s\times t}$, which yields $\ker_{\mathcal{F}}(R_1.) \cong \ker_{\mathcal{F}}(Q_2.)$ for all left D-modules $\mathcal{F}$, i.e., the linear system $R_1\,\eta = 0$ is equivalent to the linear system $Q_2\,\zeta = 0$ defined by fewer equations and fewer unknowns. Such a reduction is called *Serre's reduction* and is studied in detail in [1, 9, 11]. Theorem 2 is an extension of Theorem 4.1 of [1] for a non necessarily full row rank matrix R_1.

Example 11 With the notations $R_1 := R$, $R_2 := -Q$, $T_1 := P'$, $T_2 := Z'$, $S_1 := Z$, $S_2 := -Q'$, $Q_1 := P$, and $Q_2 := R'$, in Example 8, we proved the identity (2.21). By Theorem 2, we find again that $M := \mathrm{coker}_D(.R) \cong M' := \mathrm{coker}_D(.R')$, where R and R' have not full row rank (see Example 8), and $\ker_D(.R) \cong \ker_D(.R')$. We also have $\mathrm{coker}_D(.Z) \cong \mathrm{coker}_D(.Z')$ and $\ker_D(.Z) \cong \ker_D(.Z')$ (see (2.26) and (2.27)), $\mathrm{coker}_D(.Q) \cong \mathrm{coker}_D(.P)$ and $\ker_D(.Q) \cong \ker_D(.P)$ (see (2.28) and (2.29)), $\mathrm{coker}_D(.P') \cong \mathrm{coker}_D(.Q')$ and $\ker_D(.P') \cong \ker_D(.Q')$ (see (2.30) and (2.31)).

Example 12 We consider again Example 10. Theorem 2 then shows that we have $M := \mathrm{coker}_D(.R) \cong M' := \mathrm{coker}_D(.R')$, $\mathrm{coker}_D(.P) \cong \mathrm{coker}_D(.Q) = 0$ and $\ker_D(.P) \cong \ker_D(.Q) = 0$ since $P \in \mathrm{GL}_p(D)$ and $Q \in \mathrm{GL}_q(D)$.

Example 13 We consider again Example 9, where $S_1 := Z = (0 \quad 0)$ and $T_2 := Z'$. We can check that $\mathrm{coker}_D(.Z')$ is a free left D-module of rank 5 and $\mathrm{coker}_D(.Z) \cong D^{1\times 2}$ is a free left D-module of rank 2. Hence, the isomorphisms (2.22) of Theorem 2 only hold when we have (2.21) and not when (2.14) and (2.15) hold.

We can give a system-theoretic interpretation of Theorem 2. The hypotheses of Theorem 2 show that we can inflate the linear system $R_1\,\eta_1 = 0$ into the larger linear system $R_1\,\eta_1 + R_2\,\eta_2 = 0$ which is *flat* (see [3, 13, 21] and the references therein), i.e., which is associated with the free left D-module $E := \mathrm{coker}_D(.(R_1 \quad R_2))$ of rank $t = p - q + s$ (see Remark 3). Then, we know that the flat system admits an injective parametrization, i.e., we have $\ker_{\mathcal{F}}((R_1 \quad R_2).) = \mathrm{im}_{\mathcal{F}}((Q_1^T \quad Q_2^T)^T.)$, where $T_1\,Q_1 + T_2\,Q_2 = I_t$. For more details, see [3, 22]. Hence, we get

$$R_1\,\eta_1 + R_2\,\eta_2 = 0 \quad \Leftrightarrow \quad \begin{cases} \eta_1 = Q_1\,\xi, \\ \eta_2 = Q_2\,\xi, \end{cases}$$

for a certain $\xi \in \mathcal{F}^t$ which is such that $\xi = T_1\,\eta_1 + T_2\,\eta_2$. Now, setting $\eta_2 = 0$, we get that for $\eta_1 \in \ker_{\mathcal{F}}(R_1.)$, there exists a unique $\xi = T_1\,\eta_1 \in \mathcal{F}^t$ such that:

$$\eta_1 = Q_1\,\xi, \quad Q_2\,\xi = 0.$$

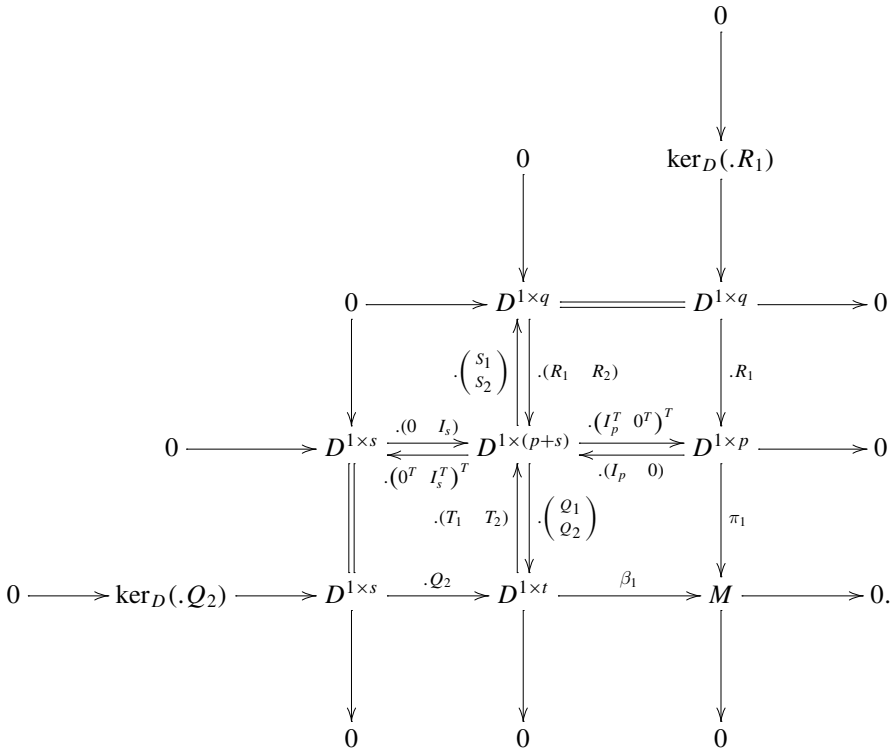
Within systems theory, we find again the first isomorphisms of (2.24) and (2.25).

For instance, the linear OD system $\dot{x}(t) = A\,x(t)$, with $A \in \mathbb{R}^{n\times n}$, is equivalent to an ODE with constant coefficients in one unknown if and only if there exists $B \in \mathbb{R}^n$ such that the control (inflated) linear system $\dot{x}(t) = A\,x(t) + B\,u(t)$ is flat, i.e., if and only if it is controllable. For more details and extensions, see [9].

Remark 5 We can give another (pictorial) proof of the first point of Theorem 2, i.e., of $\mathrm{coker}_D(.R_1) \cong \mathrm{coker}_D(.Q_2)$ and $\ker_D(.R_1) \cong \ker_D(.Q_2)$. Identities (2.21) and (2.23) are equivalent to the following split short exact sequence of left D-modules:

$$0 \longrightarrow D^{1\times q} \underset{\cdot\begin{pmatrix} S_1 \\ S_2 \end{pmatrix}}{\overset{.(R_1 \quad R_2)}{\rightleftarrows}} D^{1\times(p+s)} \underset{.(T_1 \quad T_2)}{\overset{\cdot\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}}{\rightleftarrows}} D^{1\times t} \longrightarrow 0. \tag{2.32}$$

For more details, see, e.g., [21, 24]. With the above notations, we then have the following commutative exact diagram:



(2.33)

Let us denote

$$L := \operatorname{coker}_D(.Q_2) = D^{1\times t}/(D^{1\times s}\, Q_2),\ \ M := \operatorname{coker}_D(.R_1) = D^{1\times p}/(D^{1\times q}\, R_1),$$

and $\kappa_2 : D^{1\times t} \longrightarrow L$ (resp., $\pi_1 : D^{1\times p} \longrightarrow M$) the canonical projection. Then, using (2.33), we obtain the following isomorphism:

$$\begin{aligned}
&\phi : L \longrightarrow M \\
&\kappa_2(\nu_2) \longmapsto \pi_1\left(\nu_2 \begin{pmatrix} T_1 & T_2 \end{pmatrix} \begin{pmatrix} I_p \\ 0 \end{pmatrix}\right) = \pi_1(\nu_2\, T_1), \\
&\phi^{-1} : M \longrightarrow L \\
&\pi_1(\lambda_1) \longmapsto \kappa_2\left(\lambda_1 \begin{pmatrix} I_p & 0 \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}\right) = \kappa_2(\lambda_1\, Q_1).
\end{aligned}$$

A *chase* in the commutative exact diagram (2.33) (see, e.g., [24]) yields the following isomorphism

$$\begin{array}{ll} \gamma : \ker_D(.Q_2) \longrightarrow \ker_D(.R_1) & \gamma^{-1} : \ker_D(.R_1) \longrightarrow \ker_D(.Q_2) \\ \theta_2 \longmapsto \theta_2\, S_2, & \mu_1 \longmapsto \mu_1\, R_2, \end{array}$$

i.e., we have $\ker_D(.R_1) \cong \ker_D(.Q_2)$. Finally, the other isomorphisms (2.22) stated in Theorem 2 can be proved similarly.

Corollary 4 *The following two assertions are equivalent:*

(a) The matrices $R \in D^{q\times p}$ and $R' \in D^{q\times p}$ are equivalent, namely, there exist $P \in \mathrm{GL}_p(D)$ and $Q \in \mathrm{GL}_q(D)$ such that $R' = Q^{-1}\, R\, P$.
(b) There exist $Q \in \mathrm{GL}_q(D)$ and $U \in \mathrm{GL}_{p+q}(D)$ such that:

$$(R \quad -Q)\, U = (I_q \quad 0).$$

Proof $1 \Rightarrow 2$. If R and R' are equivalent, then 2 is proved in Example 10 (see (2.17)).
$2 \Rightarrow 1$. Let us note:

$$U := \begin{pmatrix} Z & P \\ -Q' & R' \end{pmatrix}, \quad U^{-1} := \begin{pmatrix} R & -Q \\ P' & Z' \end{pmatrix}.$$

In particular, we have $R\, P = Q\, R'$, i.e., $R' = Q^{-1}\, R\, P$ since $Q \in \mathrm{GL}_q(D)$. Now, (2.22) yields $\mathrm{coker}_D(.P) \cong \mathrm{coker}_D(.Q) = 0$ and $\ker_D(.P) \cong \ker_D(.Q) = 0$ since $Q \in \mathrm{GL}_q(D)$, which shows that $P \in \mathrm{GL}_p(D)$ and proves 1. Finally, using $P'\, P + Z'\, R' = I_p$, we get $(P' + Z'\, Q^{-1}\, R)\, P = I_p$ which shows that we have:

$$P^{-1} = P' + Z'\, Q^{-1}\, R.$$

Finally, let us give an application of Theorem 2 for the study of *doubly coprime factorizations* (see, e.g., [25]). To keep the standard notations used within the *fractional representational approach* [25], we now denote the ring D by A.

Corollary 5 *Let A be an integral domain, namely, a commutative ring with no non-zero divisors, $K := \{\frac{n}{d} \mid 0 \neq d,\ n \in A\}$ the quotient field of A, $P \in K^{q\times r}$, and $P = D^{-1}\, N = \widetilde{N}\, \widetilde{D}^{-1}$ a doubly coprime factorization of P, namely, $D \in A^{q\times q}$, $N \in A^{q\times r}$, $\widetilde{D} \in A^{r\times r}$, and $\widetilde{N} \in A^{q\times r}$ satisfying the following identity*

$$\begin{pmatrix} D & -N \\ -\widetilde{Y} & \widetilde{X} \end{pmatrix} \begin{pmatrix} X & \widetilde{N} \\ Y & \widetilde{D} \end{pmatrix} = I_{q+r},$$

for some matrices $X \in A^{q\times q}$, $Y \in A^{r\times q}$, $\widetilde{X} \in A^{r\times r}$, and $\widetilde{Y} \in A^{r\times q}$. Then, we have the following isomorphisms of A-modules:

$$\left\{\begin{array}{l}\mathrm{coker}_A(.D) \cong \mathrm{coker}_A(.\widetilde{D}),\\ \mathrm{coker}_A(.X) \cong \mathrm{coker}_A(.\widetilde{X}),\\ \mathrm{coker}_A(.N) \cong \mathrm{coker}_A(.\widetilde{N}),\\ \mathrm{coker}_A(.Y) \cong \mathrm{coker}_A(.\widetilde{Y}),\end{array}\right. , \quad \left\{\begin{array}{l}\ker_A(.D) \cong \ker_A(.\widetilde{D}) = 0,\\ \ker_A(.X) \cong \ker_A(.\widetilde{X}),\\ \ker_A(.N) \cong \ker_A(.\widetilde{N}),\\ \ker_A(.Y) \cong \ker_A(.\widetilde{Y}).\end{array}\right.$$

Similar results hold for right matrix multiplication, i.e., we also have:

$$\mathrm{coker}_A(D.) \cong \mathrm{coker}_A(\widetilde{D}.),\ \ker_A(D.) \cong \ker_A(\widetilde{D}.) = 0,\ \ldots$$

References

1. Boudellioua, M.S., Quadrat, A.: Serre's reduction of linear functional systems. Math. Comput. Sci. **2–3**, 289–312 (2010)
2. Bryc, W., Letac, G.: Meixner matrix ensembles. J. Theor. Probab. **26**, 107–152 (2013)
3. Chyzak, F., Quadrat, A., Robertz, D.: Effective algorithms for parametrizing linear control systems over Ore algebras. Appl. Algebra Engrg. Commun. Comput. **16**, 319–376 (2005)
4. Chyzak, F., Quadrat, A., Robertz, D.: OreModules: a symbolic package for the study of multidimensional linear systems. Lecture Notes in Control and Information Sciences, vol. 352, pp. 233–264. Springer (2007). https://who.rocq.inria.fr/Alban.Quadrat/OreModules/index.html
5. Chyzak, F., Salvy, B.: Non-commutative elimination in Ore algebras proves multivariate identities. J. Symb. Comput. **26**, 187–227 (1998)
6. Cluzeau, T., Quadrat, A.: Factoring and decomposing a class of linear functional systems. Linear Algebra Appl. **428**, 324–381 (2008)
7. Cluzeau, T., Quadrat, A.: OreMorphisms: a homological algebraic package for factoring, reducing and decomposing linear functional systems. Lecture Notes in Control and Information Sciences vol. 388, pp. 179–194. Springer (2009). https://who.rocq.inria.fr/Alban.Quadrat/OreMorphisms/index.html
8. Cluzeau, T., Quadrat, A.: A constructive version of Fitting's theorem on isomorphisms and equivalences of linear systems. In: Proceedings of nDS'11, Poitiers (France) (05–07/09/11)
9. Cluzeau, T., Quadrat, A.: Serre's reduction of linear systems of partial differential equations with holonomic adjoints. J. Symb. Comput. **47**, 1192–1213 (2012)
10. Cluzeau, T., Quadrat, A.: Isomorphisms and Serre's reduction of linear systems. In: Proceedings of nDS13, Erlangen (Germany) (9–11/09/13)
11. Cluzeau, T., Quadrat, A.: A new insight into Serre's reduction problem. Linear Algebra Appl. **483**, 40–100 (2015)
12. Cluzeau, T., Quadrat, A., Tõnso, M.: OreAlgebraicAnalysis: A Mathematica package for the algorithmic study of linear functional systems. OreAlgebraicAnalysis project (2015). https://who.rocq.inria.fr/Alban.Quadrat/OreAlgebraicAnalysis/index.html
13. Fliess, M., Mounier, H.: Controllability and observability of linear delay systems: an algebraic approach. ESAIM COCV **3**, 301–314 (1998)
14. Hotta, R., Takeuchi, K., Tanisaki, T.: *D*-Modules, Perverse Sheaves, and Representation Theory. Progress in Mathematics, vol. 236. Birkhäuser, Basel (2008)
15. Kashiwara, M.: Algebraic Study of Systems of Partial Differential Equations. Mémoires de la Société Mathématique de France 63 (1995). English translation (Kyoto 1970)
16. Lam, T.Y.: Lectures on Modules and Rings. Graduate Texts in Mathematics, vol. 189. Springer, Berlin (1999)
17. Malgrange, B.: Systèmes différentiels à coefficients constants. Séminaire Bourbaki **1962**(63), 1–11 (1962)

18. McConnell, J.C., Robson, J.C.: Noncommutative Noetherian Rings. American Mathematical Society (2000)
19. Oberst, U.: Multidimensional constant linear systems. Acta Appl. Math. **20**, 1–175 (1990)
20. Pal, D., Pillai, H.K.: Algorithms for the theory of restrictions of scalar nD systems to proper subspaces of $\mathbb{R}^n$. Multidim. Syst. Sign. Process. **26**, 439–457 (2015)
21. Quadrat, A.: An introduction to constructive algebraic analysis and its applications. Les cours du CIRM. Journées Nationales de Calcul Formel (2010) **1**(2), 281–471 (2010). https://hal.inria.fr/inria-00506104/document
22. Quadrat, A., Robertz, D.: Computation of bases of free modules over the Weyl algebras. J. Symbolic Comput. **42**, 1113–1141 (2007)
23. Robertz, D.: Recent progress in an algebraic analysis approach to linear systems. Multidimens. Syst. Signal Process. **26**, 349–388 (2015)
24. Rotman, J.J.: An Introduction to Homological Algebra. Academic, Cambridge (1979)
25. Vidyasagar, M.: Control System Synthesis: A Factorization Approach. MIT Press, Cambridge (1985)

Chapter 3
Computing Polynomial Solutions and Annihilators of Integro-Differential Operators with Polynomial Coefficients

Alban Quadrat and Georg Regensburger

Abstract In this chapter, we study algorithmic aspects of the algebra of linear ordinary integro-differential operators with polynomial coefficients. Even though this algebra is not Noetherian and has zero divisors, Bavula recently proved that it is coherent, which allows one to develop an algebraic systems theory over this algebra. For an algorithmic approach to linear systems of integro-differential equations with boundary conditions, computing the kernel of matrices with entries in this algebra is a fundamental task. As a first step, we have to find annihilators of integro-differential operators, which, in turn, is related to the computation of polynomial solutions of such operators. For a class of linear operators including integro-differential operators, we present an algorithmic approach for computing polynomial solutions and the index. A generating set for right annihilators can be constructed in terms of such polynomial solutions. For initial value problems, an involution of the algebra of integro-differential operators then allows us to compute left annihilators, which can be interpreted as compatibility conditions of integro-differential equations with boundary conditions. We illustrate our approach using an implementation in the computer algebra system `Maple`.

Keywords Linear systems theory · Rings of integro-differential operators · Polynomial solutions · Indicial equation · Compatibility conditions · Computer algebra

Supported by the Austrian Science Fund (FWF): P27229.

A. Quadrat (✉)
Inria Paris, Institut de Mathématiques de Jussieu-Paris Rive Gauche, Sorbonne University, Paris, France
e-mail: alban.quadrat@inria.fr

G. Regensburger
Institute for Algebra, Johannes Kepler University Linz, Linz, Austria
e-mail: georg.regensburger@jku.at

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_3

3.1 Introduction

Rings of *functional operators* (e.g., rings of ordinary differential (OD) operators, partial differential (PD) operators, differential time-delay operators, differential difference operators) were recently introduced in mathematical systems theory. Since many control linear systems can be defined by means of a matrix with entries in a *skew polynomial ring*, in an *Ore algebra* or in an *Ore extension* of functional operators (i.e., classes of univariate or multivariate noncommutative polynomial rings) [16, 39], the classical *polynomial approach* to linear systems theory can be generalized yielding a *module-theoretic approach* to linear functional systems [19, 33, 34, 41, 45, 47]. Symbolic computation techniques (e.g., Gröbner basis techniques) and computer algebra systems can then be used to develop dedicated packages for algebraic systems theory [17, 29]. For more details, see Chap. 1.

Algebras of ordinary integro-differential (ID) operators have recently been studied within an algebraic approach in [8–11] and within an algorithmic approach in [22, 40, 42, 43]. The goal of the latter works is to provide an algebraic and algorithmic framework for studying *boundary value problems and Green's operators*.

The ring of ID and time-delay/dilatation operators was introduced in [37] to develop a purely algorithmic approach to standard *Artstein's transformation* of linear differential systems with delayed inputs. This work also advocates for the effective study of the ring of ID time-delay/dilatation operators. The *normal forms* of elements of this noncommutative algebra will be studied in a future publication based on the new effective techniques introduced in [22, 23]. In this paper, we focus on its subring of ID operators. We also note that effective computations over ID algebras play an important role in *parameter estimation problems* as shown in [15].

Even though linear systems of ID equations play an important role in different domains and applications (e.g., PID controllers), it does not seem that they have been extensively studied by the mathematical systems community. For *boundary value systems*, we refer to [20, 21] and the references therein. The first purpose of this paper is to introduce concepts, techniques, and results developed in the above recents works. In particular, we emphasize that the algebraic structure of the ring of ID operators with polynomial coefficients is much more involved (e.g., zero divisors, non-Noetherianity) than the one of the ring of OD operators with polynomial coefficients (the so-called *Weyl algebra*). The fundamental issue of computing left/right kernel of a matrix of ID operators has to be solved towards developing a system-theoretic approach to linear ID systems. For more details, see [16, 36].

The second goal of this paper is to study this problem for a single ID operator, that is, computing its *annihilator*. Within a *representation approach*, we show that this problem is related to the computation of polynomial solutions of ID operators, a problem that is also studied in detail. To solve this problem, we introduce the concept of a *rational indicial equation* for a linear operator acting on the polynomial ring. This approach allows us to find again and generalize standard results on the *indicial equation* classically used in the theory of linear OD equations [2, 3, 5].

This chapter is based on the conference paper [38]. It includes a self-contained introduction to ordinary integro-differential operators with polynomial coefficients with several evaluations including normal forms (Sects. 3.2–3.4). All other sections have been revised and extended.

3.2 The Ring of Ordinary Integro-Differential Operators with Polynomial Coefficients

Before discussing the ring of ID operators with polynomial coefficients, as an introducing example, we first recall two standard constructions of the ring A of OD operators with polynomial coefficients (also called *the Weyl algebra* and denoted by $A_1(k)$, where k is a field). The first construction is as the subalgebra $k\langle t, \partial\rangle$ of all linear maps on the polynomial ring $k[t]$ and the second is by means of generators and relations.

In what follows, let k denote a fixed field, which contains $\mathbb{Q}$. Let $\mathrm{end}_k(k[t])$ denote the k-algebra formed by all k-linear maps from the polynomial ring $k[t]$ to itself. We consider the k-subalgebra $k\langle t, \partial\rangle$ of $\mathrm{end}_k(k[t])$ generated by the following two k-linear maps

$$t\colon t^n \longmapsto t^{n+1} \quad \text{and} \quad \partial\colon t^n \longmapsto n\,t^{n-1}$$

defined on the basis $(t^n)_{n\in\mathbb{N}}$ of $k[t]$. They respectively correspond to the multiplication operator and the derivation on the polynomial ring $k[t]$, namely:

$$\begin{aligned} t\colon k[t] &\longrightarrow k[t] \\ p &\longmapsto t\,p, \end{aligned} \quad \text{and} \quad \begin{aligned} \partial\colon k[t] &\longrightarrow k[t] \\ p &\longmapsto \tfrac{dp}{dt}. \end{aligned} \tag{3.1}$$

One immediately verifies that we have

$$\forall\ p \in k[t], \quad (\partial \circ t)(p) = \frac{d(t\,p)}{dt} = t\,\frac{dp}{dt} + p = (t \circ \partial + \mathrm{id})(p),$$

where id (also denoted by 1) is the identity map on $k[t]$. It shows that the *Leibniz rule*

$$\partial \circ t = t \circ \partial + \mathrm{id}$$

holds in the operator algebra $k\langle t, \partial\rangle$.

Using the Leibniz rule, we can define the Weyl algebra also by *generators and relations*: let $k\langle T, D\rangle$ be the *free associative* k-algebra on the set $\{T, D\}$, that is, the k-vector space with the basis formed by all words over $\{T, D\}$ and the multiplication of basis elements defined by concatenation. Let now

$$J = (D\,T - T\,D - 1) \subseteq k\langle T, D\rangle$$

denote the two-sided ideal generated by $D\,T - T\,D - 1$ and define the k-algebra:

$$A = k\langle T, D\rangle / J.$$

By definition, the Leibniz rule

$$D\,T \equiv T\,D + 1 \mod J$$

holds in A. Using this identity, each element of $d \in A$ can uniquely be written as a finite sum

$$d \equiv \sum a_{ij}\,T^i\,D^j \mod J$$

with coefficients $a_{ij} \in k$.

To see that the two constructions above are equivalent, one can use the fact that A is a *simple ring*, that is, its only proper two-sided ideal is the zero ideal (see, for example, [18]). Hence every ring homomorphism is injective and so the k-algebra homomorphism $A \longrightarrow k\langle t, \partial\rangle$ mapping

$$T + J \longmapsto t \quad \text{and} \quad D + J \longmapsto \partial$$

is an isomorphism. In other words, each $d \in A$ can be identified with the following corresponding k-linear map

$$\begin{aligned} L_d \colon k[t] &\longrightarrow k[t], \\ p &\longmapsto d(p), \end{aligned}$$

where $d(p)$ denotes the action of d on p.

In the following, we use a similar approach to introduce and study the algebra of ID operators with polynomial coefficients. ID operators with polynomial coefficients were studied in [8, 10] as a *generalized Weyl algebra* [6, 7]. See [40] for the construction of ordinary ID operators with polynomial coefficients as a factor algebra of a *skew polynomial ring* (see, e.g., [16, 31] and the references therein). For the construction of the algebra of ID operators $\mathcal{F}_\Phi\langle\partial, \int\rangle$ defined over an ordinary ID algebra $\mathcal{F}$ and endowed with a set of *characters* (that is, multiplicative linear functionals) Φ, we refer to [42, 43]. This construction is based on a parametrized noncommutative Gröbner basis; see Sect. 3.3 for the case of polynomial coefficients. For a basis-free construction using a finite reduction system in *tensor algebras*, we refer to [22]. In contrast to [8, 10], the last two approaches allows one to have more than one point evaluation as described in Sect. 3.4, which is crucial for the study of *boundary problems*.

Definition 1 The *k-algebra of ordinary ID operators with polynomial coefficients* is defined as the k-subalgebra

$$k\langle t, \partial, \int\rangle \subseteq \mathrm{end}_k(k[t]),$$

with the operators t and ∂ defined as in (3.1) and

$$\begin{aligned} \int : k[t] &\longrightarrow k[t] \\ t^n &\longmapsto t^{n+1}/(n+1), \end{aligned}$$

defined on the basis $(t^n)_{n\in\mathbb{N}}$ of $k[t]$.

The integral operator $\int$ corresponds to the usual integral starting at 0:

$$\begin{aligned} \int : k[t] &\longrightarrow k[t] \\ p &\longmapsto \int_0^t p(s)\,ds. \end{aligned}$$

One can verify directly that the *fundamental theorem* of calculus

$$\partial \circ \int = \mathrm{id}$$

holds. Moreover, we see that

$$\mathbf{E} = \mathrm{id} - \int \circ \partial$$

corresponds to the *evaluation* at 0:

$$\begin{aligned} \mathbf{E}: k[t] &\longrightarrow k[t] \\ p &\longmapsto p(0). \end{aligned}$$

Hence, as soon as we have an integral, we also have one evaluation map to the constants k "for free", which allows us to define and study initial value problems in terms of integro-differential operators. Note that the operator $\mathbf{E}$ naturally induces the existence of *zero divisors*. For instance, we have:

$$\mathbf{E} \circ t = 0.$$

Based on the basic identities above, we can construct the algebra of integro-differential operators with polynomial coefficients also by generators and relations.

Definition 2 We define the k-algebra

$$\mathbb{I} = k\langle T, D, I, E\rangle / J,$$

where J is the two-sided ideal of relations generated by the following elements:

$$DT - TD - 1, \quad DI - 1, \quad ID + E - 1, \quad ET. \tag{3.2}$$

We note by $\overline{T} = T + J$ (resp., $\overline{D} = D + J$, $\overline{I} = I + J$, $\overline{E} = E + J$) the residue class of T (resp., D, I, E) in $\mathbb{I}$.

3.3 Normal Forms

Since we have now four defining identities for $\mathbb{I}$ (see (3.2)) instead of one as for the Weyl algebra A, it is more involved to obtain the *normal form* of an element of $\mathbb{I}$, i.e., its unique expression as a noncommutative polynomial in the operators T, D, I and E modulo the relations (3.2). In this section, we informally discuss the construction of a noncommutative Gröbner basis for the defining ideal following Buchberger's algorithm. For background on noncommutative Gröbner bases, we refer to [12, 13, 32, 46]. In the noncommutative case, note that Buchberger's algorithm does not terminate in general and the property of having a finite Gröbner basis is undecidable. However, in our case we can "guess" a parametrized Gröbner basis from the corresponding *S-polynomial computations*.

See [42, 43] for further details on a parametrized Gröbner basis for the defining relations for integro-differential operators over an ordinary ID algebra and the corresponding normal forms. An analogous finite tensor reduction system and the related S-polynomial computations using the package `TenRes` can be found in [22, 23].

We denote the S-polynomial between two polynomials of the form

$$U\,V - P \quad \text{and} \quad V\,W - Q,$$

with "leading terms" $U\,V$ and $V\,W$ by:

$$S(U\,V, V\,W) = P\,W - U\,Q.$$

In the following, we consider a graded partial order with $D > T$ and $I > T$. We first compute the S-polynomial between the polynomials

$$D\,I - 1 \quad \text{and} \quad I\,D + E - 1$$

and obtain:

$$S(D\,I, I\,D) = 1\,D - D\,(1 - E) = D\,E.$$

So we need to add the polynomial

$$D\,E$$

to the generators of our ideal, which corresponds to the evaluation mapping to k. The S-polynomial between $I\,D + E - 1$ and the new polynomial gives:

$$S(I\,D, D\,E) = (1 - E)\,E.$$

So we obtain

$$E^2 - E,$$

which corresponds to the evaluation acting as a *projector* onto k. Since

$$S(I\,D, D\,I) = (1 - E)\,I - I\,1 = -E\,I,$$

we also have to add the polynomial

$$E\,I$$

to our generators, which corresponds to the integral $\int_0^t$ evaluated at 0 being 0.

The S-polynomial between

$$I\,D - 1 + E \quad \text{and} \quad D\,T - T\,D - 1$$

is given by:

$$S(I\,D, D\,T) = (1 - E)\,T - I\,(T\,D + 1) = T - E\,T - I\,T\,D - I.$$

Using the polynomial $E\,T$ from the original generators, we see that we need to add the polynomial:

$$I\,T\,D - T + I.$$

This gives rise to new S-polynomials with $D\,T - T\,D - 1$ and one sees inductively that we need to add the family

$$\forall\, n \geq 1, \quad I\,T^n\,D - T^n + n\,I\,T^{n-1}$$

to our generators, corresponding to *integration by parts*. Computing the S-polynomials with this family and $D\,E$, we then obtain

$$I\,E - T\,E,$$

and

$$\forall\, n \geq 1, \quad I\,T^n\,E - T^{n+1}/(n+1)\,E$$

which corresponds to the k-linearity of the integral.

Finally, the S-polynomial between

$$I\,T\,D - T + I \quad \text{and} \quad D\,I - 1$$

is given by:

$$S(I\,T\,D, D\,I) = (T - I)\,I - I\,T.$$

So we obtain the polynomial

$$I^2 - T\,I + I\,T,$$

allowing to reduce an iterated integral to a sum of two single integrals. Again, this identity gives rise to an infinite family

$$\forall\, n \geq 1, \quad I\,T^n\,I - (T^{n+1}I + I\,T^{n+1})/(n+1)$$

of new generators.

Collecting all the identities above, one can verify that all parametrized S-polynomials now reduce to zero and we have indeed a Gröbner basis for the defining identities (compare with [42, Proposition 13] and [22, Theorem 5.1]).

Theorem 1 *The generators*

$$D\,T - T\,D - 1, \quad D\,I - 1, \quad I\,D + E - 1, \quad E\,T,$$
$$D\,E, \quad E^2 - E, \quad E\,I, \quad I\,E - T\,E, \quad I^2 - T\,I + I\,T,$$

and the parametrized generators

$$\forall\, n \geq 1, \quad \begin{cases} I\,T^n\,D - T^n + n\,I\,T^{n-1}, \\ I\,T^n\,E - T^{n+1}/(n+1)\,E, \\ I\,T^n\,I - (T^{n+1}\,I + I\,T^{n+1})/(n+1), \end{cases}$$

form a noncommutative Gröbner basis for the ideal J of $\mathbb{I}$ (see Definition 2) with respect to a graded partial order with $D > T$ and $I > T$.

By the normal form corresponding to the Gröbner basis from Theorem 1, using the notations of Definition 2, each $d \in \mathbb{I}$ can uniquely be written as a sum

$$d = d_1 + d_2 + d_3,$$

where

$$d_1 = \sum a_{ij}\,\overline{T}^i\,\overline{D}^j, \qquad d_2 = \sum b_{ij}\,\overline{T}^i\,\overline{I}\,\overline{T}^j, \qquad d_3 = \sum f_{ij}\,\overline{T}^i\,\overline{E}\,\overline{D}^j \tag{3.3}$$

are respectively an OD operator, an *integral operator*, and a *boundary operator*, with a_{ij}, b_{ij}, and $f_{ij} \in k$, and d_1, d_2, and d_3 contain only finitely nonzero summands.

To see that the definition of integro-differential operators via generators and relations and Definition 1 are equivalent, we can use the fact that $\mathbb{I}$ is "almost" a simple ring. The only nonzero proper two-sided ideal is the ideal $(\overline{E})$ generated by the "evaluation" $\overline{E}$. This was first proved by Bavula in [8]. Here we give an alternative proof based on the normal forms and direct sum decomposition above, which also generalizes to the more general setting including several evaluations mentioned in the next section.

Proposition 1 *The only nonzero proper two-sided ideal of $\mathbb{I}$ is $(\overline{E})$.*

Proof Let $d \in \mathbb{I} \setminus (\overline{E})$ with $d \equiv d_1 + d_2 + d_3$ as in (3.3) and $d_1 + d_2 \neq 0$ by assumption. Using the identities

$$\overline{D}\,\overline{T} = \overline{T}\,\overline{D} + 1, \quad \overline{D}\,\overline{I} = 1, \quad \overline{D}\,\overline{E} = 0,$$

we can find a $k \in \mathbb{N}$ such that

$$\overline{D}^k\, d \in A \setminus \{0\}$$

is a nonzero differential operator and the statement follows since A is a simple ring. □

Corollary 1 *The k-algebra homomorphism $\chi\colon \mathbb{I} \longrightarrow k\langle t, \partial, \int\rangle$ mapping*

$$\overline{T} \longmapsto t, \quad \overline{D} \longmapsto \partial, \quad \overline{E} \longmapsto \mathbf{E}, \quad \overline{I} \longmapsto \int$$

is an isomorphism.

In other words, we can identify each $d \in \mathbb{I}$ with the corresponding k-linear map

$$\begin{aligned} L_d\colon k[t] &\longrightarrow k[t], \\ p &\longmapsto d(p), \end{aligned} \tag{3.4}$$

where $d(p)$ denotes the action of d on p.

Finally, using (3.3), up to isomorphism, we have the following direct sum decomposition

$$\mathbb{I} = A \oplus k[t] \int k[t] \oplus (\mathbf{E})$$

with the two-sided ideal $(\mathbf{E})$ of boundary operators generated by $\mathbf{E}$.

3.4 Several Evaluations

For treating boundary problems, we allow additional point evaluations (*characters*, i.e., multiplicative linear forms) in our operator algebra. We denote the evaluation at $\alpha \in k$ by

$$\begin{aligned} \mathbf{E}_\alpha\colon k[t] &\longrightarrow k[t] \\ p &\longmapsto p(\alpha). \end{aligned}$$

The basic identities for evaluations at $\alpha, \beta \in k$ and the derivation ∂ are

$$\mathbf{E}_\alpha \circ t = \alpha\, \mathbf{E}_\alpha, \quad \mathbf{E}_\beta \circ \mathbf{E}_\alpha = \mathbf{E}_\alpha, \quad \partial \circ \mathbf{E}_\alpha = 0.$$

Definition 3 Let Φ be a subset of k with $0 \in \Phi$. Identifying $\mathbf{E}_0$ with $\mathbf{E} = \mathrm{id} - \int \circ\, \partial$, we define the k-subalgebra $k\langle t, \partial, \int, (\mathbf{E}_\alpha)_{\alpha\in\Phi}\rangle$ of $\mathrm{end}_k(k[t])$ formed by the ordinary ID operators with polynomial coefficients with characters $(\mathbf{E}_\alpha)_{\alpha\in\Phi}$.

Clearly, if $\Phi = \{0\}$, then $k\langle t, \partial, \int, (\mathbf{E}_\alpha)_{\alpha\in\Phi}\rangle = \mathbb{I}$. We now construct the algebra of integro-differential operators with a set of characters $(\mathbf{E}_\alpha)_{\alpha\in\Phi}$ by generators and relations.

Definition 4 We define the k-algebra

$$\mathbb{I}_\Phi = k\langle T, D, I, (E_\alpha)_{\alpha\in\Phi}\rangle / J_\Phi,$$

where J_Φ is the two-sided ideal generated by:

$$\begin{gathered} D\,T - T\,D - 1, \quad D\,I - 1, \quad I\,D + E_0 - 1, \\ \forall\ \alpha,\ \beta \in \Phi, \quad \begin{cases} E_\alpha\,T - \alpha E_\alpha, \\ E_\beta\,E_\alpha - E_\alpha, \\ D\,E_\alpha. \end{cases} \end{gathered} \tag{3.5}$$

We note by $\overline{T} = T + J_\Phi$ (resp., $\overline{D} = D + J_\Phi$, $\overline{I} = I + J_\Phi$, $\overline{E_\alpha} = E_\alpha + J_\Phi$ for $\alpha \in \Phi$) the residue class of T (resp., D, I, E_α) in $\mathbb{I}_\Phi$.

For obtaining a Gröbner basis for the ideal of relations J_Φ, to the defining relations (3.5) and the generators from Theorem 1, we have to add the following parametrized generators:

$$\forall\ n \geq 0, \quad \alpha \in \Phi, \quad I\,T^n\,E_\alpha - T^{n+1}/(n+1)\,E_\alpha.$$

By the corresponding normal forms, every ID operator $d \in \mathbb{I}_\Phi$ can be uniquely written as a sum $d = d_1 + d_2 + d_3$, with d_1 and d_2 as in (3.3) and a *boundary operator* of the form

$$d_3 = \sum_{\alpha\in\Phi} \Big(\sum f_{ij}\,\overline{T}^i\,\overline{E_\alpha}\,\overline{D}^j + \sum g_{ij}\,\overline{T}^i\,\overline{E_\alpha}\,\overline{I}\,\overline{T}^j \Big), \tag{3.6}$$

where f_{ij} and $g_{ij} \in k$ and d_3 contains only finitely nonzero summands. Based on the above decomposition, the proof of Proposition 1 can be generalized.

Proposition 2 *The only nonzero proper two-sided ideal of* $\mathbb{I}_\Phi$ *is* $(\{\overline{E_\alpha}\}_{\alpha\in\Phi})$*, simply denoted by* $(\overline{E})$*. Moreover, we have* $(\overline{E}) = (\overline{E_0})$.

The equality $(\overline{E}) = (\overline{E_0})$ comes from the fact that $0 \in \Phi$ and, with the notation of (3.6), from the following identity:

$$d_3 = \sum_{\alpha\in\Phi} \Big(\sum f_{ij}\,\overline{T}^i\,\overline{E_0}\,\overline{E_\alpha}\,\overline{D}^j + \sum g_{ij}\,\overline{T}^i\,\overline{E_0}\,\overline{E_\alpha}\,\overline{I}\,\overline{T}^j \Big) \in (\overline{E_0}).$$

Corollary 2 *The k-algebra homomorphism*

$$\chi\colon \mathbb{I}_\Phi \longrightarrow k\langle t, \partial, \textstyle\int, (\mathbf{E}_\alpha)_{\alpha\in\Phi}\rangle$$

mapping

$$\overline{T} \longmapsto t, \quad \overline{D} \longmapsto \partial, \quad \overline{E_\alpha} \longmapsto \mathbf{E}_\alpha, \quad \text{for } \alpha \in \Phi, \quad \overline{I} \longmapsto \textstyle\int$$

is an isomorphism.

So we can identify again each $d \in \mathbb{I}_\Phi$ with the corresponding k-linear map L_d on the polynomial ring $k[t]$ as in (3.4). For the rest of the paper, we do this identification and write ∂, $\int$, t, $\mathbf{E}_\alpha$ for both the linear operators on the polynomial ring $k[t]$ and the corresponding residue classes in $\mathbb{I}_\Phi$. So the normal form for an ID operators

$$d = d_1 + d_2 + d_3 \in \mathbb{I}_\Phi$$

from Eqs. (3.3) and (3.6) reads as

$$d_1 = \sum a_{ij}\, t^i\, \partial^j, \quad d_2 = \sum b_{ij}\, t^i \int t^j, \tag{3.7}$$

and

$$d_3 = \sum_{\alpha\in\Phi} \left(\sum f_{ij}\, t^i\, \mathbf{E}_\alpha\, \partial^j + \sum g_{ij}\, t^i\, \mathbf{E}_\alpha \int t^j \right). \tag{3.8}$$

Denoting by $(\{\mathbf{E}_\alpha\}_{\alpha\in\Phi})$ the two-sided ideal of $\mathbb{I}_\Phi$ generated by the $\mathbf{E}_\alpha$'s for $\alpha \in \Phi$, we then have $(\{\mathbf{E}_\alpha\}_{\alpha\in\Phi}) = (\mathbf{E})$, where $(\mathbf{E})$ denotes the two-sided ideal of $\mathbb{I}_\Phi$ generated by $\mathbf{E}$, and, up to isomorphism, we have the following direct sum decomposition:

$$\mathbb{I}_\Phi = A \oplus k[t] \int k[t] \oplus (\mathbf{E}).$$

In particular, the normal form tells us that the corresponding linear maps on the polynomial ring are linearly independent. Since we will need it later, we state this explicitly for the linear functionals in the normal form of boundary operators (3.8).

Lemma 1 *The k-linear functionals $\mathbf{E}_\alpha\, \partial^i$ and $\mathbf{E}_\alpha \int t^i$ on $k[t]$ for $i \in \mathbb{N}$ and $\alpha \in k$ are k-linearly independent.*

3.5 Syzygies and Annihilators

In this section, we discuss some important algebraic properties of the algebra $\mathbb{I}$ concerning finite generating sets of ideals. First, since the integral operator $\int$ is a right but not a left inverse of the derivation ∂, it is known that the algebra $\mathbb{I}$ is necessarily *non-Noetherian* [24].

More explicitly, if $\int^i = \int \cdots \int$ denotes the product of i integral operators and $\int^0 = 1$, using Theorem 1, one verifies that the following operators

$$e_{ij} = \int^i \mathbf{E}\, \partial^j : p \in k[t] \longmapsto p^{(j)}(0)\, \frac{t^i}{i!}$$

satisfy

$$e_{ij}\, e_{lm} = \int^i \mathbf{E}\, \partial^j \int^l \mathbf{E}\, \partial^m = \delta_{jl}\, e_{im}, \tag{3.9}$$

where $\delta_{jl} = 1$ for $j = l$, and 0 otherwise; see also [24] or [28, Ex. 21.26]. In particular, $\mathbb{I}$ contains infinitely many *orthogonal idempotents* e_{ii} for all $i \in \mathbb{N}$, i.e., $e_{ii}\, e_{jj} = \delta_{ij}$ for all $i,\ j \in \mathbb{N}$. Let us introduce the following operator:

$$e_k = e_{00} + e_{11} + \cdots + e_{kk} \in \mathbb{I}.$$

We note that the operator e_k acts on a polynomial p by

$$e_k(p) = \sum_{i=0}^{k} p^{(i)}(0)\, \frac{t^i}{i!},$$

which corresponds to the first k terms of the Taylor series of p at $t = 0$.

Using (3.9), we obtain:

$$\forall\ 0 \leq i \leq k, \quad e_{ii} = e_{ii}\, e_k = e_k\, e_{ii},$$

which yields $e_i\, e_j = e_j\, e_i = e_{\min(i,j)}$. In particular, we have $e_{k-1}\, e_k = e_k\, e_{k-1} = e_{k-1}$, which shows that $\mathbb{I}\, e_{k-1} \subseteq \mathbb{I}\, e_k$ and $e_{k-1}\, \mathbb{I} \subseteq e_k\, \mathbb{I}$. Since e_k is an idempotent of $\mathbb{I}$, i.e. $e_k^2 = e_k$, if we have $e_k \in \mathbb{I}\, e_{k-1}$, i.e. $e_k = \sum_{i=0}^{k-1} d_i\, e_i$ for certain $d_i \in \mathbb{I}$, then we get

$$e_{k-1} = e_k\, e_{k-1} = \sum_{i=0}^{k-1} d_i\, e_i\, e_{k-1} = \sum_{i=0}^{k-1} d_i\, e_i = e_k,$$

which yields a contradiction since $e_k(t^k) = 1$ and $e_{k-1}(t^k) = 0$, and shows that $\mathbb{I}\, e_{k-1} \subsetneq \mathbb{I} e_k$ for all $k \in \mathbb{N}$. Similarly, we have $e_{k-1}\, \mathbb{I} \subsetneq e_k\, \mathbb{I}$. Hence the increasing sequence $(I_k = \mathbb{I}\, e_k)_{k \geq 0}$ (resp., $(I_k = e_k\, \mathbb{I})_{k \geq 0}$) of principal left (resp., right) ideals of $\mathbb{I}$ is not stationary, which proves $\mathbb{I}$ is not a left (resp., a right) Noetherian ring.

Even though $\mathbb{I}$ is non-Noetherian, Bavula proved the following fundamental result stating that $\mathbb{I}$ is a *coherent ring*.

Theorem 2 ([10]) *The ring $\mathbb{I}$ is* coherent*, i.e., for every $r \geq 1$, and for all $d_1, \ldots, d_r \in \mathbb{I}$, the left (resp., right) $\mathbb{I}$-module*

$$S = \left\{ (c_1, \ldots, c_r) \in \mathbb{I}^{1 \times r} \mid \sum_{i=1}^{r} c_i\, d_i = 0 \right\}$$

(resp., $S = \left\{ (c_1, \ldots, c_r)^T \in \mathbb{I}^{r \times 1} \mid \sum_{i=1}^{r} c_i\, e_i = 0 \right\}$) is finitely generated as a left (resp., right) $\mathbb{I}$-module.

Linear systems are usually described by means of finite matrices with entries in a certain ring of functional operators $\mathcal{D}$. As explained in [35], if $\mathcal{D}$ is a coherent ring, an algebraic systems theory can be developed as if $\mathcal{D}$ were a Noetherian ring. Hence, Theorem 2 shows that an algebraic systems theory can be developed over $\mathbb{I}$. In particular, basic module-theoretic operations of *finitely presented* left/right $\mathbb{I}$-

modules namely, left/right $\mathbb{I}$-modules defined by matrices, are finitely presented, and thus, finitely generated. For more details, see, e.g., [28, 44]. It is shown in [11] that Theorem 2 cannot be generalized for more than one differential operator, i.e., for the algebra $\mathbb{I}_n$ of integro-partial differential operators with polynomial coefficients defined by the operators x_i, $\partial_i = \frac{\partial}{\partial x_i}$ and $\int^{x_i}$ for $i = 1, \ldots, n$ and $n > 1$.

Based on normal forms for generalized Weyl algebras, it is shown in [8] that $\mathbb{I}$ admits the *involution* θ defined by

$$\theta(\partial) = \int, \quad \theta(\int) = \partial, \quad \theta(t) = t\,\partial^2 + \partial = (t\,\partial + 1)\,\partial, \tag{3.10}$$

i.e., θ is a k-linear *anti-automorphism*, namely, it satisfies:

$$\forall\, d,\, e \in \mathbb{I}, \quad \theta(d\,e) = \theta(e)\,\theta(d), \quad \theta^2(d) = d.$$

We note that $\partial \int = 1$ and $\mathbf{E} = 1 - \int \partial$ yield:

$$\theta(1) = \theta(\textstyle\int)\,\theta(\partial) = \partial \int = 1, \quad \theta(\mathbf{E}) = \theta(1) - \theta(\partial)\,\theta(\textstyle\int) = 1 - \int \partial = \mathbf{E}.$$

With the notations (3.7) and (3.8), we get:

$$\begin{cases} \theta(d_1) = \sum a_{ij}\,\theta(\partial)^j\,\theta(t)^i = \sum a_{ij}\,\int^j ((t\,\partial + 1)\,\partial)^i, \\ \theta(d_2) = \sum b_{ij}\,\theta(t)^j\,\theta(\int)\,\theta(t)^i = \sum b_{ij}\,((t\,\partial + 1)\,\partial)^j\,\partial\,((t\,\partial + 1)\,\partial)^i, \\ \theta(d_3) = \displaystyle\sum_{\alpha\in\Phi} \Big(\sum f_{ij}\,\theta(\partial)^j\,\theta(\mathbf{E})\,\theta(t)^i\Big) = \sum_{\alpha\in\Phi} \Big(\sum f_{ij}\,\int^j \mathbf{E}\,((t\,\partial + 1)\,\partial)^i\Big) \\ \qquad = \displaystyle\sum_{\alpha\in\Phi} \Big(\sum f_{ij}\,\frac{t^j}{j!}\,\mathbf{E}\,((t\,\partial + 1)\,\partial)^i\Big). \end{cases}$$

In particular, we have $\theta((\mathbf{E})) \subseteq (\mathbf{E})$ and $\theta(k[t]\int k[t]) \subseteq A$. Finally, we note that:

$$\theta(t\,\partial) = \textstyle\int (t\,\partial + 1)\,\partial = t\,\partial.$$

As a consequence, many algebraic properties of left $\mathbb{I}$-modules have a right analogue and conversely. Finally, in [8–10], various algebraic properties of $\mathbb{I}$ and important results are proven amongst them a classification of *simple modules*, an analogue of *Stafford's theorem*, and of the *first conjecture of Dixmier*.

The computation of *syzygies*, namely, left/right kernel of a matrix with entries in $\mathbb{I}$ is a central task towards developing an algorithmic approach to linear systems of ID equations with boundary conditions based on module theory and homological algebra. See [16, 29, 36] and references therein. However, the the proof of Theorem 2 given in [10] is non-constructive. As a first step for computing syzygies, we discuss in the following how to find left/right annihilators of elements in $\mathbb{I}$. As we will see, this

problem leads, in turn, to computing polynomial solutions of ordinary ID equations with boundary conditions, which we discuss in Sect. 3.7.

The *left annihilator* of $d \in \mathbb{I}$ is defined by

$$\operatorname{ann}_{\mathbb{I}}(.d) := \{e \in \mathbb{I} \mid e\,d = 0\},$$

and, analogously, the *right annihilator* is defined by:

$$\operatorname{ann}_{\mathbb{I}}(d.) := \{e \in \mathbb{I} \mid d\,e = 0\}.$$

The left annihilator can be interpreted as *compatibility conditions* of the inhomogeneous ID equation $d\,y(t) = u(t)$. Indeed, for $e \in \operatorname{ann}_{\mathbb{I}}(.d)$, we have:

$$e\,u(t) = e\,d\,y(t) = 0.$$

If d is not a zero divisor, then $d\,y = u$ does not admit compatibility condition of the form $e\,u = 0$, where $e \in \mathbb{I}$.

Example 1 We first consider the following trivial example:

$$\textstyle\int_0^t y(s)\,ds = u(t).$$

The compatibility condition $u(0) = 0$ corresponds to the left annihilator $\mathbf{E}$ of $\int$, i.e., $\mathbf{E}\int = 0$ in $\mathbb{I}$. As a nontrivial example, we consider the inhomogeneous ID equation:

$$t^2\,\ddot{y}(t) - 2\,t\,\dot{y}(t) + (t+2)\,y(t) - (3\,t/5+2)\textstyle\int_0^t y(s)\,ds + 3/5\int_0^t s\,y(s)\,ds = u(t). \quad (3.11)$$

The left annihilator of the following ID operator

$$d = t^2\,\partial^2 - 2\,t\,\partial + (t+2) - (3\,t/5+2)\textstyle\int + 3/5\int t \in \mathbb{I} \quad (3.12)$$

yields the compatibility conditions of (3.11). The compatibility conditions of d will be given in Example 9.

The relation between annihilators and polynomial solutions of ordinary ID equations comes from the fact that we can identify an integro-differential operator $d \in \mathbb{I}$ with the corresponding linear map L_d on the polynomial ring $k[t]$. Hence, we have the equivalences:

$$d\,e = 0 \;\Leftrightarrow\; L_{d\,e} = L_d \circ L_e = 0 \;\Leftrightarrow\; \operatorname{im} L_e \subseteq \ker L_d. \quad (3.13)$$

Suppose that we want to compute the right annihilator of d and assume that L_d has a finite dimensional kernel. Then the image of L_e for an $e \in \operatorname{ann}_{\mathbb{I}}(d.)$ has to be finite dimensional and must be contained in $\ker L_d$. In other words, we have to compute the polynomial solutions of L_d and then find generators for all ID operators e with

$\operatorname{im} L_e \subseteq \ker L_d$. After discussing some general properties of Fredholm and finite-rank operators in the next section, we follow this strategy for ID operators including several evaluations in Sect. 3.8.

3.6 Fredholm and Finite-Rank Operators

Several properties of *Fredholm operators* can be studied in the purely algebraic setting of linear maps on infinite-dimensional vector spaces. In [11], such properties are used to investigate $\mathbb{I}$. It turns out that Fredholm operators are also very useful for an algorithmic approach to operator algebras. We review some algebraic properties of Fredholm operators in this section.

Definition 5 A k-linear map $f: V \longrightarrow W$ between two k-vector spaces is called *Fredholm* if it has finite dimensional kernel and cokernel, where $\operatorname{coker} f = W/\operatorname{im} f$. The *index* of a Fredholm operator f is defined by:

$$\operatorname{ind}_k f = \dim_k(\ker f) - \dim_k(\operatorname{coker} f).$$

We have the *long exact sequence* of k-vector spaces [44]

$$0 \longrightarrow \ker f \xrightarrow{i} V \xrightarrow{f} W \xrightarrow{p} \operatorname{coker} f \longrightarrow 0,$$

i.e., i is injective, $\ker f = \operatorname{im} i$, $\ker p = \operatorname{im} f$, and p is surjective, where $p(w)$ is the residue class of $w \in W$ in $\operatorname{coker} f$. Then, $\dim_k(\operatorname{coker} f)$ gives the number of independent k-linear compatibility conditions $g(w) = 0$ on w for the solvability of the inhomogeneous linear system $f(v) = w$ (e.g., f is surjective if and only if $\operatorname{coker} f = 0$), while $\dim_k(\ker f)$ measures the degrees of freedom in a solution ($v + u$ is solution for all $u \in \ker f$).

Example 2 Viewing the basic operators $1, t, \partial, \int \in \mathbb{I}$ as k-linear maps on $V = W = k[t]$, we get:

$$\begin{aligned} &\ker 1 = \ker t = \ker \textstyle\int = 0, && \ker \partial = k, \\ &\operatorname{im} 1 = \operatorname{im} \partial = k[t], && \operatorname{im} t = \operatorname{im} \textstyle\int = k[t]\, t. \end{aligned}$$

Hence, they are also Fredholm with index:

$$\operatorname{ind}_k 1 = 0, \quad \operatorname{ind}_k t = \operatorname{ind}_k \textstyle\int = -1, \quad \operatorname{ind}_k \partial = 1.$$

If V and W are two finite-dimensional k-vector spaces, then

$$\dim_k(\operatorname{coker} f) = \dim_k(W) - \dim_k(\operatorname{im} f)$$

and *the rank-nullity theorem* yields $\dim_k V = \dim_k(\operatorname{im} f) + \dim_k(\ker f)$, hence

$$\operatorname{ind}_k f = \dim_k V - \dim_k W, \tag{3.14}$$

i.e., $\operatorname{ind}_k f$ depends only on the dimensions of V and W.

We also recall the index formula for Fredholm operators.

Proposition 3 *Let $V' \xrightarrow{f} V \xrightarrow{g} V''$ be k-linear maps between k-vector spaces. If two of the maps f, g, and $g \circ f$ are Fredholm, then so is the third, and:*

$$\operatorname{ind}_k(g \circ f) = \operatorname{ind}_k g + \operatorname{ind}_k f. \tag{3.15}$$

Proof Considering the following commutative square

$$\begin{array}{ccc} V' & \xrightarrow{f} & V \\ \| & & \downarrow{\scriptstyle g} \\ V' & \xrightarrow{g \circ f} & V'', \end{array}$$

we obtain the following *commutative exact diagram* (see, e.g., [44]):

$$\begin{array}{ccccccccc} & & & & 0 & & & & \\ & & & & \downarrow & & & & \\ & & & & \ker g & & & & \\ & & & & \downarrow & & & & \\ 0 & \longrightarrow & \ker f & \longrightarrow & V' & \xrightarrow{f} & V & \longrightarrow & \operatorname{coker} f & \longrightarrow & 0 \\ & & & & \| & & \downarrow{\scriptstyle g} & & & & \\ 0 & \longrightarrow & \ker(g \circ f) & \longrightarrow & V' & \xrightarrow{g \circ f} & V'' & \longrightarrow & \operatorname{coker}(g \circ f) & \longrightarrow & 0. \\ & & & & \downarrow & & \downarrow & & & & \\ & & & & 0 & & \operatorname{coker} g & & & & \\ & & & & & & \downarrow & & & & \\ & & & & & & 0 & & & & \end{array}$$

A *chase* in the above commutative exact diagram shows that we have the following long exact sequence of finite-dimensional k-vector spaces [44]:

$$\begin{aligned} 0 \longrightarrow \ker f \longrightarrow \ker(g \circ f) \longrightarrow \ker g \longrightarrow \\ \operatorname{coker} f \longrightarrow \operatorname{coker}(g \circ f) \longrightarrow \operatorname{coker} g \longrightarrow 0. \end{aligned}$$

Using the *Euler-Poincaré characteristic* [44], we then get

$$\begin{aligned}&\dim_k(\ker f) - \dim_k(\ker(g \circ f)) + \dim_k(\ker g)\\ &- \dim_k(\operatorname{coker} f) + \dim_k(\operatorname{coker}(g \circ f)) - \dim_k(\operatorname{coker} g) = 0,\end{aligned}$$

which finally proves (3.15).

Definition 6 A k-linear map between two k-vector spaces is called *finite-rank* if its image is finite-dimensional.

Example 3 Let us consider $\mathbf{E} = 1 - \int\partial \in \mathbb{I}$. It has an infinite-dimensional kernel $\ker_k \mathbf{E} = k[t]\,t$, but its image $\operatorname{im}_k \mathbf{E} = k$ is one-dimensional. More generally, every boundary operator $d_3 \in \mathbb{I}_\Phi$ is obviously of finite rank since its image is contained in the k-vector space of polynomials with degree less than or equal n, where n is the maximal index i with a nonzero coefficient f_{ij} or g_{ij} in (3.8).

Clearly, composing a finite-rank map with a linear map from either side gives again finite-rank map and Proposition 3 shows that the composition of two Fredholm operators is a Fredholm operator.

Proposition 4 *Let V be a k-vector space and $\mathcal{A}$ a k-subalgebra of* $\operatorname{end}_k(V)$. *Then,*

$$\mathcal{F}_\mathcal{A} = \{a \in \mathcal{A} \mid a \text{ is Fredholm}\}$$

forms a monoid and

$$\mathcal{C}_\mathcal{A} = \{c \in \mathcal{A} \mid c \text{ is finite-rank }\}$$

is a two-sided ideal of $\mathcal{A}$.

In particular, we have another interpretation of the only proper two-sided ideal $(\mathbf{E})$ of boundary operators as finite-rank operators. All other ID operators of $\mathbb{I}_\Phi \setminus (\mathbf{E})$ are Fredholm as we will see in Proposition 6. More generally, the notion of *(strong) compact-Fredholm alternative* for an arbitrary k-algebra $\mathcal{A}$ was introduced in [10].

3.7 Polynomial Solutions of Rational Indicial Maps and Polynomial Index

Computing polynomial solutions of linear systems of OD is well-studied in symbolic computation since it appears as a subproblem of many important algorithms. See, for example, [1–5, 14]. In this section, we discuss an algebraic setting and an algorithmic approach for the computation of polynomial solutions (kernel), cokernel, and the "polynomial" index for a general class of linear operators including ID operators.

For computing the kernel and cokernel of a k-linear map $L : V \longrightarrow V'$ on infinite-dimensional k-vector spaces V and V', we can use the following simple consequence of the *snake lemma* in homological algebra (see, e.g., [44]).

Lemma 2 *Let $L\colon V \longrightarrow V'$ be a k-linear map and $U \subseteq V$, $U' \subseteq V'$ k-subspaces such that $L(U) \subseteq U'$. Let*

$$L' = L_{|U}\colon U \longrightarrow U' \quad \text{and} \quad \overline{L}\colon V/U \longrightarrow V'/U'$$

be the induced k-linear map defined by $\overline{L}(\pi(v)) = \pi'(L(v))$ for all $v \in V$, where $\pi\colon V \longrightarrow V/U$ (resp., $\pi'\colon V' \longrightarrow V'/U'$) is the canonical projection onto V/U (resp., V'/U'). Then, we have the following commutative exact diagram:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & U & \longrightarrow & V & \xrightarrow{\pi} & V/U & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle L'} & & \downarrow{\scriptstyle L} & & \downarrow{\scriptstyle \overline{L}} & & \\ 0 & \longrightarrow & U' & \longrightarrow & V' & \xrightarrow{\pi'} & V'/U' & \longrightarrow & 0. \end{array} \tag{3.16}$$

If $\overline{L}$ is an isomorphism, i.e., $V/U \cong V'/U'$, then:

$$\ker L' = \ker L, \quad \operatorname{coker} L' \cong \operatorname{coker} L.$$

Moreover, if U and U' are two finite-dimensional k-vector spaces, then L is Fredholm and $\operatorname{ind}_k L = \dim_k U - \dim_k U'$.

Proof Since $\overline{L}$ is an isomorphism, applying the standard the snake lemma (see, e.g., [44]) to the following commutative exact diagram of k-vector spaces

$$\begin{array}{ccccccccc} & & 0 & & 0 & & & & \\ & & \downarrow & & \downarrow & & & & \\ & & \ker L' & & \ker L & & 0 & & \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & U & \longrightarrow & V & \xrightarrow{\pi} & V/U & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle L'} & & \downarrow{\scriptstyle L} & & \downarrow{\scriptstyle \overline{L}} & & \\ 0 & \longrightarrow & U' & \longrightarrow & V' & \xrightarrow{\pi'} & V'/U' & \longrightarrow & 0, \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & \operatorname{coker} L' & & \operatorname{coker} L & & 0 & & \\ & & \downarrow & & \downarrow & & & & \\ & & 0 & & 0 & & & & \end{array}$$

we obtain the following long exact sequence of k-vector spaces

$$0 \longrightarrow \ker L' \longrightarrow \ker L \longrightarrow 0 \longrightarrow \operatorname{coker} L' \longrightarrow \operatorname{coker} L \longrightarrow 0,$$

and the statements about the kernel and cokernel follow. If U and U' are two finite-dimensional k-vector spaces, then so are $\ker L' = \ker L$ and $\operatorname{coker} L' \cong \operatorname{coker} L$ and $\operatorname{ind}_k L = \operatorname{ind}_k L' = \dim_k U - \dim_k U'$ by (3.14). □

Remark 1 In the language of homological algebra, the fact that $\overline{L}$ defines an isomorphism in Lemma 2 means that the following chain complex of k-vector spaces

$$\begin{array}{ccccccc} 0 & \longrightarrow & U & \longrightarrow & V & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle L'} & & \downarrow{\scriptstyle L} & & \\ 0 & \longrightarrow & U' & \longrightarrow & V' & \longrightarrow & 0 \end{array}$$

is a *quasi-isomorphism*, namely the homologies of the horizontal complexes, i.e., V/U and V'/U', are isomorphic. Hence, the complex $0 \longrightarrow V \xrightarrow{L} V' \longrightarrow 0$ of infinite-dimensional k-vector spaces, whose homologies are $\ker L$ and $\operatorname{coker} L$, is then *reduced* to the complex $0 \longrightarrow U \xrightarrow{L'} U' \longrightarrow 0$ of finite-dimensional k-vector spaces, which homologies, $\ker L'$ and $\operatorname{coker} L'$, are then isomorphic to $\ker L$ and $\operatorname{coker} L$.

From an algorithmic point of view, we want to find finite-dimensional k-subspaces U and U', and an algorithmic criterion for $\overline{L}$ being an isomorphism on the remaining infinite-dimensional parts V/U and V'/U'.

The cokernel of a k-linear map $f : V \longrightarrow W$ between two finite-dimensional k-vector spaces V and W can be characterized as follows. Choosing bases of V and W, there exists a matrix $C \in k^{m \times n}$ such that $f(v) = C\,v$ for all $v \in V \cong k^n$. Computing a basis of the finite-dimensional k-vector space $\ker C^T$ and stacking the elements of this basis into a matrix $D \in k^{l \times m}$, we get $\ker C^T = \operatorname{im} D^T$. Then, $\operatorname{coker} f \cong \operatorname{im} D$ and, more precisely, if $\pi : W \longrightarrow \operatorname{coker} f$ is the canonical projection onto $\operatorname{coker} f$, then the k-linear map $\sigma : \operatorname{coker} f \longrightarrow \operatorname{im} D$ defined by $\sigma(\pi(w)) = D\,w$ for all $w \in W$, is an isomorphism of k-vector spaces.

Let us now study when the k-linear map $\overline{L} : V/U \longrightarrow V'/U'$ is an isomorphism. In what follows, we will focus on the polynomial case, namely, $V = V' = k[t]$. To do that, let us introduce the degree filtration of $k[t]$, namely,

$$k[t] = \bigcup_{i \in \mathbb{N}} k[t]_{\leq i}, \quad k[t]_{\leq i} = \bigoplus_{j=0}^{i} k\,t^j,$$

defined by the finite-dimensional k-vector spaces $k[t]_{\leq i}$ formed by the polynomials of $k[t]$ of degree less than or equal to i (we set $k[t]_{\leq -1} = 0$). Note that this filtration is induced by any basis $\{p_i\}_{i \in \mathbb{N}}$ of $k[t]$ with $\deg p_i = i$ for all $i \in \mathbb{N}$.

For motivating the following definition, we recall that we defined the multiplication operator, derivation, and integral operator in terms of their action on the basis $(t^n)_{n \in \mathbb{N}}$ of $k[t]$; see Eq. (3.1) and Definition 1. More generally, we can easily check that the action of the summands of an ID operator in the normal form (3.7) is respec-

tively given by:

$$\begin{aligned}(t^i\,\partial^j)(t^n) &= \frac{n!}{(n-j)!}\,t^{n-j+i}, && n \geq j,\\ (t^i\,\partial^j)(t^n) &= 0, && n < j,\\ (t^i \int t^j)(t^n) &= \frac{1}{n+j+1}\,t^{i+j+n+1}.\end{aligned}$$

So the action on a basis element t^n for n large enough is given by a rational function in the exponent n and a shift in the exponent.

Definition 7 A k-linear map $L\colon k[t] \longrightarrow k[t]$ is called *rational indicial* if there exist a nonzero rational function $q \in k(n)$, an integer $s \in \mathbb{Z}$, s a bound $M \in \mathbb{N}$, and nonzero constants $c_n \in k^*$ such that

$$L(t^n) = c_n\,q(n)\,t^{n+s} + \text{lower degree terms},$$

for all $n \geq M \geq -s$. Then, we call the pair

$$\operatorname{rsym}(L) = (s,\ q)$$

its *rational symbol*.

Example 4 The rational symbols of the defining ID operators are:

$$\begin{aligned}\operatorname{rsym}(1) &= (0,\ 1), & \operatorname{rsym}(t) &= (1,\ 1),\\ \operatorname{rsym}(\partial) &= (-1,\ n), & \operatorname{rsym}(\textstyle\int) &= \left(1,\ \frac{1}{n+1}\right).\end{aligned}$$

Operators such as shift and dilation operators on $k[t]$ are also rational indicial. For instance, if $a \in k \setminus \{0\}$ and χ_a is the dilation operator defined by $\chi_a(t^n) = (a\,t)^n$ for all $n \geq 0$, then we get $c_n = a^n$, $q = 1$, $s = 0$, and $M = 0$.

Example 5 The sum of a rational indicial map and a finite-rank map is also rational indicial with the same symbol for a large enough bound M. For instance, if we consider $L_1 = 1 + t^3\,\mathbf{E}_0$, then we have $L_1(1) = t^3 + 1$ and $L_1(t^n) = t^n$ for $n \geq 1$, which shows that $M = 1$, $s = 0$, $q = 1$, and $c_n = 1$ (compare with $L_0 = 1$ which is such that $M = 0$, $s = 0$, $c_n = 1$, and $q = 1$). Finally, if we consider $L_2 = 1 + t^3\,\mathbf{E}_0\,\partial^2$, then we have $L_2(1) = 1$, $L_2(t) = t$, $L_2(t^2) = 2\,t^3 + t^2$, and $L_2(t^n) = t^n$ for $n \geq 3$, which shows that $M = 3$, $s = 0$, $q = 1$, and $c_n = 1$.

Let us now state a result for the computation of the kernel and cokernel of rational indicial maps (compare with Lemma 6.5 of [10]).

Proposition 5 *Let $L\colon k[t] \longrightarrow k[t]$ be a k-linear map. Let*

$$-1 \leq N,\ -(N+1) \leq s, \quad U = k[t]_{\leq N}, \quad U' = k[t]_{\leq N+s}$$

be such that $L(U) \subseteq U'$. *Let* $L' = L_{|U} : U \longrightarrow U'$ *be the induced map. If* $\deg L(t^n) = n + s$ *for all* $n \geq N + 1$, *then:*

$$\ker L' = \ker L, \quad \operatorname{coker} L' \cong \operatorname{coker} L.$$

Moreover, L *is a Fredholm operator with* $\operatorname{ind}_k L = -s$.

Proof Let $V = V' = k[t]$ and $\pi : V \longrightarrow V/U$ (resp., $\pi' : V' \longrightarrow V'/U'$) be the canonical projection onto V/U (resp., V'/U'). Then, $\overline{L}(\pi(t^n)) = \pi'(L(t^n))$ for all $n \in \mathbb{N}$.

Let us note $T_n = \pi(t^n)$ and $S_n = \pi'(t^n)$ for all $n \geq 0$. Then, we get:

$$V/U = k[t]/k[t]_{\leq N} = \bigoplus_{i \geq N+1} k\, T_i, \quad V'/U' = k[t]/k[t]_{\leq N+s} = \bigoplus_{i \geq N+s+1} k\, S_i.$$

Moreover, if $p = \sum_{i=N+1}^{N+r} p_i\, t^i \in k[t]$, where $p_i \in k$, then we have

$$L(p) = \sum_{i=N+1}^{N+r} p_i\, L(t^i) = \sum_{i=N+1}^{N+r} p_i\, (c_i\, q(i)\, t^{i+s} + \ldots) = \sum_{i=N+1}^{N+r} p_i\, c_i\, q(i)\, t^{i+s} + \ldots,$$

where $\ldots$ denotes lower degree terms. Note that we have $\pi(p) = \sum_{i=N+1}^{N+r} p_i\, T_i$ and $\pi'(L(p)) = \sum_{i=N+1}^{N+r} p_i\, c_i\, q(i)\, S_{i+s} + \ldots$, which shows that $\overline{L}$ corresponds to the following linear operator:

$$\begin{array}{rcl} \overline{L} : V/U = \bigoplus_{i \geq N+1} k\, T_i & \longrightarrow & V'/U' = \bigoplus_{i \geq N+s+1} k\, S_i \\ \sum_{i=N+1}^{N+r} p_i\, T_i & \longmapsto & \sum_{i=N+1}^{N+r} p_i\, c_i\, q(i)\, S_{i+s} + \ldots \end{array}$$

Considering the coefficients of the elements of V/U (resp., V'/U') in the basis $\{T_i\}_{i \geq N+1}$ (resp., $\{S_j\}_{j \geq N+s+1}$), up to isomorphism of k-vector spaces, we obtain:

$$\begin{array}{rcl} \overline{L} : \bigoplus_{i \geq N+1} k & \longrightarrow & \bigoplus_{i \geq N+s+1} k \\ (p_{N+1},\ p_{N+2}, \ldots, p_{N+r}, 0, \ldots) & \longmapsto & (p_{N+1}\, c_{N+1}\, q(N+1) + \ldots, \\ & & \ldots, p_{N+r}\, c_{N+r}\, q(N+r) + \ldots, 0, \ldots). \end{array}$$

We note that $\overline{L}$ is defined by an upper triangular infinite matrix which determinant is $\prod_{i=N+1}^{N+r} c_i\, q(i) \neq 0$. Hence, the linear operator $\overline{L}$ is invertible, and thus defines an isomorphism of k-vector spaces, i.e. $V/U \cong V'/U'$.

Finally, the result follows from Lemma 2 after noting that:

$$\dim_k U - \dim_k U' = N + 1 - (N + 1 + s) = -s.$$

□

Example 6 Let us consider the Fredholm operator $L = t\int + \int t$. Then, we get $L(t^n) = \left(\frac{1}{n+1} + \frac{1}{n+2}\right) t^{n+2}$ for all $n \geq 0$, which shows that $\operatorname{rsym}(L) = \left(2, \frac{2n+3}{(n+1)(n+2)}\right)$. Hence, if we consider $N = 0$, $s = 2$, $V = V' = k[t]$, $U = k$, $U' = k[t]_{\leq 2}$, and $L' = L_{|U}$, i.e., $L'(u) = 3\,u\,t^2/2$ for all $u \in k$, then $\ker L' = 0$ and $\operatorname{coker} L' = k[t]_{\leq 2}/(t^2) \cong k + k\,t$. Let us note $T_i = \pi(t^i)$ and $S_i = \pi'(t^i)$ for all $i \in \mathbb{N}$. If $p = \sum_{i=0}^{r} p_i\,t^i \in k[t]$, then using that $q(i) \neq 0$ for all $i \in \mathbb{N}$, we obtain the following isomorphism of k-vector spaces

$$\begin{aligned} \overline{L} : V/U = k[t]/k = \textstyle\bigoplus_{i\geq 1} k\,T_i &\longrightarrow V'/U' = k[t]/k[t]_{\leq 2} = \textstyle\bigoplus_{i\geq 3} k\,S_i \\ \pi(p) = \textstyle\sum_{i\geq 1}^{r} p_i\,T_i &\longmapsto \pi'(L(p)) = \textstyle\sum_{i\geq 1}^{r} p_i\,q(i)\,S_{i+2}, \end{aligned}$$

which, up to isomorphism, corresponds to the isomorphism of k-vector spaces:

$$(p_1, \ldots, p_r, 0, \ldots) \longmapsto (p_1\,q(1), \ldots, p_r\,q(r), 0, \ldots).$$

By Proposition 5, we obtain $\ker L = \ker L' = 0$ and $\operatorname{coker} L \cong \operatorname{coker} L' \cong k + k\,t$.

Similarly, we let the reader compute the polynomial solutions of $L = \frac{2}{3}\,t\int - \int t$.

Given a rational indicial operator with rational symbol $(s,\ q)$ and bound M, we obtain a bound N for Proposition 5 by computing the largest nonnegative integer root l of q and taking $N = \max(l, M)$. Hence computing the kernel and cokernel of $L\colon k[t] \longrightarrow k[t]$ reduces to the same problem for the k-linear map $L' = L_{|U}\colon U \longrightarrow U'$ between two finite-dimensional k-vector spaces, which can be solved using basic linear algebra techniques. We have implemented in `Maple` the computation of kernel and cokernel of rational indicial maps.

Corollary 3 *A rational indicial operator with rational symbol $(s,\ q)$ is Fredholm with index $-s$ and its kernel and cokernel can be effectively computed.*

We can explicitly compute the rational symbol $(s,\ q)$ for $d \notin (\mathbf{E})$ from its normal form. For computing the index of OD equations with analytic coefficients, we have the Komatsu–Malgrange index theorem [25, 30]. The following proposition is a purely algebraic version of an *index theorem*. Compare with [10, Proposition 6.1].

Proposition 6 *Let $d = \sum a_{ij}\,t^i\,\partial^j + \sum b_{ij}\,t^i\int t^j + d_3 \in \mathbb{I}_\Phi$ be an ID operator, where $d_3 \in (\mathbf{E})$, such that $d \notin (\mathbf{E})$. Then, the k-linear map*

$$\begin{aligned} L_d\colon k[t] &\longrightarrow k[t], \\ p &\longmapsto d(p), \end{aligned}$$

is rational indicial with rational symbol (s, q) given by

$$s = -\operatorname{ind}_k d = \max(\{i - j \mid a_{ij} \neq 0\} \cup \{i + j + 1 \mid b_{ij} \neq 0\}),$$

and:

$$q(n) = \sum_{i-j=s} a_{ij} \frac{n!}{(n-j)!} + \sum_{i+j+1=s} b_{ij} \frac{1}{n+j+1}.$$

3.8 Polynomial Solutions and Annihilators

In his proof of Theorem 2, stating that $\mathbb{I}$ is a coherent ring, Bavula [10] uses that the left and right annihilators are finitely generated $\mathbb{I}$-modules, for which a non-constructive argument is given.

Theorem 3 ([10]) *Let $d \in \mathbb{I}$. Then, the left (resp., right) annihilator* $\operatorname{ann}_{\mathbb{I}}(.d)$ *(resp.,* $\operatorname{ann}_{\mathbb{I}}(d.)$*) of d is a finitely generated left (resp., right) $\mathbb{I}$-module.*

In this section, we generalize this result to right annihilators of Fredholm operators $d \in \mathbb{I}_\Phi$ with several evaluations using a constructive approach. As outlined at the end of Sect. 3.5, our approach is based on the fact that we can identify integro-differential operators with the corresponding linear map on the polynomial ring (see Corollary 2). To characterize the right annihilator $\operatorname{ann}_{\mathbb{I}_\Phi}(d.)$, we use the equivalences (3.13). If d is Fredholm, i.e., $d \in \mathbb{I}_\Phi \setminus (\mathbf{E})$, then $\ker L_d$ is a finite-dimensional k-vector space, and thus, e has to be finite-rank and hence must be a boundary operator $e \in (\mathbf{E})$. Thus, we have to compute polynomial solutions of the Fredholm operator d, i.e., $\ker L_d$, and then find generators for all the e's satisfying $\operatorname{im} L_e \subseteq \ker L_d$.

We first describe the image of a finite-rank operator L_e for a boundary operator $e \in (\mathbf{E})$. By (3.8), e is a finite $k[t]$-linear combination of terms of the form $\mathbf{E}_\alpha \, \partial^i$ and $\mathbf{E}_\alpha \int t^i$ with $\alpha \in \Phi$, namely

$$e = \sum_{\alpha \in \Phi} \left(\sum_{i=0}^{l} p_{\alpha,i} \, \mathbf{E}_\alpha \, \partial^i + \sum_{i=0}^{m} q_{\alpha,i} \, \mathbf{E}_\alpha \int t^i \right), \tag{3.17}$$

where $p_{\alpha,i}, \, q_{\alpha,i} \in k[t]$. With Lemma 1, we can now apply the following general fact for linear functionals on arbitrary vector spaces; see, e.g., [27, pp. 71–72].

Lemma 3 *Let V be a k-vector space and $\lambda_1, \ldots, \lambda_n \in V^*$ k-linear functionals. Then, the λ_i are k-linearly independent iff there exist $v_1, \ldots, v_n \in V$ such that:*

$$\forall \, i, j = 1, \ldots, n, \quad \lambda_i(v_j) = \delta_{ij}.$$

Proposition 7 *Let $e \in (\mathbf{E})$ be as in (3.17). Then, we have:*

$$\operatorname{im} L_e = \sum_{\alpha \in \Phi} \sum_{i=0}^{l} k \, p_{\alpha,i} + \sum_{\alpha \in \Phi} \sum_{i=0}^{m} k \, q_{\alpha,i}.$$

Proof The inclusion $\subseteq$ is obvious since $\mathbf{E}_\alpha \, \partial^i$ and $\mathbf{E}_\alpha \int t^i$ are functionals. Let $\mathbf{E}_\alpha \, \partial^i$ or $\mathbf{E}_\alpha \int t^i$ be a linear functional corresponding to a nonzero summand in (3.17).

Since these linear functional forms are k-linearly independent by Lemma 1, using Lemma 3 with $V = k[t]$, there exists a polynomial $p \in k[t]$ such that $(\mathbf{E}_\alpha\, \partial^i)(p) = 1$ (resp., $(\mathbf{E}_\alpha \int t^i)(p) = 1$) and $(\mathbf{E}_\beta\, \partial^j)(p) = 0$ (resp., $(\mathbf{E}_\beta \int t^j)(p) = 0$) for all other functionals corresponding to nonzero summands of (3.17). Then, we get $L_e(p) = e(p) = p_{\alpha,i}$ or $L_e(p) = e(p) = q_{\alpha,i}$, which proves the reverse inclusion. □

Theorem 4 *Let Φ be a subset of k with $0 \in \Phi$. Let $d \in \mathbb{I}_\Phi$ be Fredholm with*

$$\ker L_d = \sum_{i=1}^{n} k\, r_i,$$

where $r_i \in k[t]$. Then, we have:

$$\operatorname{ann}_{\mathbb{I}_\Phi}(d.) = \sum_{i=1}^{n} (r_i\, \mathbf{E})\, \mathbb{I}_\Phi.$$

In particular, $\operatorname{ann}_{\mathbb{I}_\Phi}(d.)$ *is a finitely generated right $\mathbb{I}_\Phi$-module.*

Proof Since $\operatorname{im} L_{r_i\,\mathbf{E}} = k\, r_i \subseteq \ker L_d$, the inclusion $\supseteq$ follows by (3.13). Conversely, let $e \in \mathbb{I}_\Phi$ as in (3.17) with $d\, e = 0$. Then, by (3.13) and Proposition 7, we have:

$$\operatorname{im} L_e = \sum_{\alpha\in\Phi} \sum_{i=0}^{l} k\, p_{\alpha,i} + \sum_{\alpha\in\Phi} \sum_{i=0}^{m} k\, q_{\alpha,i} \subseteq \ker L_d = \sum_{i=1}^{n} k\, r_i.$$

Hence, every nonzero $p_{\alpha,i}$ and $q_{\alpha,i}$ can be written as a k-linear combination of the r_i's, i.e., $p_{\alpha,i} = \sum_{j=1}^{n} u_{\alpha,i,j}\, r_j$ and $q_{\alpha,i} = \sum_{j=1}^{n} v_{\alpha,i,j}\, r_j$ for certain $u_{\alpha,i,j},\ v_{\alpha,i,j} \in k$. Using (3.17) and $\mathbf{E}\,\mathbf{E}_\alpha = \mathbf{E}_\alpha$, we then get

$$\begin{aligned}
e &= \sum_{\alpha\in\Phi} \sum_{j=1}^{n} \left(\sum_{i=0}^{l} u_{\alpha,i,j}\, r_j\, \mathbf{E}_\alpha\, \partial^i + \sum_{i=0}^{m} v_{\alpha,i,j}\, r_j\, \mathbf{E}_\alpha \int t^i \right) \\
&= \sum_{\alpha\in\Phi} \sum_{j=1}^{n} \left(\sum_{i=0}^{l} u_{\alpha,i,j}\, r_j\, \mathbf{E}\,\mathbf{E}_\alpha\, \partial^i + \sum_{i=0}^{m} v_{\alpha,i,j}\, r_j\, \mathbf{E}\,\mathbf{E}_\alpha \int t^i \right) \\
&= \sum_{j=1}^{n} r_j\, \mathbf{E} \left(\sum_{\alpha\in\Phi} \sum_{i=0}^{l} v_{\alpha,i,j}\, \mathbf{E}_\alpha\, \partial^i + \sum_{\alpha\in\Phi} \sum_{i=0}^{m} u_{\alpha,i,j}\, \mathbf{E}_\alpha \int t^i \right) \in \sum_{j=1}^{n} (r_j\, \mathbf{E})\, \mathbb{I}_\Phi,
\end{aligned}$$

which proves the reverse inclusion $\subseteq$ and thus the result. □

Example 7 If $d = \partial^2$, then we have $\ker L_d = k + k\, t$, which shows that $\operatorname{ann}_{\mathbb{I}_\Phi}(\partial^2.) = \mathbf{E}\,\mathbb{I}_\Phi + t\,\mathbf{E}\,\mathbb{I}_\Phi$. We can check again that $\partial^2\,(t\,\mathbf{E}) = (t\,\partial^2 + 2\,\partial)\,\mathbf{E} = 0$.

Lemma 4 (Corollary 3.2 of [10]) *If $d \in \mathbb{I}$ is Fredholm, then so is $\theta(d)$.*

Proof Let $d \in \mathbb{I}$ be Fredholm, i.e., $d \in \mathbb{I} \setminus (\mathbf{E})$. Suppose that $\theta(d) \in (\mathbf{E})$. At the end of Sect. 3.5, we show that $\theta((\mathbf{E})) \subset (\mathbf{E})$. Thus, $d = \theta(\theta(d)) \in (\mathbf{E})$, which is a contradiction and proves that $\theta(d) \in \mathbb{I} \setminus (\mathbf{E})$, i.e., $\theta(d)$ is Fredohlm.

The following corollary of Theorem 4 gives a way to compute a set of generators of the left annihilator $\text{ann}_{\mathbb{I}}(.d)$.

Corollary 4 *Let Φ be a subset of k with $0 \in \Phi$. Let $d \in \mathbb{I}_\Phi$ be Fredholm with* $\ker L_{\theta(d)} = \sum_{i=1}^{n} k\, r_i$, *where $r_i \in k[t]$. Then, we have*

$$\text{ann}_{\mathbb{I}_\Phi}(.d) = \sum_{i=1}^{n} \mathbb{I}_\Phi\, \mathbf{E}\, r_i((t\,\partial + 1)\,\partial) = \sum_{i=1}^{n} \mathbb{I}_\Phi\, \mathbf{E}\, \hat{r}_i(\partial),$$

where the polynomial $\hat{r}_i$ is defined by substituting t^i by $i!\,\partial^i$ into r_i.

Proof By Theorem 4, we have $\text{ann}_{\mathbb{I}_\Phi}(\theta(d).) = \sum_{i=1}^{n} (r_i\, \mathbf{E})\, \mathbb{I}_\Phi$. Applying θ to $r_i\, \mathbf{E}$, we get $\theta(r_i\, \mathbf{E}) = \theta(\mathbf{E})\, \theta(r_i) = \mathbf{E}\, r_i((t\,\partial + 1)\,\partial)$. We have $\theta(r_i\, \mathbf{E})\, d = \theta(\theta(d)\, r_i\, \mathbf{E}) = \theta(0) = 0$, which proves the inclusion $\supseteq$. Conversely, if $e \in \text{ann}_{\mathbb{I}_\Phi}(.d)$, i.e., $e\, d = 0$, then $\theta(d)\, \theta(e) = 0$, and thus $\theta(e) = \sum_{i=1}^{n} r_i\, \mathbf{E}\, d_i$ for certain $d_i \in \mathbb{I}_\Phi$, which yields $e = \theta^2(e) = \sum_{i=1}^{n} \theta(d_i)\, \mathbf{E}\, \theta(r_i)$, which proves the inclusion $\subseteq$ and the first equality. Finally, we note that $\mathbf{E}\, \theta(t)^j = \mathbf{E}\, ((t\,\partial + 1)\,\partial)^j = j!\, \mathbf{E}\, \partial^j$ for $j \in \mathbb{N}$, and thus $\mathbf{E} \sum_{j=0}^{r} s_j\, \theta(t)^j = \mathbf{E} \sum_{j=0}^{r} s_j\, j!\, \partial^j$, where $s_j \in k$, which proves the second equality. □

Example 8 If $d' = \int^2$, then $\theta(d') = \partial^2$ and using Example 7, we obtain $\text{ann}_{\mathbb{I}_\Phi}(\partial^2.) = \mathbf{E}\, \mathbb{I}_\Phi + t\, \mathbf{E}\, \mathbb{I}_\Phi$, which shows that $\text{ann}_{\mathbb{I}_\Phi}(\int^2.) = \mathbb{I}_\Phi\, \mathbf{E} + \mathbb{I}_\Phi\, \mathbf{E}\, (t\,\partial + 1)\,\partial = \mathbb{I}_\Phi\, \mathbf{E} + \mathbb{I}_\Phi\, \mathbf{E}\, \partial$. We can check again that $\mathbf{E}\, (t\,\partial + 1)\,\partial \int^2 = \mathbf{E}\, (t\,\partial + 1) \int = \mathbf{E}\, (t + \int) = 0$. Finally, according to the comments above Example 1, we obtain that the compatibility conditions of the inhomogeneous equation $\int_0^t \left(\int_0^\tau y(x)\, dx \right) d\tau = u(t)$, where u is a fixed enough regular function, are generated by $u(0) = 0$ and $(t\, \ddot{u}(t) + \dot{u}(t))(0) = \dot{u}(0) = 0$.

Similarly, we let the reader check that we have $\text{ann}_{\mathbb{I}_\Phi}(.(t\,\partial - 1)\,\partial^2) = \mathbf{E}\, \partial$.

All necessary steps for computing right and left annihilators have been implemented based on the `Maple` package *IntDiffOp* [26] for ID operators and boundary problems.

Example 9 Let us compute the compatibility conditions of (3.11). Note that

$$\text{rsym}(\theta(d)) = (0, n^2 - 3\,n + 2),$$

where:

$$\theta(d) = (t^2 + t - 3/5)\,\partial^2 - (2\,t + 1)\,\partial + 2.$$

The largest nonnegative integer root of q is 2. With this bound N for Proposition 5, we get for the following kernel:

$$\ker L_{\theta(d)} = k\,(t^2 + 3/5) + k\,(t + 1/2).$$

By Theorem 4, we obtain:

$$\operatorname{ann}_{\mathbb{I}}(\theta(d).) = ((t^2 + 3/5)\,\mathbf{E})\,\mathbb{I} + ((t + 1/2)\,\mathbf{E})\,\mathbb{I}.$$

Computing the involution of these generators yield the left annihilator

$$\operatorname{ann}_{\mathbb{I}}(.d) = \mathbb{I}\,(2\,\mathbf{E}\,\partial^2 + 3/5\,\mathbf{E}) + \mathbb{I}\,(\mathbf{E}\,\partial + 1/2\,\mathbf{E})$$

for (3.11), which correspond to the following compatibility conditions:

$$2\,\ddot{u}(0) + 3/5\,u(0) = 0, \quad \dot{u}(0) + 1/2\,u(0) = 0.$$

References

1. Abramov, S.A., Bronstein, M.: On solutions of linear functional systems. In: Proceedings of ISSAC 2001, pp. 1–6. ACM, New York (2001)
2. Abramov, S.A., Bronstein, M., Petkovšek, M.: On polynomial solutions of linear operator equations. In: Proceedings of ISSAC 1995, pp. 290–296. ACM, New York (1995)
3. Abramov, S.A., Kvansenko, K.Y.: Fast algorithms to search for the rational solutions of linear differential equations with polynomial coefficients. In: Proceedings of ISSAC 1991, pp. 267–270. ACM, New York (1991)
4. Abramov, S.A., Petkovšek, M., Ryabenko, A.: Special formal series solutions of linear operator equations. Discrete Math. **210**(1–3), 3–25 (2000)
5. Barkatou, M.A.: On rational solutions of systems of linear differential equations. J. Symb. Comput. **28**(4–5), 547–567 (1999)
6. Bavula, V.V.: Finite-dimensionality of Ext^n and Tor_n of simple modules over a class of algebras. Funktsional. Anal. Prilozhen. **25**(3), 80–82 (1991)
7. Bavula, V.V.: Generalized Weyl algebras and their representations. Algebra i Analiz **4**(1), 75–97 (1992)
8. Bavula, V.V.: The algebra of integro-differential operators on a polynomial algebra. J. Lond. Math. Soc. **83**(2), 517–543 (2011)
9. Bavula, V.V.: An analogue of the conjecture of Dixmier is true for the algebra of polynomial integro-differential operators. J. Algebra **372**, 237–250 (2012)
10. Bavula, V.V.: The algebra of integro-differential operators on an affine line and its modules. J. Pure Appl. Algebra **217**, 495–529 (2013)
11. Bavula, V.V.: The algebra of polynomial integro-differential operators is a holonomic bimodule over the subalgebra of polynomial differential operators. Algebr. Represent. Theory **17**, 275–288 (2014)
12. Bergman, G.M.: The diamond lemma for ring theory. Adv. Math. **29**, 178–218 (1978)
13. Bokut, L.A., Chen, Y.: Gröbner-Shirshov bases and their calculation. Bull. Math. Sci. **4**, 325–395 (2014)
14. Bostan, A., Cluzeau, T., Salvy, B.: Fast algorithms for polynomial solutions of linear differential equations. In: Proceedings of ISSAC 2005, pp. 45–52. ACM, New York (2005)
15. Boulier, F., Lemaire, F., Rosenkranz, M., Ushirobira, R., Verdière, N.: On symbolic approaches to integro-differential equations. In: Quadrat, A., Zerz, E. (eds.) Algebraic and Symbolic Computation Methods in Dynamical Systems, pp. 161–182. Springer, Switzerland (2020)

16. Chyzak, F., Quadrat, A., Robertz, D.: Effective algorithms for parametrizing linear control systems over Ore algebras. Appl. Algebra Engrg. Comm. Comput. **16**(5), 319–376 (2005)
17. Chyzak, F., Quadrat, A., Robertz, D.: OreModules: A symbolic package for the study of multidimensional linear systems. Applications of time delay systems. LNCIS, vol. 352, pp. 233–264. Springer, Berlin (2007)
18. Coutinho, S.C.: A Primer of Algebraic *D*-Modules, vol. 33. Cambridge University Press, Cambridge (1995)
19. Fliess, M.: Some basic structural properties of generalized linear systems. Syst. Control Lett. **15**(5), 391–396 (1990)
20. Gohberg, I., Kaashoek, M.A.: Time varying linear systems with boundary conditions and integral operators. I. The transfer operator and its properties. Integr. Equ. Oper. Theory **7**(3), 325–391 (1984)
21. Gohberg, I., Kaashoek, M.A., Lerer, L.: Minimality and irreducibility of time-invariant linear boundary value systems. Int. J. Control **44**(2), 363–379 (1986)
22. Hossein Poor, J., Raab, C.G., Regensburger, G.: Algorithmic operator algebras via normal forms for tensors. In: Rosenkranz, M. (ed.) Proceedings of ISSAC 2016, pp. 397–404. ACM, New York (2016)
23. Hossein Poor, J., Raab, C.G., Regensburger, G.: Normal forms for operators via Gröbner bases in tensor algebras. In: Greuel, G.M., Koch, T., Paule, P., Sommese, A. (eds.) Proceedings of ICMS 2016, Lecture Notes in Computing Science, vol. 9725, pp. 505–513. Springer, Berlin (2016)
24. Jacobson, N.: Some remarks on one-sided inverses. Proc. Am. Math. Soc. **1**, 352–355 (1950)
25. Komatsu, H.: On the index of ordinary differential operators. J. Fac. Sci. Univ. Tokyo Sect. IA Math. **18**, 379–398 (1971)
26. Korporal, A., Regensburger, G., Rosenkranz, M.: Symbolic computation for ordinary boundary problems in Maple. ACM Commun. Comput. Algebra **46**, 154–156 (2012). Software presentation at ISSAC 2012
27. Köthe, G.: Topological Vector Spaces. I, Die Grundlehren der mathematischen Wissenschaften, vol. 159. Springer, New York (1969)
28. Lam, T.Y.: A First Course in Noncommutative Rings. Springer, New York (1991)
29. Levandovskyy, V., Zerz, E.: Algebraic systems theory and computer algebraic methods for some classes of linear control systems. In: Proceedings of MTNS 2006. Kyoto (2006)
30. Malgrange, B.: Sur les points singuliers des équations différentielles. Enseign. Math. **2**(20), 147–176 (1974)
31. McConnell, J.C., Robson, J.C.: Noncommutative Noetherian Rings, Graduate Studies in Mathematics, vol. 30, revised edn. American Mathematical Society, Providence, RI (2001)
32. Mora, T.: An introduction to commutative and noncommutative Gröbner bases. Theoret. Comput. Sci. **134**, 131–173 (1994)
33. Oberst, U.: Multidimensional constant linear systems. Acta Appl. Math. **20**(1–2), 1–175 (1990)
34. Pommaret, J.F., Quadrat, A.: A functorial approach to the behaviour of multidimensional control systems. Int. J. Appl. Math. Comput. Sci. **13**(1), 7–13 (2003). Multidimensional systems *n*D and iterative learning control (Czocha Castle, 2000)
35. Quadrat, A.: The fractional representation approach to synthesis problems: an algebraic analysis viewpoint. I. (Weakly) doubly coprime factorizations. SIAM J. Control Optim. **42**(1), 266–299 (electronic) (2003)
36. Quadrat, A.: An introduction to constructive algebraic analysis and its applications (2010). http://hal.archives-ouvertes.fr/inria-00506104/fr/
37. Quadrat, A.: A constructive algebraic analysis approach to Artstein's reduction of linear time-delay systems. IFAC-PapersOnLine **48**(12), 209–214 (2015)
38. Quadrat, A., Regensburger, G.: Polynomial solutions and annihilators of ordinary integro-differential operators. In: Proceedings of SSSC (5th IFAC Symposium on System Structure and Control) 2013, pp. 308–313. IFAC, New York (2013)

39. Quadrat, A., Ushirobira, R.: Algebraic analysis for the Ore extension ring of differential time-varying delay operators. In: Proceedings of the 22nd International Symposium on Mathematical Theory of Networks and Systems (MTNS), University of Minnesota, Minneapolis, 12-15/07/2016 (2016)
40. Regensburger, G., Rosenkranz, M., Middeke, J.: A skew polynomial approach to integro-differential operators. In: Proceedings of ISSAC 2009, pp. 287–294. ACM, New York (2009)
41. Robertz, D.: Recent progress in an algebraic analysis approach to linear systems. Multidimens. Syst. Signal Process. **26**(2), 349–388 (2015)
42. Rosenkranz, M., Regensburger, G.: Solving and factoring boundary problems for linear ordinary differential equations in differential algebras. J. Symb. Comput. **43**(8), 515–544 (2008)
43. Rosenkranz, M., Regensburger, G., Tec, L., Buchberger, B.: Symbolic analysis for boundary problems: From rewriting to parametrized Gröbner bases. In: Numerical and Symbolic Scientific Computing: Progress and Prospects, pp. 273–331. SpringerWienNew York, Vienna (2012)
44. Rotman, J.: An Introduction to Homological Algebra, 2nd edn. Springer, New York (2009)
45. Seiler, W.M., Zerz, E.: Algebraic theory of linear systems: A survey. In: Surveys in differential-algebraic equations. II, Differential-Algebraic Equations Forum, pp. 287–333. Springer, Cham (2015)
46. Ufnarovski, V.: Introduction to noncommutative Gröbner bases theory. In: Gröbner bases and applications, pp. 259–280. Cambridge University Press, Cambridge (1998)
47. Wood, J.: Modules and behaviours in nD systems theory. Multidimens. Syst. Signal Process. **11**(1–2), 11–48 (2000). Recent progress in multidimensional control theory and applications

Part II
Symbolic Methods for Nonlinear Dynamical Systems and for Applications to Observation and Estimation Problems

Chapter 4
Thomas Decomposition and Nonlinear Control Systems

Markus Lange-Hegermann and Daniel Robertz

Abstract This paper applies the Thomas decomposition technique to nonlinear control systems, in particular to the study of the dependence of the system behavior on parameters. Thomas' algorithm is a symbolic method which splits a given system of nonlinear partial differential equations into a finite family of so-called simple systems which are formally integrable and define a partition of the solution set of the original differential system. Different simple systems of a Thomas decomposition describe different structural behavior of the control system in general. The paper gives an introduction to the Thomas decomposition method and shows how notions such as invertibility, observability and flat outputs can be studied. A Maple implementation of Thomas' algorithm is used to illustrate the techniques on explicit examples.

Keywords Thomas decomposition · Differential elimination · Nonlinear control systems · Flatness · Observability · Invertibility · Parameters in nonlinear control system

4.1 Introduction

This paper gives an introduction to the Thomas decomposition method and presents first steps in applying it to the structural study of nonlinear control systems. It extends and refines our earlier work [28].

Symbolic computation allows to study many structural aspects of control systems, e.g., controllability, observability, input-output behavior, etc. In contrast to a

M. Lange-Hegermann
Department of Electrical Engineering and Computer Science, Ostwestfalen-Lippe University of Applied Sciences and Arts, Campusallee 12, 32657 Lemgo, Germany
e-mail: markus.lange-hegermann@th-owl.de

D. Robertz (✉)
Centre for Mathematical Sciences, Plymouth University, 2-5 Kirkby Place, Drake Circus, Plymouth PL4 8AA, UK
e-mail: daniel.robertz@plymouth.ac.uk

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_4

numerical treatment, the dependence of the results on parameters occurring in the system is accessible to symbolic methods.

An algebraic approach for treating nonlinear control systems has been developed during the last decades, e.g., by M. Fliess and coworkers, J.-F. Pommaret and others, cf., e.g., [13, 20, 37], and the references therein. In particular, the notion of flatness has been studied extensively and has been applied to many interesting control problems (cf., e.g., [2, 14, 31]). The approach of Diop [10, 11] builds on the characteristic set method (cf. [24, 47]). The Rosenfeld-Gröbner algorithm (cf. [7]) can be used to perform the relevant computations effectively; implementations of related techniques are available, e.g., as Maple packages `DifferentialAlgebra` (by F. Boulier and E. S. Cheb-Terrab), formerly `diffalg` (by F. Boulier and E. Hubert), and `RegularChains` (by F. Lemaire, M. Moreno Maza, and Y. Xie) [29]; cf. also [46] for alternative approaches. As an example of an application of the Rosenfeld-Gröbner algorithm we refer to [34], where it is demonstrated how to compute a block feedforward form and a generalized controller form for a nonlinear control system.

So far the dependence of nonlinear control systems on parameters has not been studied by a rigorous method such as Thomas decomposition. This paper demonstrates how the Thomas decomposition method can be applied in this context. In particular, Thomas' algorithm can detect certain structural properties of control systems by performing elimination and it can separate singular cases of behavior in control systems from the generic case due to splitting into disjoint solution sets. We also consider the Thomas decomposition method as a preprocessing technique for the study of a linearization of a nonlinear system (cf. [42, Sect. 5.5]), an aspect that we do not pursue here.

Dependence of control systems on parameters has been examined, in particular, by J.-F. Pommaret and A. Quadrat in [37, 38]. For linear systems, stratifications of the space of parameter values have been studied using Gröbner bases in [30].

In the 1930s the American mathematician J. M. Thomas designed an algorithm which decomposes a polynomially nonlinear system of partial differential equations into so-called simple systems. The algorithm uses, in contrast to the characteristic set method, inequations to provide a disjoint decomposition of the solution set (cf. [44]). It precedes work by E. R. Kolchin [24] and A. Seidenberg [43], who followed J. F. Ritt [40]. Recently a new algorithmic approach to the Thomas decomposition method has been developed (cf. [4, 17, 41]), building also on ideas of the French mathematicians C. Riquier [39] and M. Janet [23]. Implementations as Maple packages of the algebraic and differential parts of Thomas' algorithm are available due to work by T. Bächler and M. Lange-Hegermann [5]. The implementation of the differential part is available in the Computer Physics Communications library [19] and has also been incorporated into Maple's standard library since Maple 2018. An earlier implementation of the algebraic part was given by D. Wang [45].

Section 4.2 introduces the Thomas decomposition method for algebraic and differential systems and discusses the main properties of its output. The algorithm for the differential case builds on the algebraic part. Section 4.3 explains how the Thomas decomposition technique can be used to solve elimination problems that occur in our study of nonlinear control systems. Finally, Sect. 4.4 addresses concepts of

nonlinear control theory, such as invertibility, observability, and flat outputs, possibly depending on parameters of the control system, and gives examples using a Maple implementation of Thomas' algorithm.

4.2 Thomas Decomposition

This section gives an introduction to the Thomas decomposition method for algebraic and differential systems. The case of differential systems, discussed in Sect. 4.2.2, builds on the case of algebraic systems which is dealt with in the first subsection. For more details on Thomas' algorithm, we refer to [3, 4, 17, 26, 35], and [41, Sect. 2.2].

4.2.1 Algebraic Systems

Let K be a field of characteristic zero and $R = K[x_1, \ldots, x_n]$ the polynomial algebra with indeterminates $x_1, \ldots, x_n$ over K. We denote by $\overline{K}$ an algebraic closure of K.

Definition 1 An *algebraic system* S, defined over R, is given by finitely many equations and inequations

$$p_1 = 0, \quad p_2 = 0, \quad \ldots, \quad p_s = 0, \quad q_1 \neq 0, \quad q_2 \neq 0, \quad \ldots, \quad q_t \neq 0, \qquad (4.1)$$

where $p_1, \ldots, p_s, q_1, \ldots, q_t \in R$ and $s, t \in \mathbb{Z}_{\geq 0}$. The *solution set* of S in $\overline{K}^n$ is

$$\mathrm{Sol}_{\overline{K}}(S) := \{\, a \in \overline{K}^n \mid p_i(a) = 0 \text{ and } q_j(a) \neq 0 \text{ for all } 1 \leq i \leq s,\ 1 \leq j \leq t \,\}.$$

We fix a total ordering $>$ on the set $\{x_1, \ldots, x_n\}$ allowing us to consider every non-constant element p of R as a univariate polynomial in the greatest variable with respect to $>$ which occurs in p, with coefficients which are themselves univariate polynomials in lower ranked variables, etc. Without loss of generality we may assume that $x_1 > x_2 > \cdots > x_n$. The choice of $>$ corresponds to a choice of projections

$$\begin{aligned}
\pi_1 &: \overline{K}^n \longrightarrow \overline{K}^{n-1} : (a_1, a_2, \ldots, a_n) \longmapsto (a_2, a_3, a_4, \ldots, a_n),\\
\pi_2 &: \overline{K}^n \longrightarrow \overline{K}^{n-2} : (a_1, a_2, \ldots, a_n) \longmapsto \quad (a_3, a_4, \ldots, a_n),\\
&\ \ \vdots \qquad\qquad\qquad\qquad\quad \vdots\\
\pi_{n-1} &: \overline{K}^n \longrightarrow \overline{K} : \quad (a_1, a_2, \ldots, a_n) \longmapsto \qquad\qquad a_n.
\end{aligned}$$

Thus, the recursive representation of polynomials is motivated by considering each $\pi_{k-1}(\mathrm{Sol}_{\overline{K}}(S))$ as fibered over $\pi_k(\mathrm{Sol}_{\overline{K}}(S))$, for $k = 1, \ldots, n-1$, where $\pi_0 := \mathrm{id}_{\overline{K}^n}$ (cf. also [35]). The purpose of a Thomas decomposition of $\mathrm{Sol}_{\overline{K}}(S)$, to be defined below, is to clarify this fibration structure. The solution set $\mathrm{Sol}_{\overline{K}}(S)$ is partitioned

into subsets $\mathrm{Sol}_{\overline{K}}(S_1)$, …, $\mathrm{Sol}_{\overline{K}}(S_r)$ in such a way that, for each $i = 1, \ldots, r$ and $k = 1, \ldots, n-1$, the fiber cardinality $|\pi_k^{-1}(\{a\})|$ does not depend on the choice of $a \in \pi_k(\mathrm{Sol}_{\overline{K}}(S_i))$. In terms of the defining equations and inequations in (4.1), the fundamental obstructions to this uniform behavior are zeros of the leading coefficients of p_i or q_j and zeros of p_i or q_j of multiplicity greater than one.

Definition 2 Let $p \in R \setminus K$.

(a) The greatest variable with respect to $>$ which occurs in p is referred to as the *leader* of p and is denoted by $\mathrm{ld}(p)$.
(b) For $v = \mathrm{ld}(p)$ we denote by $\deg_v(p)$ the degree of p in v.
(c) The coefficient of the highest power of $\mathrm{ld}(p)$ occurring in p is called the *initial* of p and is denoted by $\mathrm{init}(p)$.
(d) The *discriminant* of p is defined as

$$\mathrm{disc}(p) := (-1)^{d(d-1)/2} \operatorname{res}\left(p, \frac{\partial p}{\partial\, \mathrm{ld}(p)}, \mathrm{ld}(p)\right) / \mathrm{init}(p), \qquad d = \deg_{\mathrm{ld}(p)}(p),$$

where $\operatorname{res}(p, q, v)$ is the resultant of p and q with respect to the variable v. (Note that $\mathrm{disc}(p)$ is a polynomial because $\mathrm{init}(p)$ divides $\operatorname{res}(p, \partial p/\partial\, \mathrm{ld}(p), \mathrm{ld}(p))$, since the Sylvester matrix, whose determinant is $\operatorname{res}(p, \partial p/\partial\, \mathrm{ld}(p), \mathrm{ld}(p))$, has a column all of whose entries are divisible by $\mathrm{init}(p)$.)

Both $\mathrm{init}(p)$ and $\mathrm{disc}(p)$ are elements of the polynomial algebra $K[x \mid x < \mathrm{ld}(p)]$. The zeros of a univariate polynomial which have multiplicity greater than one are the common zeros of the polynomial and its derivative. The solutions of $\mathrm{disc}(p) = 0$ in $\overline{K}^{n-k}$, where $\mathrm{ld}(p) = x_k$, are therefore those tuples $(a_{k+1}, a_{k+2}, \ldots, a_n)$ for which the substitution $x_{k+1} = a_{k+1}$, $x_{k+2} = a_{k+2}$, …, $x_n = a_n$ in p results in a univariate polynomial with a zero of multiplicity greater than one.

Definition 3 An algebraic system S, defined over R, as in (4.1) is said to be *simple* (with respect to $>$) if the following three conditions hold.

(a) For all $i = 1, \ldots, s$ and $j = 1, \ldots, t$ we have $p_i \notin K$ and $q_j \notin K$.
(b) The leaders of the left hand sides of the equations and inequations in S are pairwise different, i.e., $|\{\mathrm{ld}(p_1), \ldots, \mathrm{ld}(p_s), \mathrm{ld}(q_1), \ldots, \mathrm{ld}(q_t)\}| = s + t$.
(c) For every $r \in \{p_1, \ldots, p_s, q_1, \ldots, q_t\}$, if $\mathrm{ld}(r) = x_k$, then neither of the equations $\mathrm{init}(r) = 0$ and $\mathrm{disc}(r) = 0$ has a solution $(a_{k+1}, a_{k+2}, \ldots, a_n)$ in $\pi_k(\mathrm{Sol}_{\overline{K}}(S))$.

Subsets of non-constant polynomials in R with pairwise different leaders (i.e., satisfying (a) and (b)) are also referred to as triangular sets (cf., e.g., [1, 21, 46]).

Remark 1 A simple algebraic system S admits the following solution procedure, which also shows that its solution set is not empty. Let $S_{<k}$ be the subset of S consisting of the equations $p = 0$ and inequations $q \neq 0$ with $\mathrm{ld}(p) < x_k$ and $\mathrm{ld}(q) < x_k$. The fibration structure implied by (c) ensures that, for every $k = 1, \ldots, n-1$, every

solution $(a_{k+1}, a_{k+2}, \ldots, a_n)$ of $\pi_k(\mathrm{Sol}_{\overline{K}}(S)) = \pi_k(\mathrm{Sol}_{\overline{K}}(S_{<k}))$ can be extended to a solution $(a_k, a_{k+1}, \ldots, a_n)$ of $\pi_{k-1}(\mathrm{Sol}_{\overline{K}}(S))$. If S contains an equation $p = 0$ with leader x_k, then there exist exactly $\deg_{x_k}(p)$ such elements $a_k \in \overline{K}$ (because zeros with multiplicity greater than one are excluded by the non-vanishing discriminant). If S contains an inequation $q \neq 0$ with leader x_k, all $a_k \in \overline{K}$ except $\deg_{x_k}(q)$ elements define a tuple $(a_k, a_{k+1}, \ldots, a_n)$ as above. If no equation and no inequation in S has leader x_k, then $a_k \in \overline{K}$ can be chosen arbitrarily.

Definition 4 Let S be an algebraic system, defined over R. A *Thomas decomposition* of S (or of $\mathrm{Sol}_{\overline{K}}(S)$) with respect to $>$ is a collection of finitely many simple algebraic systems S_1, …, S_r, defined over R, such that $\mathrm{Sol}_{\overline{K}}(S)$ is the disjoint union of the solution sets $\mathrm{Sol}_{\overline{K}}(S_1)$, …, $\mathrm{Sol}_{\overline{K}}(S_r)$.

We outline Thomas' algorithm for computing a Thomas decomposition of algebraic systems.

Remark 2 Given S as in (4.1) and a total ordering $>$ on $\{x_1, \ldots, x_n\}$, a Thomas decomposition of S with respect to $>$ can be constructed by combining Euclid's algorithm with a splitting strategy.

First of all, if S contains an equation $c = 0$ with $0 \neq c \in K$ or the inequation $0 \neq 0$, then S is discarded because it has no solutions. Moreover, from now on the equation $0 = 0$ and inequations $c \neq 0$ with $0 \neq c \in K$ are supposed to be removed from S.

An elementary step of the algorithm applies a pseudo-division to a pair p_1, p_2 of non-constant polynomials in R with the same leader x_k and $\deg_{x_k}(p_1) \geq \deg_{x_k}(p_2)$. The result is a pseudo-remainder

$$r = c_1 \cdot p_1 - c_2 \cdot p_2, \tag{4.2}$$

where $c_1, c_2 \in R$ and r is constant or has leader less than x_k or has leader x_k and $\deg_{x_k}(r) < \deg_{x_k}(p_1)$. Since the coefficients of p_1 and p_2 are polynomials in lower ranked variables, multiplication of p_1 by a non-constant polynomial c_1 may be necessary in general to perform the reduction in R (and not in its field of fractions). The choice of c_1 as a suitable power of $\mathrm{init}(p_2)$ always achieves this.

In order to turn S into a triangular set, the algorithm deals with three kinds of subsets of S of cardinality two. Firstly, each pair of equations $p_1 = 0$, $p_2 = 0$ in S with $\mathrm{ld}(p_1) = \mathrm{ld}(p_2)$ is replaced with the single equation $r = 0$, where r is the result of applying Euclid's algorithm to p_1 and p_2, considered as univariate polynomials in their leader, using the above pseudo-division. (If this computation was stable under substitution of values for lower ranked variables in p_1 and p_2, then r would be the greatest common divisor of the specialized polynomials.)

The solution set of the system is supposed not to change, when the equation $p_1 = 0$ is replaced with the equation $r = 0$ given by the pseudo-reduction (4.2). Therefore, we assume that the polynomial c_1, and hence $\mathrm{init}(p_2)$, does not vanish on the solution set of the system. In order to ensure this condition, a preparatory step splits the system into two, if necessary, and adds the inequation $\mathrm{init}(p_2) \neq 0$ to one

of them and the equation $\mathrm{init}(p_2) = 0$ to the other. The algorithm then deals with both systems separately. These case distinctions also allow to arrange for the part of condition (c) in Definition 3 which concerns initials.

Secondly, let $p = 0$, $q \neq 0$ be in S with $\mathrm{ld}(p) = \mathrm{ld}(q) = x_k$. If $\deg_{x_k}(p) \leq \deg_{x_k}(q)$, then $q \neq 0$ is replaced with $r \neq 0$, where r is the result of applying the pseudo-division (4.2) to q and p. Otherwise, Euclid's algorithm is applied to p and q, keeping track of the coefficients used for the reductions as in (4.2). Given the result r, the system is then split into two, adding the conditions $r \neq 0$ and $r = 0$, respectively. The inequation $q \neq 0$ is removed from the first new system, because $p = 0$ and $q \neq 0$ have no common solution in that case. The assumption $r = 0$ and the bookkeeping allows to divide p by the common factor of p and q (modulo left hand sides of equations with smaller leader). The left hand side of $p = 0$ is replaced with that quotient in the second new system. Not all of these cases need a closer inspection. For instance, if p divides q, then the solution set of S is empty and S is discarded.

Thirdly, for a pair $q_1 \neq 0, q_2 \neq 0$ in S with $\mathrm{ld}(q_1) = \mathrm{ld}(q_2)$, Euclid's algorithm is applied to q_1 and q_2 in the same way as above. Keeping track of the coefficients used in intermediate steps allows to determine the least common multiple m of q_1 and q_2, which again depends on distinguishing the cases whether the result of Euclid's algorithm vanishes or not. The pair $q_1 \neq 0, q_2 \neq 0$ is then replaced with $m \neq 0$.

The part of condition (c) in Definition 3 regarding discriminants is taken care of by applying Euclid's algorithm as above to p and $\partial p / \partial \,\mathrm{ld}(p)$, where p is the left hand side of an equation or inequation. Bookkeeping allows to determine the square-free part of p, which depends again on case distinctions.

Expressions tend to grow very quickly when performing these reductions, so that an appropriate strategy is essential for dealing with non-trivial systems. Apart from dividing by the content (in K) of polynomials, in intermediate steps of Euclid's algorithm the coefficients should be reduced modulo equations in the system with lower ranked leaders. In practice, subresultant computations (cf., e.g., [32]) allow to diminish the growth of coefficients significantly.

Termination of the procedure sketched above depends on the organization of its steps. One possible strategy is to maintain an intermediate triangular set, reduce new equations and inequations modulo the equations in the triangular set, and select among these results the one with smallest leader and least degree, preferably an equation, for insertion into the triangular set. If the set already contains an equation or inequation with the same leader, then the pair is treated as discussed above. Since equations are replaced with equations of smaller degree and inequations are replaced with equations if possible or with the least common multiple of inequations, this strategy terminates after finitely many steps.

For more details on the algebraic part of Thomas' algorithm, we refer to [3, 4], and [41, Subsect. 2.2.1].

An implementation of Thomas' algorithm for algebraic systems has been developed by T. Bächler as Maple package `AlgebraicThomas` [5].

In what follows, variables are underlined to emphasize that they are leaders of polynomials with respect to the fixed total ordering $>$.

Example 1 Let us compute a Thomas decomposition of the algebraic system

$$x^2 + y^2 - 1 = 0$$

consisting of one equation, defined over $R = \mathbb{Q}[x, y]$, with respect to $x > y$. We set $p_1 := x^2 + y^2 - 1$. Then we have $\mathrm{ld}(p_1) = x$ and $\mathrm{init}(p_1) = 1$ and

$$\mathrm{disc}(p_1) = -4\,y^2 + 4.$$

We distinguish the cases whether or not $p_1 = 0$ has a solution which is also a zero of $\mathrm{disc}(p_1)$, or equivalently, of $y^2 - 1$. In other words, we replace the original algebraic system with two algebraic systems which are obtained by adding the inequation $y^2 - 1 \neq 0$ or the equation $y^2 - 1 = 0$. The first system is readily seen to be simple, whereas the second one is transformed into a simple system by taking the difference of the two equations and computing a square-free part. Clearly, the solution sets of the two resulting simple systems form a partition of the solution set of $p_1 = 0$. We obtain the Thomas decomposition

$$\boxed{\begin{aligned} \underline{x}^2 + y^2 - 1 &= 0 \\ \underline{y}^2 - 1 &\neq 0 \end{aligned}} \qquad \boxed{\begin{aligned} \underline{x} &= 0 \\ \underline{y}^2 - 1 &= 0 \end{aligned}}$$

In this example, all points of $\mathrm{Sol}_{\overline{K}}(\{\, p_1 = 0 \,\})$ for which the projection π_1 onto the y-axis has fibers of an exceptional cardinality have real coordinates, and the significance of the above case distinction can be confirmed graphically.

As a further illustration let us augment the original system by the equation which expresses the coordinate t of the point of intersection of the line through the two points $(0, 1)$ and (x, y) on the circle with the x-axis (stereographic projection):

$$\begin{cases} x^2 + y^2 - 1 = 0 \\ (1 - y)\,t - x = 0 \end{cases}$$

A Thomas decomposition with respect to $x > y > t$ is obtained as follows. We set $p_2 := x + t\,y - t$. Since $\mathrm{ld}(p_1) = \mathrm{ld}(p_2)$, we apply polynomial division:

$$p_1 - (x - t\,y + t)\,p_2 = (1 + t^2)\,\underline{y}^2 - 2\,t^2\,\underline{y} + t^2 - 1 = (\underline{y} - 1)\,((1 + t^2)\,\underline{y} - t^2 + 1).$$

Replacing p_1 with the remainder of this division does not alter the solution set of the algebraic system. It is convenient (but not necessary) to split the system into two systems according to the factorization of the remainder:

$$\begin{cases} \underline{x} + t\,y - t = 0 \\ (1 + t^2)\,\underline{y} - t^2 + 1 = 0 \\ \underline{y} - 1 \neq 0 \end{cases} \qquad \begin{cases} \underline{x} + t\,y - t = 0 \\ \\ \underline{y} - 1 = 0 \end{cases}$$

Another polynomial division reveals that the equation and the inequation with leader y in the first system have no common solutions. Therefore, the inequation can be omitted from that system. The initial of the equation has to be investigated. In fact, the assumption $1 + t^2 = 0$ leads to a contradiction. Finally, the equation with leader y can be used to eliminate y in the equation with leader x:

$$(1 + t^2)\,(\underline{x} + t\,y - t) - t\,((1 + t^2)\,\underline{y} - t^2 + 1) = (1 + t^2)\,\underline{x} - 2\,t.$$

A similar simplification can be applied to the second system above. We obtain the Thomas decomposition

$$\boxed{\begin{aligned}(1 + t^2)\,\underline{x} - 2\,t &= 0\\ (1 + t^2)\,\underline{y} - t^2 + 1 &= 0\\ \underline{t}^2 + 1 &\neq 0\end{aligned}} \qquad \boxed{\begin{aligned}\underline{x} &= 0\\ \underline{y} - 1 &= 0\end{aligned}}$$

from which a rational parametrization of the circle can be read off.

Remark 3 A Thomas decomposition of an algebraic system is not uniquely determined. It depends on the chosen total ordering $>$, the order in which intermediate systems are dealt with and other choices, such as whether factorizations of left hand sides of equations are taken into account or not.

According to Hilbert's Nullstellensatz (cf., e.g., [12]), the solution sets V of systems of polynomial equations in $x_1, \ldots, x_n$ in $\overline{K}^n$, defined over R, are in one-to-one correspondence with their vanishing ideals in R

$$\mathcal{I}_R(V) := \{\, p \in R \mid p(a) = 0 \text{ for all } a \in V \,\},$$

and these are the radical ideals of R, i.e., the ideals I of R which equal their radicals

$$\sqrt{I} := \{\, p \in R \mid p^r \in I \text{ for some } r \in \mathbb{Z}_{\geq 0} \,\}.$$

The solution sets V can then be considered as the closed subsets of $\overline{K}^n$ with respect to the Zariski topology.

The fibration structure of a simple algebraic system S allows to deduce that the polynomials in R which vanish on $\mathrm{Sol}_{\overline{K}}(S)$ are precisely those polynomials in R whose pseudo-remainders modulo $p_1, \ldots, p_s$ are zero, where $p_1 = 0, \ldots, p_s = 0$ are the equations in S. If E is the ideal of R generated by $p_1, \ldots, p_s$ and q the product of all $\mathrm{init}(p_i)$, then these polynomials form the saturation ideal

$$E : q^\infty := \{\, p \in R \mid q^r \cdot p \in E \text{ for some } r \in \mathbb{Z}_{\geq 0} \,\}.$$

In particular, simple algebraic systems admit an effective way to decide membership of a polynomial to the associated radical ideal (cf. also Proposition 3 below).

Proposition 1 ([41], Prop. 2.2.7) *Let S be a simple algebraic system as in (4.1), E the ideal of R generated by $p_1, \ldots, p_s$, and q the product of all* $\mathrm{init}(p_i)$. *Then $E : q^\infty$ consists of all polynomials in R which vanish on* $\mathrm{Sol}_{\overline{K}}(S)$. *In particular, $E : q^\infty$ is a radical ideal. Given $p \in R$, we have $p \in E : q^\infty$ if and only if the pseudo-remainder of p modulo $p_1, \ldots, p_s$ is zero.*

4.2.2 Differential Systems

Definition 5 A *differential field* K with commuting derivations $\delta_1, \ldots, \delta_n$ is a field K endowed with maps $\delta_i : K \to K$, satisfying

$$\delta_i(k_1 + k_2) = \delta_i(k_1) + \delta_i(k_2), \quad \delta_i(k_1\, k_2) = \delta_i(k_1)\, k_2 + k_1\, \delta_i(k_2) \quad \text{for all } k_1, k_2 \in K,$$

$i = 1, \ldots, n$, and $\delta_i \circ \delta_j = \delta_j \circ \delta_i$ for all $1 \le i, j \le n$.

In what follows, let K be the differential field of (complex) meromorphic functions on an open and connected subset Ω of $\mathbb{C}^n$. The derivations on K are given by the partial differential operators $\delta_1, \ldots, \delta_n$ with respect to the coordinates of $\mathbb{C}^n$. Moreover, let $R = K\{u_1, \ldots, u_m\}$ be the differential polynomial ring in the differential indeterminates $u_1, \ldots, u_m$. These indeterminates give rise to symbols $(u_k)_J$, where $J = (j_1, \ldots, j_n) \in (\mathbb{Z}_{\ge 0})^n$, which represent the partial derivatives of m infinitely differentiable functions. More precisely, R is the polynomial algebra $K[(u_k)_J \mid 1 \le k \le m,\ J \in (\mathbb{Z}_{\ge 0})^n]$ over K in infinitely many indeterminates $(u_k)_J$, endowed with commuting derivations $\partial_1, \ldots, \partial_n$ such that

$$\partial_j\,((u_k)_J) = (u_k)_{J+1_j}, \qquad \partial_j|_K = \delta_j \quad \text{for all } j = 1, \ldots, n,$$

where 1_j is the jth standard basis vector of $\mathbb{Z}^n$. For $k \in \{1, \ldots, m\}$, we identify $(u_k)_{(0,\ldots,0)}$ and u_k. We set $\Delta := \{\, \partial_1, \ldots, \partial_n \,\}$, and for any subset $\{\, \partial_{i_1}, \ldots, \partial_{i_r} \,\}$ of Δ we define the free commutative monoid of all monomials in $\partial_{i_1}, \ldots, \partial_{i_r}$

$$\mathrm{Mon}(\{\, \partial_{i_1}, \ldots, \partial_{i_r} \,\}) := \{\, \partial_{i_1}^{e_1} \ldots \partial_{i_r}^{e_r} \mid e \in (\mathbb{Z}_{\ge 0})^r \,\}.$$

Definition 6 A *differential system* S, defined over $R = K\{u_1, \ldots, u_m\}$, is given by finitely many equations and inequations

$$p_1 = 0, \quad p_2 = 0, \quad \ldots, \quad p_s = 0, \quad q_1 \neq 0, \quad q_2 \neq 0, \quad \ldots, \quad q_t \neq 0, \qquad (4.3)$$

where $p_1, \ldots, p_s, q_1, \ldots, q_t \in R$ and $s, t \in \mathbb{Z}_{\ge 0}$. The *solution set* of S is

$$\mathrm{Sol}_\Omega(S) := \{\, f = (f_1, \ldots, f_m) \mid f_k : \Omega \to \mathbb{C} \text{ analytic},\ k = 1, \ldots, m,\ p_i(f) = 0,\ q_j(f) \neq 0,\ i = 1, \ldots, s,\ j = 1, \ldots, t \,\}.$$

Remark 4 Since each component f_k of a solution of (4.3) is assumed to be analytic, the equations $p_i = 0$ and inequations $q_j \neq 0$ (and their consequences) can be translated into algebraic conditions on the Taylor coefficients of power series expansions of $f_1, \ldots, f_m$ (around a point in Ω). An inequation $q \neq 0$ then turns into a disjunction of algebraic inequations for all coefficients which result from substitution of power series expansions for $u_1, \ldots, u_m$ in q. (This approach leads to the definition of the differential counting polynomial, a fine invariant of a differential system [27].)

An appropriate choice of $\Omega \subseteq \mathbb{C}^n$ can often only be made after the formal treatment of a given differential system by Thomas' algorithm (as, e.g., singularities of coefficients in differential consequences will only be detected during that process). In general, we assume that Ω is chosen in such a way that the given systems have analytic solutions on Ω.

Clearly, by neglecting the derivations on $R = K\{u_1, \ldots, u_m\}$, a differential system can be considered as an algebraic system in the finitely many variables $(u_i)_J$ which occur in the equations and inequations. The same recursive representation of polynomials as in the algebraic case is employed, but the total ordering on the set of variables $(u_i)_J$ is supposed to respect the action of the derivations.

Definition 7 A *ranking* $>$ on $R = K\{u_1, \ldots, u_m\}$ is a total ordering on the set

$$\mathrm{Mon}(\Delta)\, u := \{\, (u_k)_J \mid 1 \leq k \leq m,\ J \in (\mathbb{Z}_{\geq 0})^n \,\}$$

such that for all $j \in \{1, \ldots, n\}$, $k, k_1, k_2 \in \{1, \ldots, m\}$, $J_1, J_2 \in (\mathbb{Z}_{\geq 0})^n$ we have

(a) $\partial_j\, u_k > u_k$ and
(b) $(u_{k_1})_{J_1} > (u_{k_2})_{J_2}$ implies $\partial_j\, (u_{k_1})_{J_1} > \partial_j\, (u_{k_2})_{J_2}$.

Remark 5 Every ranking $>$ on R is a well-ordering (cf., e.g., [24, Ch. 0, Sect. 17, Lemma 15]), i.e., every descending sequence of elements of $\mathrm{Mon}(\Delta)\, u$ terminates.

Example 2 On $K\{u\}$ (i.e., $m = 1$) with commuting derivations $\partial_1, \ldots, \partial_n$ the *degree-reverse lexicographical ranking* (with $\partial_1 u > \partial_2 u > \cdots > \partial_n u$) is defined for u_J, $u_{J'}$, $J = (j_1, \ldots, j_n)$, $J' = (j'_1, \ldots, j'_n) \in (\mathbb{Z}_{\geq 0})^n$, by

$$u_J > u_{J'} \quad :\Longleftrightarrow \quad \begin{cases} j_1 + \cdots + j_n > j'_1 + \cdots + j'_n & \text{or} \\ \big(j_1 + \cdots + j_n = j'_1 + \cdots + j'_n \ \text{ and } \ J \neq J' \ \text{ and} \\ j_i < j'_i \ \text{ for } \ i = \max\{\, 1 \leq k \leq n \mid j_k \neq j'_k \,\} \big). \end{cases}$$

For instance, if $n = 3$, we have $u_{(1,2,1)} > u_{(1,2,0)} > u_{(2,0,1)}$.

In what follows, we assume that a ranking $>$ on $R = K\{u_1, \ldots, u_m\}$ is fixed.

Remark 6 Let $p_1, p_2 \in R$ be two non-constant differential polynomials. If p_1 and p_2 have the same leader $(u_k)_J$ and the degree of p_1 in $(u_k)_J$ is greater than or equal to the degree of p_2 in $(u_k)_J$, then the same pseudo-division as in (4.2) yields a remainder which is either zero, or has leader less than $(u_k)_J$, or has leader $(u_k)_J$ and smaller degree in $(u_k)_J$ than p_1.

More generally, if $\mathrm{ld}(p_1) = \theta\,\mathrm{ld}(p_2)$ for some $\theta \in \mathrm{Mon}(\Delta)$, then this pseudo-division can be applied with p_2 replaced with $\theta\,p_2$. Note that, by condition (b) of the definition of a ranking, we have $\mathrm{ld}(\theta\,p_2) = \theta\,\mathrm{ld}(p_2)$, and that, if $\theta \neq 1$, the degree of $\theta\,p_2$ in $\theta\,\mathrm{ld}(p_2)$ is one, so that the reduction can be applied without assumption on the degree of p_2 in $\mathrm{ld}(p_2)$. Then c_1 in (4.2) is again chosen as a suitable power of $\mathrm{init}(\theta\,p_2)$. In case $\theta \neq 1$ we have

$$\mathrm{init}(\theta\,p_2) = \frac{\partial p_2}{\partial\,\mathrm{ld}(p_2)} =: \mathrm{sep}(p_2),$$

and this differential polynomial is referred to as the *separant* of p_2.

In order not to change the solution set of a differential system, when $p_1 = 0$ is replaced with $r = 0$, where r is the result of a reduction of p_1 modulo p_2 or $\theta\,p_2$ as above, it is assumed that $\mathrm{init}(p_2)$ and $\mathrm{sep}(p_2)$ do not vanish on the solution set of the system. By definition of the separant and the discriminant (cf. Definition 2 (d)), non-vanishing of $\mathrm{sep}(p_2)$ follows from non-vanishing of $\mathrm{disc}(p_2)$, as ensured by the algebraic part of Thomas' algorithm (cf. Remark 2).

We assume now that the given differential system is simple as an algebraic system; it could be one of the systems resulting from the algebraic part of Thomas' algorithm.

Remark 7 The symmetry of the second derivatives $\partial_i\,\partial_j\,u_k = \partial_j\,\partial_i\,u_k$ (and similarly for higher order derivatives) imposes necessary conditions on the solvability of a system of partial differential equations. Taking identities like these into account and forming linear combinations of (derivatives of) the given equations may produce differential consequences with lower ranked leaders. In order to obtain a complete set of algebraic conditions on the Taylor coefficients of an analytic solution, the system has to be augmented by these integrability conditions in general. If a system of partial differential equations admits a translation into algebraic conditions on the Taylor coefficients such that no further integrability conditions have to be taken into account, then it is said to be *formally integrable*.

A simple differential system, to be defined in Definition 11, will be assumed to be formally integrable. The construction of simple differential systems, and therefore, the computation of a Thomas decomposition, as presented in [4, 41], employs techniques which can be traced back to C. Riquier [39] and M. Janet [23]. The main idea is to turn the search for new differential consequences (i.e., integrability conditions) into a systematic procedure by singling out for each differential equation those derivations (called "non-admissible" here) which need to be applied to it in this investigation. The notion of Janet division, as discussed next, establishes a sense of direction in combining the given equations and deriving consequences. It is a particular case of an involutive division on sets of monomials, a concept developed by V. P. Gerdt and Y. A. Blinkov and others (cf., e.g., [18]).

Definition 8 Given a finite subset M of $\mathrm{Mon}(\Delta)$, *Janet division* associates with each $\theta \in M$ a subset of *admissible derivations* $\mu(\theta, M)$ of $\Delta = \{\partial_1, \ldots, \partial_n\}$ as follows. Let $\theta = \partial_1^{i_1} \cdots \partial_n^{i_n}$. Then $\partial_k \in \mu(\theta, M)$ if and only if

$$i_k = \max\{\, j_k \mid \partial_1^{j_1} \cdots \partial_n^{j_n} \in M \text{ with } j_1 = i_1,\ j_2 = i_2, \ldots,\ j_{k-1} = i_{k-1} \,\}.$$

The subset $\overline{\mu}(\theta, M) := \Delta \setminus \mu(\theta, M)$ consists of the *non-admissible derivations* for the element θ of M.

Example 3 Let $\Delta = \{\, \partial_1, \partial_2, \partial_3 \,\}$ and $M = \{\, \partial_1^2\, \partial_2,\ \partial_1^2\, \partial_3,\ \partial_2^2\, \partial_3,\ \partial_2\, \partial_3^2 \,\}$. Then Janet division associates the sets $\mu(\theta, M)$ of admissible derivations to the elements $\theta \in M$ as indicated in the following table, where we replace non-admissible derivations in the set Δ with the symbol '$*$'.

$$\begin{array}{l} \partial_1^2\, \partial_2,\ \{\partial_1\,,\ \partial_2\,,\ \partial_3\} \\ \partial_1^2\, \partial_3,\ \{\partial_1\,,\ *\,,\ \partial_3\} \\ \partial_2^2\, \partial_3,\ \{\,*\,,\ \partial_2\,,\ \partial_3\} \\ \partial_2\, \partial_3^2,\ \{\,*\,,\ *\,,\ \partial_3\} \end{array}$$

Definition 9 A finite subset M of $\mathrm{Mon}(\Delta)$ is said to be *Janet complete* if

$$\bigcup_{\theta \in M} \mathrm{Mon}(\mu(\theta, M))\, \theta = \bigcup_{\theta \in M} \mathrm{Mon}(\Delta)\, \theta,$$

i.e., if every monomial which is divisible by some monomial in M is obtained by multiplying a certain $\theta \in M$ by admissible derivations for θ only. (Recall that the left hand side of the above equation is a disjoint union.)

Example 4 The set M in Example 3 is not Janet complete because, e.g., the monomial $\partial_1\, \partial_2^2\, \partial_3$ is not obtained as a multiple of any $\theta \in M$ when multiplication is restricted to admissible derivations for θ. By adding this monomial and the monomial $\partial_1\, \partial_2\, \partial_3^2$ to M, we obtain the following Janet complete superset of M in $\mathrm{Mon}(\Delta)$.

$$\begin{array}{l} \partial_1^2\, \partial_2,\ \{\partial_1\,,\ \partial_2\,,\ \partial_3\} \\ \partial_1^2\, \partial_3,\ \{\partial_1\,,\ *\,,\ \partial_3\} \\ \partial_1\, \partial_2^2\, \partial_3,\ \{\,*\,,\ \partial_2\,,\ \partial_3\} \\ \partial_1\, \partial_2\, \partial_3^2,\ \{\,*\,,\ *\,,\ \partial_3\} \\ \partial_2^2\, \partial_3,\ \{\,*\,,\ \partial_2\,,\ \partial_3\} \\ \partial_2\, \partial_3^2,\ \{\,*\,,\ *\,,\ \partial_3\} \end{array}$$

Remark 8 Every finite subset M of $\mathrm{Mon}(\Delta)$ can be augmented to a Janet complete finite set by adding certain monomials which are products of some $\theta \in M$ and a monomial which is divisible by at least one non-admissible derivation for θ.

For more details on Janet division, we refer to, e.g., [4, 18, 41].

Each equation $p_i = 0$ in a differential system is assigned the set of admissible derivations $\mu(\theta_i, M_k)$, where $\mathrm{ld}(p_i) = \theta_i\, u_k$ and

$$M_k := \{\, \theta \in \mathrm{Mon}(\Delta) \mid \theta\, u_k \in \{\, \mathrm{ld}(p_1), \ldots, \mathrm{ld}(p_s) \,\} \,\} \tag{4.4}$$

is the set of all monomials which define leaders of the equations $p_1 = 0, \ldots, p_s = 0$ in the system involving the same differential indeterminate u_k. We refer to $d\ p_i$ for $d \in \mathrm{Mon}(\mu(\theta_i, M_k))$ as the *admissible derivatives* of p_i.

Formal integrability of a differential system is then decided by applying to each equation $p_i = 0$ every of its non-admissible derivations $d \in \overline{\mu}(\theta_i, M_k)$ and computing the pseudo-remainder of $d\ p_i$ modulo $p_1, \ldots, p_s$ and their admissible derivatives. The restriction of the pseudo-division to admissible derivatives requires M_k to be Janet complete. If one of these pseudo-remainders is non-zero, then it is added as a new equation to the system, and the augmented system has to be treated by the algebraic part of Thomas' algorithm again.

Definition 10 A system of partial differential equations $\{\, p_1 = 0, \ldots, p_s = 0 \,\}$, where $p_1, \ldots, p_s \in R \setminus K$, is said to be *passive* if the following two conditions hold for $\mathrm{ld}(p_1) = \theta_1\, u_{k_1}, \ldots, \mathrm{ld}(p_s) = \theta_s\, u_{k_s}$, where $\theta_i \in \mathrm{Mon}(\Delta)$, $k_i \in \{1, \ldots, m\}$.

(a) For all $k \in \{1, \ldots, m\}$, the set M_k defined in (4.4) is Janet complete.
(b) For all $i \in \{1, \ldots, s\}$ and all $d \in \overline{\mu}(\theta_i, M_{k_i})$, the pseudo-remainder of $d\ p_i$ modulo $p_1, \ldots, p_s$ and their admissible derivatives is zero.

Definition 11 A differential system S, defined over R, as in (4.3) is said to be *simple* (with respect to $>$) if the following three conditions hold.

(a) The system S is simple as an algebraic system (in the finitely many variables $(u_i)_J$ which occur in the equations and inequations of S, totally ordered by $>$).
(b) The system $\{\, p_1 = 0, \ldots, p_s = 0 \,\}$ is passive.
(c) The left hand sides of the inequations $q_1 \neq 0, \ldots, q_t \neq 0$ equal their pseudo-remainders modulo $p_1, \ldots, p_s$ and their derivatives.

Definition 12 Let S be a differential system, defined over R. A *Thomas decomposition* of S (or of $\mathrm{Sol}_\Omega(S)$) with respect to $>$ is a collection of finitely many simple differential systems $S_1, \ldots, S_r$, defined over R, such that $\mathrm{Sol}_\Omega(S)$ is the disjoint union of the solution sets $\mathrm{Sol}_\Omega(S_1), \ldots, \mathrm{Sol}_\Omega(S_r)$.

Remark 9 Given S as in (4.3) and a ranking on R, a Thomas decomposition of S with respect to $>$ can be computed by interweaving the algebraic part discussed in Sect. 4.2.1 and differential reduction and completion with respect to Janet division.

First of all, a Thomas decomposition of S, considered as an algebraic system, is computed. Each of the resulting simple algebraic systems is then treated as follows. Differential pseudo-division is applied to pairs of distinct equations with leaders $\theta_1\, u_k$ and $\theta_2\, u_k$ such that $\theta_1 \mid \theta_2$ until either a non-zero pseudo-remainder is obtained or no such further reductions are possible. Non-zero pseudo-remainders are added to the system, the algebraic part of Thomas' algorithm is applied again, and the process is repeated. Once the system is auto-reduced in this sense, then it is possibly augmented with certain derivatives of equations so that the sets M_k defined in (4.4) are Janet complete. Then it is checked whether the system is passive. If a non-zero remainder is obtained by a pseudo-division of a non-admissible derivative modulo the equations and their admissible derivatives, then the algebraic part of Thomas' algorithm is

applied again to the augmented system. Otherwise, the system is passive. Finally, the left hand side of each inequation is replaced with its pseudo-remainder modulo the equations and their derivatives, in order to ensure condition (c) of Definition 11. The main reason why this procedure terminates is Dickson's Lemma, which shows that the ascending sequence of ideals of the semigroup $\mathrm{Mon}(\Delta)$ formed by the monomials θ defining leaders of equations (for each differential indeterminate) becomes stationary after finitely many steps.

For more details on the differential part of Thomas' algorithm, we refer to [4, 26], and [41, Subsect. 2.2.2].

An implementation of Thomas' algorithm for differential systems has been developed by M. Lange-Hegermann as Maple package `DifferentialThomas` [5].

We also use a simpler notation for the indeterminates $(u_k)_J$ of the differential polynomial ring. In case $m = 1$ we use the symbol u as a synonym for u_1. In addition, if the derivations $\partial_1, \partial_2, \partial_3$ represent the partial differential operators with respect to x, y, z, respectively, then we write

$$u\underbrace{x,\ldots,x}_{i}\underbrace{y,\ldots,y}_{j}\underbrace{z,\ldots,z}_{k}$$

instead of $u_{(i,j,k)}$.

When displaying a simple differential system we indicate next to each equation its set of admissible derivations.

Example 5 Let us consider the ordinary differential equation (which is discussed in [22, Example in Sect. 4.7])

$$\left(\frac{\partial u}{\partial x}\right)^3 - 4\,x\,u(x)\,\frac{\partial u}{\partial x} + 8\,u(x)^2 = 0.$$

The left hand side is represented by the element $p := u_x^3 - 4\,x\,u\,u_x + 8\,u^2$ of the differential polynomial ring $R = K\{u\}$ with one derivation ∂_x, where $K = \mathbb{Q}(x)$ is the field of rational functions in x, endowed with differentiation with respect to x.

The initial of p is constant, the separant of p is $3\,u_x^2 - 4\,x\,u$. The algebraic part of Thomas' algorithm only distinguishes the cases whether the discriminant of p vanishes or not. We have

$$\mathrm{disc}(p) = -\mathrm{res}(p, \mathrm{sep}(p), u_x) = -64\,\underline{u}^3\,(27\,\underline{u} - 4\,x^3).$$

This case distinction leads to the Thomas decomposition

$\underline{u_x}^3 - 4\,x\,u\,\underline{u_x} + 8\,u^2 = 0,\ \{\partial_x\}$ $(27\,\underline{u} - 4\,x^3)\,\underline{u} \neq 0$	$(27\,\underline{u} - 4\,x^3)\,\underline{u} = 0,\ \{\partial_x\}$

Since both systems contain only one equation, no differential reductions are necessary. The second simple system could be split into two with equations $27\,u - 4\,x^3 = 0$ and $u = 0$, respectively. The solutions of the first simple system are given by $u(x) = c\,(x - c)^2$, where c is an arbitrary non-zero constant. The solutions $u(x) = 0$ and $u(x) = \frac{4}{27}\,x^3$ of the second simple system are called *singular solutions*, the latter one being an envelope of the general solution.

Example 6 Let us compute a Thomas decomposition of the system of (nonlinear) partial differential equations

$$\begin{cases} \dfrac{\partial^2 u}{\partial x^2} - \dfrac{\partial^2 u}{\partial y^2} = 0, \\ \dfrac{\partial u}{\partial x} - u^2 = 0 \end{cases}$$

for one unknown function $u(x, y)$. The left hand sides are expressed as elements $p_1 := u_{x,x} - u_{y,y}$ and $p_2 := u_x - u^2$ of the differential polynomial ring $R = \mathbb{Q}\{u\}$ with commuting derivations ∂_x, ∂_y. We choose the degree-reverse lexicographical ranking $>$ on R with $\partial_x\,u > \partial_y\,u$ (cf. Example 2).

Since the monomial ∂_x defining the leader of p_2 divides the monomial ∂_x^2 defining the leader of p_1, differential pseudo-division is applied and p_1 is replaced with

$$p_3 := p_1 - \partial_x\,p_2 - 2\,u\,p_2 = -u_{y,y} + 2\,u^3.$$

Janet division associates the sets of admissible derivations to the equations of the resulting system as follows:

$$\begin{cases} \underline{u_x} - u^2 = 0, & \{\partial_x, \partial_y\} \\ \underline{u_{y,y}} - 2\,u^3 = 0, & \{*, \partial_y\} \end{cases}$$

The set of monomials $\{\partial_x, \partial_y^2\}$ defining the leaders u_x and $u_{y,y}$ is Janet complete. The check whether the above system is passive involves the following reduction:

$$\partial_x\,p_3 + \partial_y^2\,p_2 - 6\,u^2\,p_2 - 2\,u\,p_3 = -2\,(\underline{u_y} + u^2)\,(\underline{u_y} - u^2).$$

This non-zero remainder is a differential consequence which is added as an equation to the system. In fact, the system can be split into two systems according to the given factorization. For both systems a differential reduction of p_3 modulo the chosen factor is applied because the monomial ∂_y defining the new leader divides the monomial $\partial_{y,y}$ defining $\mathrm{ld}(p_3)$. In both cases the remainder is zero, the sets of monomials defining leaders are Janet complete, and the passivity check confirms formal integrability. We obtain the Thomas decomposition

$$\underline{u_x} - u^2 = 0,\ \{\partial_x, \partial_y\}$$
$$\underline{u_y} + u^2 = 0,\ \{*, \partial_y\}$$

$$\underline{u_x} - u^2 = 0,\ \{\partial_x, \partial_y\}$$
$$\underline{u_y} - u^2 = 0,\ \{*, \partial_y\}$$
$$\underline{u} \neq 0.$$

If the above factorization is ignored, then the discriminant of $p_4 := u_y^2 - u^4$ needs to be considered, which implies vanishing or non-vanishing of the separant $2\,u_y$. This case distinction leads to a different Thomas decomposition.

A Thomas decomposition of a differential system is not uniquely determined, as the previous example shows (cf. also Remark 3 for the algebraic case). In the special case of a system S of *linear* partial differential equations no case distinctions are necessary, and the single simple system in any Thomas decomposition of S is a Janet basis for S (cf., e.g., [18, 23, 36, 41]). Pseudo-reduction of a differential polynomial modulo the equations of a simple differential system and their derivatives decides membership to the corresponding saturation ideal (cf. also Proposition 1).

Proposition 2 ([41], Prop. 2.2.50) *Let S be a simple differential system, defined over R, with equations $p_1 = 0, \ldots, p_s = 0$. Moreover, let E be the differential ideal of R generated by $p_1, \ldots, p_s$ and define the product q of the initials and separants of all $p_1, \ldots, p_s$. Then $E : q^\infty$ is a radical differential ideal. Given $p \in R$, we have $p \in E : q^\infty$ if and only if the pseudo-remainder of p modulo $p_1, \ldots, p_s$ and their derivatives is zero.*

Similarly to the algebraic case, the Nullstellensatz for analytic functions (due to J. F. Ritt and H. W. Raudenbush, cf. [40, Sects. II.7–11, IX.27]) establishes a one-to-one correspondence of solution sets $V := \mathrm{Sol}_\Omega(S)$ of systems of partial differential equations $S = \{\, p_1 = 0, \ldots, p_s = 0 \,\}$ for m unknown functions, defined over R, and their vanishing ideals in $R = K\{u_1, \ldots, u_m\}$

$$\mathcal{I}_R(V) := \{\, p \in R \mid p(f) = 0 \text{ for all } f \in V \,\}.$$

These are the radical differential ideals of R. The Nullstellensatz implies that, with the notation of Proposition 2, we have $\mathcal{I}_R(\mathrm{Sol}_\Omega(S)) = E : q^\infty$.

The following proposition allows to decide whether a given differential equation $p = 0$ is a consequence of a (not necessarily simple) differential system S by applying pseudo-division to p modulo each of the simple systems in a Thomas decomposition of S. It follows from the previous proposition and the Nullstellensatz and it also applies to algebraic systems by ignoring the separants.

Proposition 3 ([41], Prop. 2.2.72) *Let S be a (not necessarily simple) differential system as in (4.3) and $S_1, \ldots, S_r$ a Thomas decomposition of S with respect to any ranking on R. Moreover, let E be the differential ideal of R generated by $p_1, \ldots, p_s$ and define the product q of $q_1, \ldots, q_t$. For $i \in \{1, \ldots, r\}$, let $E^{(i)}$ be the differential ideal of R generated by the equations in S_i and define the product $q^{(i)}$ of the initials and separants of all these equations. Then we have*

$$\sqrt{E : q^\infty} = \left(E^{(1)} : (q^{(1)})^\infty\right) \cap \cdots \cap \left(E^{(r)} : (q^{(r)})^\infty\right).$$

4.3 Elimination

Thomas' algorithm can be used to solve various differential elimination problems. This section presents results on certain rankings on the differential polynomial ring $R = K\{u_1, \ldots, u_m\}$ which allow to compute all differential consequences of a given differential system involving only a specified subset of the differential indeterminates $u_1, \ldots, u_m$. In other words, this technique allows to determine all differential equations which are satisfied by certain components of the solution tuples. We adopt the notation from the previous section.

Definition 13 Let I_1, I_2, ..., I_k form a partition of $\{1, 2, \ldots, m\}$ such that $i_1 \in I_{j_1}$, $i_2 \in I_{j_2}$, $i_1 \leq i_2$ implies $j_1 \leq j_2$. Let $B_j := \{u_i \mid i \in I_j\}$, $j = 1, \ldots, k$. Moreover, fix some degree-reverse lexicographical ordering $>$ on $\mathrm{Mon}(\Delta)$. Then the *block ranking* on R *with blocks* $B_1, \ldots, B_k$ (with $u_1 > u_2 > \cdots > u_m$) is defined for $\theta_1 u_{i_1}$, $\theta_2 u_{i_2} \in \mathrm{Mon}(\Delta)\, u$, where $u_{i_1} \in B_{j_1}$, $u_{i_2} \in B_{j_2}$, by

$$\theta_1 u_{i_1} > \theta_2 u_{i_2} \quad :\Longleftrightarrow \quad \left\{ \begin{array}{l} j_1 < j_2 \quad \text{or} \quad \Big(\; j_1 = j_2 \quad \text{and} \quad \big(\theta_1 > \theta_2 \quad \text{or} \\ \quad (\theta_1 = \theta_2 \quad \text{and} \quad i_1 < i_2)\big) \Big). \end{array} \right.$$

Such a ranking is said to satisfy $B_1 \gg B_2 \gg \cdots \gg B_k$.

Example 7 With respect to the block ranking on $K\{u_1, u_2, u_3\}$ with blocks $\{u_1\}$, $\{u_2, u_3\}$ (and $u_1 > u_2 > u_3$) we have $(u_1)_{(0,1)} > u_1 > (u_2)_{(1,2)} > (u_3)_{(1,2)} > (u_2)_{(0,1)}$.

In the situation of the previous definition, for every $i \in \{1, \ldots, k\}$, we consider $K\{B_i, \ldots, B_k\} := K\{u \mid u \in B_i \cup \cdots \cup B_k\}$ as a differential subring of R, endowed with the restrictions of the derivations ∂_1, ..., ∂_n to $K\{B_i, \ldots, B_k\}$.

For any algebraic or differential system S we denote by $S^=$ (resp. $S^{\neq}$) the set of the left hand sides of all equations (resp. inequations) in S.

Proposition 4 ([41], Prop. 3.1.36) *Let S be a simple differential system, defined over R, with respect to a block ranking with blocks B_1, ..., B_k. Moreover, let E be the differential ideal of R generated by $S^=$ and q the product of the initials and separants of all elements of $S^=$. For every $i \in \{1, \ldots, k\}$, let E_i be the differential ideal of $K\{B_i, \ldots, B_k\}$ generated by $P_i := S^= \cap K\{B_i, \ldots, B_k\}$ and let q_i be the product of the initials and separants of all elements of P_i. Then, for every $i \in \{1, \ldots, k\}$, we have*

$$(E : q^\infty) \cap K\{B_i, \ldots, B_k\} = E_i : q_i^\infty.$$

In other words, the differential equations implied by S which involve only the differential indeterminates in $B_i \cup \cdots \cup B_k$ are precisely those whose pseudo-remainders modulo the elements of $S^= \cap K\{B_i, \ldots, B_k\}$ and their derivatives are zero.

Example 8 The Cauchy–Riemann equations for a complex function of $z = x + i\,y$ with real part u and imaginary part v are

$$\begin{cases} \dfrac{\partial u}{\partial x} - \dfrac{\partial v}{\partial y} = 0, \\ \dfrac{\partial u}{\partial y} + \dfrac{\partial v}{\partial x} = 0. \end{cases}$$

The left hand sides are represented by the elements $p_1 := u_x - v_y$ and $p_2 := u_y + v_x$ of the differential polynomial ring $R = \mathbb{Q}\{u, v\}$ with derivations ∂_x and ∂_y. Choosing a block ranking on R satisfying $\{u\} \gg \{v\}$, the passivity check yields the equation

$$\partial_x\, p_2 - \partial_y\, p_1 = v_{x,x} + v_{y,y} = 0.$$

Similarly, the choice of a block ranking on R satisfying $\{v\} \gg \{u\}$ yields the consequence $u_{x,x} + u_{y,y} = 0$. These computations confirm that the real and imaginary parts of a holomorphic function are harmonic functions.

Corollary 1 ([41], Cor. 3.1.37) *Let S be a (not necessarily simple) differential system, defined over R, and $S_1, \ldots, S_r$ a Thomas decomposition of S with respect to a block ranking with blocks $B_1, \ldots, B_k$. Moreover, let E be the differential ideal of R generated by $S^{=}$ and q the product of all elements of $S^{\neq}$. Let $i \in \{1, \ldots, k\}$ be fixed. For every $j \in \{1, \ldots, r\}$, let $E^{(j)}$ be the differential ideal of $K\{B_i, \ldots, B_k\}$ generated by $P_j := S_j^{=} \cap K\{B_i, \ldots, B_k\}$ and let $q^{(j)}$ be the product of the initials and separants of all elements of P_j. Then we have*

$$\sqrt{E : q^{\infty}} \cap K\{B_i, \ldots, B_k\} = (E_1 : q_1^{\infty}) \cap \ldots \cap (E_r : q_r^{\infty}).$$

4.4 Control-Theoretic Applications

In order to apply the Thomas decomposition method to nonlinear control systems, we assume that the control system is given by differential equations and inequations whose left hand sides are polynomials. Structural information about certain configurations of the control system is obtained from each simple system of a Thomas decomposition of the given differential equations and inequations. The choice of ranking on the differential polynomial ring depends on the question at hand, although a Thomas decomposition with respect to any ranking, e.g., the degree-reverse lexicographical ranking, may give hints on how to adapt the ranking for further investigations in a certain direction.

Let $R = K\{U\}$ be the differential polynomial ring in the differential indeterminates $U := \{u_1, \ldots, u_m\}$ over a differential field K of (complex) meromorphic functions on an open and connected subset Ω of $\mathbb{C}^n$ (cf. Sect. 4.2.2). (No distinction is made a priori between state variables, input, output, etc.)

We assume that S is a simple differential system, defined over R, with respect to some ranking $>$. Let E be the differential ideal of R generated by the set $S^=$ of the left hand sides of the equations in S and define the product q of the initials and separants of all elements of $S^=$.

Definition 14 Let $x \in U$ and $Y \subseteq U \setminus \{x\}$. Then x is said to be *observable with respect to* Y if there exists $p \in (E : q^\infty) \setminus \{0\}$ such that p is a polynomial in x (not involving any proper derivative of x) with coefficients in $K\{Y\}$ and such that neither its leading coefficient nor $\partial p/\partial x$ is an element of $E : q^\infty$.

Remark 10 Let p be a polynomial as in the previous definition. Then the implicit function theorem allows to solve $p = 0$ locally for x in the sense that the component of $(f_1, \ldots, f_m) \in \mathrm{Sol}_\Omega(S)$ corresponding to x can locally be expressed as an analytic function of the components corresponding to the differential indeterminates in Y.

If $>$ satisfies $U \setminus (Y \cup \{x\}) \gg \{x\} \gg \{Y\}$, then by Proposition 4, there exists a polynomial p in $(E : q^\infty) \setminus \{0\}$ as above if and only if there exists such a polynomial in $S^= \cap K\{Y \cup \{x\}\}$. For a not necessarily simple differential system S, a Thomas decomposition with respect to a ranking as above allows to decide the existence of such a polynomial among the left hand sides of the differential consequences of S by inspecting each simple system (cf. Corollary 1).

Definition 15 A subset Y of U is called a *flat output* of S if $(E : q^\infty) \cap K\{Y\} = \{0\}$ and every $x \in U \setminus Y$ is observable with respect to Y.

Remark 11 Let $>$ satisfy $U \setminus Y \gg Y$. Then Proposition 4 allows to decide whether the conditions in Definition 15 are satisfied by checking that $S^= \cap K\{Y\} = \emptyset$ holds and that for every $x \in U \setminus Y$ there exists a polynomial $p \in S^= \cap K\{Y \cup \{x\}\}$ satisfying the conditions in Definition 14.

If the differential ideal $I := E : q^\infty$ is prime, then the field of fractions $\mathrm{Quot}(R/I)$ can be considered as a differential extension field of K. Let us assume that Y is a flat output of S and let L be the differential subfield of $\mathrm{Quot}(R/I)$ which is generated by $\{\, y + I \mid y \in Y \,\}$. Then, by Definition 15, L/K is a purely differentially transcendental extension of differential fields, and for every $x \in U \setminus Y$, the element $x + I$ of $\mathrm{Quot}(R/I)$ is algebraic over L. Hence, $\{\, y + I \mid y \in Y \,\}$ is a differential transcendence basis of $\mathrm{Quot}(R/I)/K$, and the system is flat in the sense of [14, Sect. 3.2].

Remark 12 Following [14], a system which is defined by a differential field extension is called flat if it is equivalent by endogenous feedback to a system which is defined by a purely differentially transcendental extension of differential fields. As opposed to checking whether Y is a flat output of S using the method described above, deciding whether S is flat is a difficult problem in general.

As a first illustration of how differential elimination methods can be applied to nonlinear control systems, we consider inversion, i.e., the problem of expressing the input variables in terms of the output variables (and their derivatives).

Remark 13 Using the same notation as above, we assume that disjoint subsets Y and Z of U are specified, where the differential indeterminates in Y and Z are interpreted as input and output variables of the system, respectively. We achieve an *inversion* of the system S if and only if we can exhibit for each $z \in Z$ a $p \in (E : q^\infty) \setminus \{0\}$ such that p is a polynomial in z (not involving any proper derivative of z) with coefficients in $K\{Y\}$ and such that neither its leading coefficient nor $\partial p/\partial z$ is an element of $E : q^\infty$.

If $>$ satisfies $U \setminus (Y \cup \{z\}) \gg \{z\} \gg Y$, then by Proposition 4, there exists such a polynomial p in $(E : q^\infty) \setminus \{0\}$ if and only if there exists such a polynomial in $S^= \cap K\{Y \cup \{z\}\}$. A block ranking $U \setminus (Y \cup Z) \gg Z \gg Y$ may allow to find such polynomials p for all $z \in Z$ by computing only one Thomas decomposition (cf. the following example), for instance, if all these polynomials p have degree one.

For displaying simple differential systems resulting from Thomas decompositions in a concise way, we use the following command `Print`, which makes use of both the Maple packages `Janet` [6] and `DifferentialThomas` [5], where `ivar` and `dvar` are the lists of independent and dependent variables, respectively.

```
> with(Janet):
> Print := S->Diff2Ind(
> PrettyPrintDifferentialSystem(S), ivar, dvar):
```

The sets of admissible derivations for the equations in a simple system are not reproduced here. Note that the implementation uses factorization and may, for convenience, return simple systems containing several inequations with the same leader (thus, not strictly complying with Condition (b) of Definition 3).

Example 9 The following system of ordinary differential equations models a unicycle as described in [9, Examples 3.20, 4.18, 5.10] (cf. also, e.g., [33, Example 2.35]).

$$\begin{cases} \dot{x}_1 = \cos(x_3)\, u_1, \\ \dot{x}_2 = \sin(x_3)\, u_1, \\ \dot{x}_3 = u_2. \end{cases}$$

Here x_1, x_2, x_3 are considered as state variables, where (x_1, x_2) is the position of the middle of the axis in the plane and x_3 the angle of its rotation, and the velocities u_1, u_2 are considered as inputs. Moreover, the following outputs y_1, y_2 are given:

$$\begin{cases} y_1 = x_1\,, \\ y_2 = x_2. \end{cases}$$

The task is to try to invert the system, i.e., to express u_1, u_2 in terms of y_1, y_2 and their derivatives.

In order to translate the given equations into differential polynomials, we represent $\cos(x_3)$ and $\sin(x_3)$ by differential indeterminates *cx3* and *sx3* and add the generating relations

$$cx3^2 + sx3^2 = 1, \quad cx3_t = -sx3\,(x_3)_t, \quad sx3_t = cx3\,(x_3)_t$$

to the system. More precisely speaking, we adjoin to the differential polynomial ring $\mathbb{Q}\{x_1, x_2, x_3, u_1, u_2, y_1, y_2\}$ with derivation ∂_t the differential indeterminates *cx3* and *sx3* and define the differential ideal E of the resulting differential polynomial ring which is generated by $(x_1)_t - cx3\,u_1$, $(x_2)_t - sx3\,u_1$, $(x_3)_t - u_2$, $cx3^2 + sx3^2 - 1$, $cx3_t + sx3\,(x_3)_t$ and $sx3_t - cx3\,(x_3)_t$. We then apply elimination properties of the differential Thomas decomposition method to $\sqrt{E}$ (see Proposition 3 and Corollary 1).

(Alternatively, if one accepts neglecting the particular case of movement of the unicycle in the direction of the x_2-coordinate axis without rotation, one could assume that $\cos(x_3)$ is not the zero function, multiply both sides of the equation $\dot{x}_3 = u_2$ by $\cos(x_3)$, read the resulting left hand side, using the chain rule, as the derivative of $\sin(x_3)$, and obtain the equation $sx3_t = cx3\,u_2$. This would allow to dispose of the differential indeterminate x_3 and the computation of the Thomas decomposition below would essentially yield the first five of the seven simple systems below.)

In the following computations concerning the model of a unicycle the differential polynomial ring is $\mathbb{Q}\{x_1, x_2, cx3, sx3, x_3, u_1, u_2, y_1, y_2\}$ with one derivation ∂_t.

```
> with(DifferentialThomas):
> ivar := [t]:
> dvar := [x1,x2,cx3,sx3,x3,u1,u2,y1,y2]:
```

We specify the block ranking $>$ satisfying $\{x_1, x_2, cx3, sx3, x_3\} \gg \{u_1, u_2\} \gg \{y_1, y_2\}$ as well as $x_1 > x_2 > cx3 > sx3 > x_3$ and $u_1 > u_2$ and $y_1 > y_2$.

```
> ComputeRanking(ivar,
> [[x1,x2,cx3,sx3,x3],[u1,u2],[y1,y2]]):
```

If the left hand sides of the system are written in jet notation, then a conversion into the format expected by the package `DifferentialThomas` is accomplished by the following sequence of commands.

```
> L := [x1[t]-cx3*u1, x2[t]-sx3*u1, x3[t]-u2,
> y1-x1, y2-x2, cx3^2+sx3^2-1, cx3[t]+sx3*x3[t],
> sx3[t]-cx3*x3[t]];
> LL := Diff2JetList(Ind2Diff(L, ivar, dvar));
```

$LL := [(x_1)_1 - cx3_0\,(u_1)_0,\quad (x_2)_1 - sx3_0\,(u_1)_0,\quad (x_3)_1 - (u_2)_0,\quad (y_1)_0 - (x_1)_0,$
$(y_2)_0 - (x_2)_0,\quad {cx3_0}^2 + {sx3_0}^2 - 1,\quad cx3_1 + sx3_0\,(x_3)_1,\quad sx3_1 - cx3_0\,(x_3)_1]$

We compute a Thomas decomposition with respect to $>$ of the given system of ordinary differential equations.

```
> TD := DifferentialThomasDecomposition(LL,[]);
```

$TD :=$ [*DifferentialSystem*, *DifferentialSystem*, *DifferentialSystem*,
DifferentialSystem, *DifferentialSystem*, *DifferentialSystem*, *DifferentialSystem*]

The first simple differential system is given as follows.

```
> Print(TD[1]);
```

$$[\underline{x_1} - y_1 = 0,\quad \underline{x_2} - y_2 = 0,\quad u_1\,\underline{cx3} - (y_1)_t = 0,\quad u_1\,\underline{sx3} - (y_2)_t = 0,$$
$$(y_1)_t^2\,\underline{(x_3)_t} + (y_2)_t^2\,\underline{(x_3)_t} - (y_1)_t\,(y_2)_{t,t} + (y_2)_t\,(y_1)_{t,t} = 0,$$
$$\underline{{u_1}^2} - (y_1)_t^2 - (y_2)_t^2 = 0,\quad (y_1)_t^2\,\underline{u_2} + (y_2)_t^2\,\underline{u_2} - (y_1)_t\,(y_2)_{t,t} + (y_2)_t\,(y_1)_{t,t} = 0,$$
$$\underline{(y_2)_t} \neq 0,\quad \underline{(y_1)_t} \neq 0,\quad \underline{{(y_1)_t}^2} + {(y_2)_t}^2 \neq 0,\quad (y_2)_t\,\underline{(y_1)_{t,t}} - (y_1)_t\,(y_2)_{t,t} \neq 0]$$

```
> collect(%[7], u2, factor);
```

$$\left((y_1)_t^2 + (y_2)_t^2\right) u_2 - (y_1)_t\,(y_2)_{t,t} + (y_2)_t\,(y_1)_{t,t} = 0$$

Thus, the equations with leader u_1 and u_2 in $TD[1]$ allow to express u_1 and u_2 in terms of y_1 and y_2. (Up to solving these equations for u_1 and u_2 explicitly, it is the same result as in [9, Example 5.12].)

The remaining six simple differential systems describe particular configurations, which exhibit obstructions to invertibility.

```
> Print(TD[2]);
```

$$[\underline{x_1} - y_1 = 0,\quad \underline{x_2} - y_2 = 0,\quad u_1\,\underline{cx3} - (y_1)_t = 0,\quad u_1\,\underline{sx3} - (y_2)_t = 0,\quad (x_3)_t = 0,$$
$$\underline{{u_1}^2} - (y_1)_t^2 - (y_2)_t^2 = 0,\quad \underline{u_2} = 0,\quad (y_2)_t\,\underline{(y_1)_{t,t}} - (y_1)_t\,(y_2)_{t,t} = 0,$$
$$\underline{(y_2)_t} \neq 0,\quad \underline{(y_1)_t} \neq 0,\quad \underline{{(y_1)_t}^2} + {(y_2)_t}^2 \neq 0]$$

The vanishing of the Wronskian determinant of $(y_1)_t$ and $(y_2)_t$ expresses that one of the velocities $\dot{x}_1$ and $\dot{x}_2$ is a constant multiple of the other. Hence, no rotation is allowed, which forces the input u_2 to be the zero function. Due to the inequations, the vector $(\dot{x}_1, \dot{x}_2)$ is non-zero and not parallel to any of the x_1- or x_2-coordinate axes.

```
> Print(TD[3]);
```

$$[\underline{x_1} - y_1 = 0,\quad \underline{x_2} - y_2 = 0,\quad \underline{cx3} + 1 = 0,\quad \underline{sx3} = 0,\quad \underline{(x_3)_t} = 0,$$
$$\underline{u_1} + (y_1)_t = 0,\quad \underline{u_2} = 0,\quad \underline{(y_2)_t} = 0,\quad \underline{(y_1)_t} \neq 0]$$

```
> Print(TD[4]);
```

$$[\underline{x_1} - y_1 = 0,\quad \underline{x_2} - y_2 = 0,\quad \underline{cx3} - 1 = 0,\quad \underline{sx3} = 0,\quad \underline{(x_3)_t} = 0,$$
$$\underline{u_1} - (y_1)_t = 0,\quad \underline{u_2} = 0,\quad \underline{(y_2)_t} = 0,\quad \underline{(y_1)_t} \neq 0]$$

The previous two simple systems describe cases in which only movement in any of the two directions defined by the x_1-coordinate axis is allowed and no rotation.

```
> Print(TD[5]);
```

$$[\underline{x_1} - y_1 = 0,\quad \underline{x_2} - y_2 = 0,\quad \underline{cx3}^2 + \underline{sx3}^2 - 1 = 0,\quad \underline{sx3}_t - u_2\,\underline{cx3} = 0,$$
$$\underline{(x_3)_t} - u_2 = 0,\quad \underline{u_1} = 0,\quad \underline{(y_1)_t} = 0,\quad \underline{(y_2)_t} = 0,\quad \underline{sx3} + 1 \neq 0,\quad \underline{sx3} - 1 \neq 0]$$

The fifth simple system describes configurations which only allow rotation, and the input u_1 is forced to be the zero function. (Similarly to Example 1, the inequation $sx3^2 - 1 \neq 0$ is introduced here to ensure that $cx3^2 + sx3^2 - 1$ has no multiple roots as polynomial in $cx3$. It is included in the simple system in factorized form.)

```
> Print(TD[6]);
```

$$[\underline{x_1} - y_1 = 0,\quad \underline{x_2} - y_2 = 0,\quad \underline{cx3} = 0,\quad \underline{sx3} + 1 = 0,\quad \underline{(x_3)_t} = 0,$$
$$\underline{u_1} + (y_2)_t = 0,\quad \underline{u_2} = 0,\quad \underline{(y_1)_t} = 0]$$

```
> Print(TD[7]);
```

$$[\underline{x_1} - y_1 = 0,\quad \underline{x_2} - y_2 = 0,\quad \underline{cx3} = 0,\quad \underline{sx3} - 1 = 0,\quad \underline{(x_3)_t} = 0,$$
$$\underline{u_1} - (y_2)_t = 0,\quad \underline{u_2} = 0,\quad \underline{(y_1)_t} = 0]$$

The last two simple systems cover the cases of movement in any of the two directions defined by the x_2-coordinate axis and no rotation.

Next we consider the detection of flat outputs.

Example 10 A model of a 2-D crane is given by the following system of ordinary differential equations (cf. [14, Sect. 4.1] and the references therein), where $x(t)$ and $z(t)$ are the coordinates of the load of mass m, $\theta(t)$ is the angle between the rope and the z-axis, $d(t)$ the trolley position, $T(t)$ the tension of the rope, $R(t)$ the rope length, and g the gravitational constant.

$$\begin{cases} m\,\ddot{x} = -T\,\sin\theta, \\ m\,\ddot{z} = -T\,\cos\theta + m\,g, \\ x = R\,\sin\theta + d, \\ z = R\,\cos\theta. \end{cases}$$

The task is to decide whether $\{x, z\}$ is a flat output of the system.

Similarly to the previous example, we represent $\cos\theta$ and $\sin\theta$ by differential indeterminates c and s and add the generating relation $c^2 + s^2 = 1$ to the system. In this example the given equations depend on θ only through $\cos\theta$ and $\sin\theta$. Therefore, we do not include θ as a differential indeterminate and do not need to add the relations $c_t = -s\,\theta_t$ and $s_t = c\,\theta_t$ to the system. (Note that, if I is the differential ideal of $\mathbb{Q}\{\theta, c, s\}$ with derivation ∂_t which is generated by $c^2 + s^2 - 1$ and $c_t + s\,\theta_t$ and $s_t - c\,\theta_t$, then $I \cap \mathbb{Q}\{c, s\}$ is the differential ideal which is generated by $c^2 + s^2 - 1$.)

```
> with(DifferentialThomas):
> ivar := [t]:
> dvar := [T,c,s,d,R,x,z]:
```

We set up the block ranking $>$ which satisfies $\{T, c, s, d, R\} \gg \{x, z\}$ as well as $T > c > s > d > R$ and $x > z$.

```
> ComputeRanking(ivar, [[T,c,s,d,R],[x,z]]):
```

We compute a Thomas decomposition with respect to $>$. (As is customary in Maple, the symbols m and g are treated here as algebraically independent over $\mathbb{Q}$. More precisely, the ground field for the following computation is the differential field $\mathbb{Q}(m, g)$ with trivial derivation.)

```
> TD := DifferentialThomasDecomposition(
> [m*x[2]+T[0]*s[0], m*z[2]+T[0]*c[0]-m*g,
> x[0]-R[0]*s[0]-d[0], z[0]-R[0]*c[0],
> c[0]^2+s[0]^2-1], []);
```

$TD :=$ [*DifferentialSystem*, *DifferentialSystem*, *DifferentialSystem*, *DifferentialSystem*, *DifferentialSystem*, *DifferentialSystem*, *DifferentialSystem*]

The second simple differential system is given as follows.

```
> Print(TD[2]);
```

$$[z\,\underline{T} + m\,z_{t,t}\,R - m\,g\,R = 0,\quad R\,\underline{c} - z = 0,\quad z_{t,t}\,R\,\underline{s} - g\,R\,\underline{s} - z\,x_{t,t} = 0,$$
$$z_{t,t}\,\underline{d} - g\,\underline{d} + z\,x_{t,t} - x\,z_{t,t} + g\,x = 0,$$
$$z_{t,t}{}^2\,\underline{R}^2 - 2\,g\,z_{t,t}\,\underline{R}^2 + g^2\,\underline{R}^2 - z^2\,x_{t,t}^2 - z^2\,z_{t,t}^2 + 2\,g\,z^2\,z_{t,t} - g^2\,z^2 = 0,$$
$$\underline{z} \neq 0,\quad \underline{z_{t,t}} - g \neq 0,\quad \underline{x_{t,t}} \neq 0,\quad \underline{x_{t,t}}{}^2 + z_{t,t}{}^2 - 2\,g\,z_{t,t} + g^2 \neq 0]$$

```
> collect(%[5], R, factor);
```

$$\left(z_{t,t} - g\right)^2 R^2 - z^2\left(x_{t,t}{}^2 + z_{t,t}{}^2 - 2\,g\,z_{t,t} + g^2\right) = 0$$

We observe that this simple system S contains no equation involving derivatives of x and z only. Moreover, the equations in S show that T, c, s, d, R are observable with respect to $\{x, z\}$. Hence, $\{x, z\}$ is a flat output of S.

The remaining six simple differential systems describe particular configurations for which $\{x, z\}$ is not a flat output. In fact, the movement of the load is restricted by some constraint in these cases (e.g., $x_{t,t} = 0$ or $z = 0$, one reason being, e.g., that vanishing rope tension implies constant acceleration of the load, another being a constant rope length of zero allowing no vertical movement of the load). We do not consider the system to be controllable under these conditions.

```
> Print(TD[1]);
```

$$[\underline{T} = 0,\quad R\,\underline{c} - z = 0,\quad R\,\underline{s} + d - x = 0,\quad \underline{d}^2 - 2\,x\,\underline{d} + x^2 - R^2 + z^2 = 0,$$
$$\underline{x_{t,t}} = 0,\quad \underline{z_{t,t}} - g = 0,\quad \underline{z} \neq 0,\quad \underline{R} \neq 0,\quad \underline{R} + z \neq 0,\quad \underline{R} - z \neq 0]$$

```
> Print(TD[3]);
```

$$[\underline{T} - m\,z_{t,t} + m\,g = 0,\quad \underline{c} + 1 = 0,\quad \underline{s} = 0,\quad \underline{d} - x = 0,\quad \underline{R} + z = 0,$$
$$\underline{x_{t,t}} = 0,\quad \underline{z} \neq 0]$$

```
> Print(TD[4]);
```

$$[\underline{T} + m\,z_{t,t} - m\,g = 0,\quad \underline{c} - 1 = 0,\quad \underline{s} = 0,\quad \underline{d} - x = 0,\quad \underline{R} - z = 0,$$
$$\underline{x_{t,t}} = 0,\quad \underline{z} \neq 0]$$

```
> Print(TD[5]);
```

$$[s\,\underline{T} + m\,x_{t,t} = 0,\quad x_{t,t}\,\underline{c} + g\,s = 0,\quad g^2\,\underline{s}^2 + x_{t,t}^2\,\underline{s}^2 - x_{t,t}^2 = 0,\quad \underline{d} - x = 0,\quad \underline{R} = 0,$$
$$\underline{z} = 0,\quad \underline{x_{t,t}} \neq 0,\quad \underline{x_{t,t}}{}^2 + g^2 \neq 0]$$

```
> Print(TD[6]);
```

$$[\underline{T} + m\,g = 0,\quad \underline{c} + 1 = 0,\quad \underline{s} = 0,\quad \underline{d} - x = 0,\quad \underline{R} = 0,\quad \underline{x_{t,t}} = 0,\quad \underline{z} = 0]$$

```
> Print(TD[7]);
```

$$[\underline{T} - m\,g = 0,\quad \underline{c} - 1 = 0,\quad \underline{s} = 0,\quad \underline{d} - x = 0,\quad \underline{R} = 0,\quad \underline{x_{t,t}} = 0,\quad \underline{z} = 0]$$

We give two examples which demonstrate how the Thomas decomposition technique can be used to study the dependence of structural properties of a nonlinear control system on parameters.

Example 11 A model of a continuous stirred-tank reactor (cf. [25, Example 1.2]) is given by the differential system

$$\begin{cases} \dot{V}(t) = F_1(t) + F_2(t) - k\sqrt{V(t)}, \\ \overline{c(t)\,\dot{V}(t)} = c_1\,F_1(t) + c_2\,F_2(t) - c(t)\,k\,\sqrt{V(t)}. \end{cases}$$

A dissolved material has concentration $c(t)$ in the tank and it is fed through two inputs with constant concentrations c_1 and c_2 and flow rates $F_1(t)$ and $F_2(t)$, respectively. There exists an outward flow with a flow rate proportional to the square root of the volume $V(t)$ of liquid in the tank. Moreover, k is an experimental constant.

In order to eliminate the square root of the volume in the given equations, we represent $\sqrt{V(t)}$ as a differential indeterminate sV and substitute other occurrences of $V(t)$ by sV^2. We investigate the dependence of the behavior on parameter configurations by considering c_1 and c_2 as differential indeterminates as well and adding the conditions $\dot{c}_1 = 0$ and $\dot{c}_2 = 0$.

```
> with(DifferentialThomas):
> ivar := [t]:
> dvar := [F1,F2,sV,c,c1,c2]:
```

We define $R = \mathbb{Q}\{F_1, F_2, sV, c, c_1, c_2\}$ and choose the block ranking $>$ on R with blocks $\{F_1, F_2\}$, $\{sV, c\}$, $\{c_1, c_2\}$, i.e., satisfying $\{F_2, F_2\} \gg \{sV, c\} \gg \{c_1, c_2\}$ and $F_1 > F_2$ and $sV > c$ and $c_1 > c_2$.

```
> ComputeRanking(ivar, [[F1,F2],[sV,c],[c1,c2]]):
> L := [2*sV[t]*sV-F1-F2+k*sV,
> c[t]*sV^2-c2*F2+c*k*sV-c1*F1+2*c*sV[t]*sV,
> c1[t], c2[t]]:

> LL := Diff2JetList(Ind2Diff(L, ivar, dvar));
```

$$LL := [2\,sV_1\,sV_0 - (F_1)_0 - (F_2)_0 + k\,sV_0,$$
$$c_1\,{sV_0}^2 - (c_2)_0\,(F_2)_0 + c_0\,k\,sV_0 - (c_1)_0\,(F_1)_0 + 2\,c_0\,sV_1\,sV_0, \quad (c_1)_1, \quad (c_2)_1]$$

We compute a Thomas decomposition with respect to $>$ of the given system of ordinary differential equations, to which we add the inequations $\sqrt{V} \neq 0$, $c_1 \neq 0$, $c_2 \neq 0$ to exclude trivial cases.

```
> TD := DifferentialThomasDecomposition(LL,
> [sV[0],c1[0],c2[0]]);
```

$$TD := [\mathit{DifferentialSystem}, \mathit{DifferentialSystem}, \mathit{DifferentialSystem}]$$

The first simple differential system is given as follows.

```
> Print(TD[1]);
```

$$[c_2\,\underline{F_1} - c_1\,\underline{F_1} + 2\,c\,sVsV_t - 2\,c_2\,sVsV_t + c_t\,sV^2 + c\,k\,sV - c_2\,k\,sV = 0,$$
$$c_1\,\underline{F_2} - c_2\,\underline{F_2} + 2\,c\,sVsV_t - 2\,c_1\,sVsV_t + c_t\,sV^2 + c\,k\,sV - c_1\,k\,sV = 0,$$
$$\underline{(c_1)_t} = 0,\quad \underline{(c_2)_t} = 0,\quad \underline{c_2} \neq 0,\quad \underline{c_1} \neq 0,\quad \underline{c_1} - c_2 \neq 0,\quad \underline{sV} \neq 0]$$

```
> collect(%[1], F1);
```

$$(c_2 - c_1)\,F_1 + 2\,c\,sVsV_t - 2\,c_2\,sVsV_t + c_t\,sV^2 + c\,k\,sV - c_2\,k\,sV = 0$$

```
> collect(%%[2], F2);
```

$$(c_1 - c_2)\,F_2 + 2\,c\,sVsV_t - 2\,c_1\,sVsV_t + c_t\,sV^2 + c\,k\,sV - c1\,k\,sV = 0$$

The first two equations in the first simple system S show that F_1 and F_2 are observable with respect to $\{c, sV\}$. (Although c_1 and c_2 are represented by differential indeterminates here, we consider these still as parameters.) Let E be the differential ideal of R generated by $S^=$ and q the product of the initials (and separants) of all elements of $S^=$. Due to the choice of the block ranking, we conclude that we have $(E : q^\infty) \cap \mathbb{Q}\{sV, c\} = \{0\}$ (cf. Proposition 4). Hence, $\{c, sV\}$ is a flat output of S.

The remaining two simple systems describe configurations of the system in which the two concentrations c_1 and c_2 are equal. Since both input feeds are identical and constant, this condition precludes control of the concentration in the tank. These particular systems do not admit $\{c, sV\}$ as a flat output. In fact, by inspecting the equations of these systems, we observe that we have $(E : q^\infty) \cap \mathbb{Q}\{sV, c\} \neq \{0\}$.

```
> Print(TD[2]);
```

$$[c\,\underline{F_1} - c_2\,\underline{F_1} + c\,F_2 - c_2\,F_2 + c_t\,sV^2 = 0,$$
$$2\,c\,\underline{sV_t} - 2\,c_2\,\underline{sV_t} + c_t\,sV + c\,k - c_2\,k = 0,\quad \underline{c_1} - c_2 = 0,\quad \underline{(c_2)_t} = 0,$$
$$\underline{c_2} \neq 0,\quad \underline{c} - c_2 \neq 0,\quad \underline{sV} \neq 0]$$

```
> Print(TD[3]);
```

$$[\underline{F_1} + F_2 - 2\,sVsV_t - k\,sV = 0,\quad \underline{c} - c_2 = 0,\quad \underline{c_1} - c_2 = 0,\quad \underline{(c_2)_t} = 0,$$
$$\underline{c_2} \neq 0,\quad \underline{sV} \neq 0]$$

Example 12 Let us consider the following system of linear partial differential equations for functions ξ_1, ξ_2, ξ_3 of $\mathbf{x} = (x_1, x_2, x_3)$ involving a parametric function $a(x_2)$

$$\begin{cases} -a(x_2)\dfrac{\partial \xi_1(\mathbf{x})}{\partial x_1} + \dfrac{\partial \xi_3(\mathbf{x})}{\partial x_1} - \left(\dfrac{\partial}{\partial x_2} a(x_2)\right) \xi_2(\mathbf{x}) + \dfrac{1}{2}\, a(x_2)\,(\nabla \cdot \xi(\mathbf{x})) = 0, \\ -a(x_2)\dfrac{\partial \xi_1(\mathbf{x})}{\partial x_2} + \dfrac{\partial \xi_3(\mathbf{x})}{\partial x_2} = 0, \\ -a(x_2)\dfrac{\partial \xi_1(\mathbf{x})}{\partial x_3} + \dfrac{\partial \xi_3(\mathbf{x})}{\partial x_3} - \dfrac{1}{2}\,(\nabla \cdot \xi(\mathbf{x})) = 0, \end{cases}$$

which describe infinitesimal transformations associated to a certain Pfaffian system [38, Example 4]. In order to study the influence of the parametric function a on the system using the package `DifferentialThomas`, a is included in the list of dependent variables and its dependence on merely x_2 is taken into account by adding the following two equations to the system:

$$\frac{\partial}{\partial x_1} a(x_1, x_2, x_3) = 0, \qquad \frac{\partial}{\partial x_3} a(x_1, x_2, x_3) = 0.$$

Let R be the differential polynomial ring $\mathbb{Q}\{\xi_1, \xi_2, \xi_3, a\}$, endowed with the partial differential operators $\partial_1, \partial_2, \partial_3$ with respect to x_1, x_2, x_3.

```
> with(DifferentialThomas):
> ivar := [x1,x2,x3]:
> dvar := [xi1,xi2,xi3,a]:
```

We choose a block ranking $>$ on R with blocks $\{\xi_1, \xi_2, \xi_3\}, \{a\}$.

```
> ComputeRanking(ivar, [[xi1,xi2,xi3],[a]]):
> L := [-a*xi1[x1]+xi3[x1]-a[x2]*xi2
> +(1/2)*a*(xi1[x1]+xi2[x2]+xi3[x3]),
> -a*xi1[x2]+xi3[x2], -a*xi1[x3]+xi3[x3]
> -(1/2)*(xi1[x1]+xi2[x2]+xi3[x3]), a[x1], a[x3]]:
> LL := Diff2JetList(Ind2Diff(L, ivar, dvar));
```

$$\begin{aligned} LL := [&-a_{0,0,0}\,(\xi_1)_{1,0,0} + (\xi_3)_{1,0,0} + \tfrac{1}{2}\,a_{0,0,0}\left((\xi_1)_{1,0,0} + (\xi_2)_{0,1,0} + (\xi_3)_{0,0,1}\right) \\ &-a_{0,1,0}\,(\xi_2)_{0,0,0}, \quad -a_{0,0,0}\,(\xi_1)_{0,1,0} + (\xi_3)_{0,1,0}, \\ &-a_{0,0,0}\,(\xi_1)_{0,0,1} + \tfrac{1}{2}\,(\xi_3)_{0,0,1} - \tfrac{1}{2}\,(\xi_1)_{1,0,0} - \tfrac{1}{2}\,(\xi_2)_{0,1,0}, \quad a_{1,0,0}, \quad a_{0,0,1}] \end{aligned}$$

We compute a Thomas decomposition with respect to $>$ of the given system of partial differential equations.

```
> TD := DifferentialThomasDecomposition(LL, []);
```

$$TD := [\mathit{DifferentialSystem}, \mathit{DifferentialSystem}, \mathit{DifferentialSystem}]$$

The resulting three simple differential systems are given as follows.

```
> Print(TD[1]);
```

$$\begin{aligned} [&a\,\underline{(\xi_1)_{x_2}} - (\xi_3)_{x_2} = 0, \quad a^2\,\underline{(\xi_1)_{x_3}} + (\xi_3)_{x_1} = 0, \quad \underline{\xi_2} = 0, \\ &a\,\underline{(\xi_1)_{x_1}} - 2\,(\xi_3)_{x_1} - a\,(\xi_3)_{x_3} = 0, \quad \underline{a_{x_1}} = 0, \quad \underline{a_{x_3}} = 0, \quad \underline{a} \neq 0] \end{aligned}$$

```
> Print(TD[2]);
```

$$\begin{aligned} [&a\,\underline{(\xi_1)_{x_2}} - (\xi_3)_{x_2} = 0, \quad a^2\,\underline{(\xi_1)_{x_3}} + a\,(\xi_2)_{x_2} - a_{x_2}\,\xi_2 + (\xi_3)_{x_1} = 0, \\ &a\,\underline{(\xi_1)_{x_1}} - a\,(\xi_2)_{x_2} + 2\,a_{x_2}\,\xi_2 - 2\,(\xi_3)_{x_1} - a\,(\xi_3)_{x_3} = 0, \quad \underline{a_{x_1}} = 0, \\ &\underline{a_{x_2,x_2}} = 0, \quad \underline{a_{x_2,x_3}} = 0, \quad \underline{a_{x_3}} = 0, \quad \underline{a} \neq 0, \quad \underline{\xi_2} \neq 0] \end{aligned}$$

```
> Print(TD[3]);
```

$$\begin{aligned} [&\underline{(\xi_1)_{x_1,x_1}} + (\xi_2)_{x_1,x_2} = 0, \quad \underline{(\xi_1)_{x_1,x_2}} + (\xi_2)_{x_2,x_2} = 0, \quad \underline{(\xi_3)_{x_1}} = 0, \quad \underline{(\xi_3)_{x_2}} = 0, \\ &\underline{(\xi_1)_{x_1}} + (\xi_2)_{x_2} - (\xi_3)_{x_3} = 0, \quad \underline{a} = 0] \end{aligned}$$

With regard to the parametric function a, the first simple system is the most generic one, in the sense that $a = a(x_2)$ is only assumed to be non-zero, whereas in the second and third simple systems a is subject to further equations. In particular, the additional condition $a_{x_2,x_2} = 0$ derived in [38] to ensure formal integrability of the system is exhibited in the second simple system of the Thomas decomposition.

4.5 Conclusion

In this paper the Thomas decomposition technique for systems of nonlinear partial differential equations and inequations has been applied to nonlinear control systems. The method splits a given differential system into a finite family of simple differential systems which are formally integrable and define a partition of the solution set of the original differential system. This symbolic approach allows to deal with both differential equations and inequations, which may involve parameters.

Using elimination properties of the Thomas decomposition technique, structural properties of nonlinear control systems have been investigated. In particular, notions such as invertibility, observability and flat outputs can be studied. In the presence of parameters, different simple systems of a Thomas decomposition in general represent different structural behavior of the control system. A Maple implementation of Thomas' algorithm has been used to illustrate the techniques on explicit examples.

At the time of this writing it is unclear how to adapt or generalize the techniques to nonlinear differential time-delay systems or even systems of nonlinear difference equations in full generality. An analog of the notion of Thomas decomposition is not known for systems of nonlinear difference equations. However, note that, generalizing work of J. F. Ritt [40] and R. M. Cohn [8] and others, characteristic set methods have been developed for ordinary difference polynomial systems and differential-difference polynomial systems (cf., e.g., [15, 16]).

Acknowledgements The first author was partially supported by Schwerpunkt SPP 1489 of the Deutsche Forschungsgemeinschaft. The authors would like to thank an anonymous referee for several useful remarks. They would also like to thank S. L. Rueda for pointing out reference [34].

References

1. Aubry, P., Lazard, D., Moreno Maza, M.: On the theories of triangular sets. J. Symb. Comput. **28**(1–2), 105–124 (1999)
2. Avanessoff, D., Pomet, J.-B.: Flatness and Monge parameterization of two-input systems, control-affine with 4 states or general with 3 states. ESAIM Control Optim. Calc. Var. **13**(2), 237–264 (2007)
3. Bächler, T.: Counting solutions of algebraic systems via triangular decomposition. PhD thesis, RWTH Aachen University, Germany (2014). http://publications.rwth-aachen.de/record/444946?ln=en
4. Bächler, T., Gerdt, V.P., Lange-Hegermann, M., Robertz, D.: Algorithmic Thomas decomposition of algebraic and differential systems. J. Symb. Comput. **47**(10), 1233–1266 (2012)
5. Bächler, T., Lange-Hegermann, M.: Algebraic Thomas and Differential Thomas: Thomas decomposition of algebraic and differential systems. http://wwwb.math.rwth-aachen.de/thomasdecomposition
6. Blinkov, Y.A., Cid, C.F., Gerdt, V.P., Plesken, W., Robertz, D.: The MAPLE package "Janet": I. Polynomial systems. II. Linear partial differential equations. In: Ganzha, V.G., Mayr, E.W., Vorozhtsov, E.V. (eds.) Proceedings of the 6th International Workshop on Computer Algebra in Scientific Computing, Passau, Germany, pp. 31–40 resp. pp. 41–54 (2003). http://wwwb.math.rwth-aachen.de/Janet

7. Boulier, F., Lazard, D., Ollivier, F., Petitot, M.: Computing representations for radicals of finitely generated differential ideals. Appl. Algebra Eng. Commun. Comput. **20**(1), 73–121 (2009)
8. Cohn, R.M.: Difference Algebra. Wiley-Interscience, New York (1965)
9. Conte, G., Moog, C.H., Perdon, A.M.: Nonlinear Control Systems. Lecture Notes in Control and Information Sciences, vol. 242. Springer, London (1999)
10. Diop, S.: Differential-algebraic decision methods and some applications to system theory. Theor. Comput. Sci. **98**(1), 137–161 (1992)
11. Diop, S.: Elimination in control theory. Math. Control. Signals Syst. **4**(1), 17–32 (1991)
12. Eisenbud, D.: Commutative Algebra – with a View Toward Algebraic Geometry. Graduate Texts in Mathematics, vol. 150. Springer, New York (1995)
13. Fliess, M., Glad, S.T.: An algebraic approach to linear and nonlinear control. In: Trentelman, H.L., Willems, J.C. (eds.) Essays on Control: Perspectives in the Theory and its Applications, pp. 223–267. Birkhäuser, Boston (1993)
14. Fliess, M., Lévine, J., Martin, P., Rouchon, P.: Flatness and defect of non-linear systems: introductory theory and examples. Int. J. Control. **61**(6), 1327–1361 (1995)
15. Gao, X.-S., Luo, Y., Yuan, C.M.: A characteristic set method for ordinary difference polynomial systems. J. Symb. Comput. **44**(3), 242–260 (2009)
16. Gao, X.-S., van der Hoeven, J., Yuan, C.M., Zhang, G.L.: Characteristic set method for differential-difference polynomial systems. J. Symb. Comput. **44**(9), 1137–1163 (2009)
17. Gerdt, V.P.: On decomposition of algebraic PDE systems into simple subsystems. Acta Appl. Math. **101**(1–3), 39–51 (2008)
18. Gerdt, V.P., Blinkov, Y.A.: Involutive bases of polynomial ideals. Math. Comput. Simul. **45**, 519–541 (1998)
19. Gerdt, V.P., Lange-Hegermann, M., Robertz, D.: The Maple package TDDS for computing Thomas decompositions of systems of nonlinear PDEs. Comput. Phys. Commun. 234, 202–215 (2019)
20. Glad., S.T.: Differential algebraic modelling of nonlinear systems. In: Kaashoek, M.A., van Schuppen, J.H., Ran, A.C.M. (eds.) Realization and Modelling in System Theory, pp. 97–105. Birkhäuser, Boston (1989)
21. Hubert, E.: Notes on triangular sets and triangulation-decomposition algorithms. I. Polynomial systems. II. Differential systems. In: Winkler, F., Langer, U. (eds.) Symbolic and Numerical Scientific Computation, Hagenberg (2001), pp. 1–39 resp. 40–87. Lecture Notes in Computer Science, vol. 2630. Springer, Berlin (2003)
22. Ince, E.L.: Ordinary Differential Equations. Dover Publications, New York (1956)
23. Janet, M.: Leçons sur les systèmes d'équations aux dérivées partielles. Cahiers Scientifiques IV. Gauthiers-Villars, Paris (1929)
24. Kolchin, E.R.: Differential Algebra and Algebraic Groups. Pure and Applied Mathematics, vol. 54. Academic, New York (1973)
25. Kwakernaak, H., Sivan, R.: Linear Optimal Control Systems. Wiley-Interscience, New York (1972)
26. Lange-Hegermann, M.: Counting solutions of differential equations. PhD thesis, RWTH Aachen University, Germany (2014). http://publications.rwth-aachen.de/record/229056?ln=en
27. Lange-Hegermann, M.: The differential counting polynomial. Found. Comput. Math. **18**(2), 291–308 (2018)
28. Lange-Hegermann, M., Robertz, D.: Thomas decompositions of parametric nonlinear control systems. In: Proceedings of the 5th Symposium on System Structure and Control, Grenoble, France, pp. 291–296 (2013)
29. Lemaire, F., Moreno Maza, M., Xie, Y.: The RegularChains library in MAPLE. SIGSAM Bull. **39**, 96–97 (2005). September
30. Levandovskyy, V., Zerz, E.: Obstructions to genericity in study of parametric problems in control theory. In: Park, H., Regensburger, G. (eds.) Gröbner Bases in Control Theory and Signal Processing. Radon Series on Computational and Applied Mathematics, vol. 3, pp. 127–149. Walter de Gruyter, Berlin (2007)

31. Lévine, J.: On necessary and sufficient conditions for differential flatness. Appl. Algebra Eng. Commun. Comput. **22**(1), 47–90 (2011)
32. Mishra, B.: Algorithmic Algebra. Texts and Monographs in Computer Science. Springer, New York (1993)
33. Nijmeijer, H., van der Schaft, A.: Nonlinear Dynamical Control Systems. Springer, New York (1990)
34. Picó-Marco, E.: Differential algebra for control systems design: constructive computation of canonical forms. IEEE Control Syst. Mag. **33**(2), 52–62 (2013)
35. Plesken, W.: Counting solutions of polynomial systems via iterated fibrations. Arch. Math. (Basel) **92**(1), 44–56 (2009)
36. Pommaret, J.-F.: Partial Differential Equations and Group Theory. Mathematics and Its Applications, vol. 293. Kluwer Academic Publishers Group, Dordrecht (1994)
37. Pommaret, J.-F.: Partial Differential Control Theory. Mathematics and Its Applications, vol. 530. Kluwer Academic Publishers Group, Dordrecht (2001)
38. Pommaret, J.-F., Quadrat, A.: Formal obstructions to the controllability of partial differential control systems. In: Proceedings of IMACS, Berlin, Germany, vol. 5, pp. 209–214 (1997)
39. Riquier, C.: Les systèmes d'équations aux dérivées partielles. Gauthiers-Villars, Paris (1910)
40. Ritt, J.F.: Differential Algebra. American Mathematical Society Colloquium Publications, vol. XXXIII. American Mathematical Society, New York (1950)
41. Robertz, D.: Formal Algorithmic Elimination for PDEs. Lecture Notes in Mathematics, vol. 2121. Springer, Cham (2014)
42. Robertz, D.: Recent progress in an algebraic analysis approach to linear systems. Multidimens. Syst. Signal Process. **26**(2), 349–388 (2015)
43. Seidenberg, A.: An elimination theory for differential algebra. Univ. California Publ. Math. (N.S.) **3**, 31–65 (1956)
44. Thomas, J. M.: Differential Systems. American Mathematical Society Colloquium Publications, vol. XXI. American Mathematical Society, New York (1937)
45. Wang, D.: Decomposing polynomial systems into simple systems. J. Symb. Comput. **25**(3), 295–314 (1998)
46. Wang, D.: Elimination Methods. Texts and Monographs in Symbolic Computation. Springer, Vienna (2001)
47. Wu, W.T.: Mathematics Mechanization. Mathematics and Its Applications, vol. 489. Kluwer Academic Publishers Group, Dordrecht; Science Press, Beijing (2000)

Chapter 5
Some Control Observation Problems and Their Differential Algebraic Partial Solutions

Sette Diop

Abstract Observation problems in control systems literature generally refer to problems of estimation of state variables (or identification of model parameters) from two sources of information: dynamic models of systems consisting in first order differential equations relating all system quantities, and online measurements of some of these quantities. For nonlinear systems the classical approach stems from the work of R. E. Kalman on the distinguishability of state space points given the knowledge of time histories of the output and input. In the differential algebraic approach observability is rather viewed as the ability to recover trajectories. This approach turns out to be a particularly suitable language to describe observability and related questions as structural properties of control systems. The present paper is an update on the latter approach initiated in the late eighties and early nineties by J. F. Pommaret, M. Fliess, S. T. Glad and the author.

Keywords Control observation problems · State estimation · Differential algebraic decision methods · Differential algebraic geometry

5.1 Introduction

Observation problems in control systems literature generally refer to problems of estimation of state variables x from two sources of information: online measurements of external variables u and y, and first order dynamic models

$$\begin{cases} \dot{x} = f(t, u, x)\,, \\ y = h(t, u, x)\,, \end{cases} \tag{5.1}$$

relating x to u and y. See for instance [1, 2].

S. Diop (✉)
Laboratoire des Signaux & Systmes, CNRS-CentraleSuplec-Univ Paris-Sud, Universit Paris-Saclay, Plateau de Moulon 3 rue Joliot Curie, 91192 Gif sur Yvette cedex, France
e-mail: Diop@L2S.CentraleSupelec.fr

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_5

By using tools from differential algebraic geometry

- dynamic models are allowed to be implicit and of arbitrary order but restricted to be polynomial in variables and their derivatives,
- and may be considered more general situations of estimating one subset of system variables with respect to another subset of the system variables.

Specifically, given a dynamic system described by algebraic differential equations

$$\begin{cases} P_i(w, z, \zeta) = 0 \quad (i = 1, 2, \ldots), \\ Q(w, z, \zeta) \neq 0, \end{cases} \tag{5.2}$$

one observation problem consists of the *online* estimation of $z(t) \in \mathbb{R}^{\nu}$ from the knowledge of the P_i's and Q and time histories $([t_0, t] \ni \tau \longmapsto w(\tau) \in \mathbb{R}^{\mu})$ of w. Here the P_i's and Q are differential polynomials in w, z and ζ, and ν, μ are natural integers. This problem is under investigation since the pioneering work of R. E. Kalman in the late fifties addressing its linear context. A complete nonlinear answer is still lacking. A general approach consists of a two part theory: one of *observability*, that is, derivation of conditions on the P_i's and Q guaranteeing the ability to some how estimate z from the supposedly known data, and the other part of the theory, the *observer design*, searches algorithms for such an estimation of z.

Though central the previous observation problem (observability and observer design) is not the only one. For instance, closely related to it, are two problems of *robustness* with respect to model and measurements uncertainties. Another observation problem with important practical application consists of determining subsets w of systems variables which make a given subset z observable.

Starting from the mid eighties (see [3–6]) differential algebra and differential algebraic decision methods have been shown to provide a quite consistent language to describe some of these observation problems along with some of their solutions.

An account of this is proposed here. Some of the many open problems will be described.

Reviewers of the present paper suggested appending to it materials of differential algebra which are invoked throughout. Such an account has been done already in [6] and would double the space of the present paper. These are the reasons why we prefer referring the reader to the appendix of [6] instead of duplicating here those materials of differential algebra, differential algebraic geometry and differential algebraic decision methods.

5.2 The Differential Algebraic Approach

A thorough introduction to the differential algebraic approach is available in [6]. For the sake of completeness the following definition is recalled from there.

A (*differential*) (*algebraic*) *system* $\mathcal{X}$ with s variables, and with coefficients in a differential field $\mathbf{k}$ is a *proper* differential quasi-affine variety $\mathcal{X} \subseteq \overline{\mathbf{k}}^{s}$ defined

over $\mathbf{k}$ where $\overline{\mathbf{k}}$ is a differential closure of $\mathbf{k}$. In observation problems, the system variable is partitioned into the *data*, or *observations*, $w = w_1, \ldots, w_\mu$, the variable being observed (or estimated) $z = z_1, \ldots, z_n$ and the remaining variables, ζ. In the classical observation problem, the data consist exclusively of (u, y), the control u and the measurements y. When the variable ζ is present, the projection $\mathcal{X}_{w,z}$ of $\mathcal{X}$ *along the variable* ζ is considered. It is the set of elements $(\overline{w}, \overline{z}) \in \overline{\mathbf{k}}^{\mu} \times \overline{\mathbf{k}}^{n}$ such that there is at least $\overline{\zeta}$ such that $(\overline{w}, \overline{z}, \overline{\zeta}) \in \mathcal{X}$.

In terms of equations, previously defined systems are those described by

$$\begin{cases} P_i(w, z, \zeta) = 0\,, & i = 1, 2, \ldots, \\ Q(w, z, \zeta) \neq 0\,, \end{cases} \tag{5.3}$$

where the P_i's and Q are finitely many polynomials in w, z, ζ and their derivatives.

For a system $\mathcal{X}$ the variable z is said to be *(algebraically) observable with respect to* w if the projection map $\pi : \mathcal{X}_{w,z} \to \mathcal{X}_w$ (sending every trajectory $(\overline{w}, \overline{z})$ of $\mathcal{X}_{w,z}$ onto the corresponding observation $\overline{w}$) is *generically finite*. If z is observable with respect to w then the degree of π is called the *observability degree* of z with respect to w, and is denoted by $\mathrm{d}^{\circ}_{w} z$. The variable z is said to be *rationally* observable with respect to w if it is observable with respect to w with observability degree one. State systems of the form (5.1) are said to be *observable* if x is observable with respect to (u, y).

It was first proved in [5] (see [6] for more details) that the previous definition has a differential algebraic translation, namely: z is observable with respect to w iff z is algebraic over $\mathbf{k}\langle w \rangle$, that is, for each component, z_i of z there is a polynomial equation

$$H_i\,(z_i, w, \dot{w}, \ldots) = 0 \tag{5.4}$$

in z_i, and finitely many time derivatives of the data w, with coefficients in $\mathbf{k}$.

The reader is referred to [6] for more details on differential algebraic geometry terms or notations used here without explanations.

5.3 How Does It Compare to the Classical Theory?

Formal definitions of observability can be found in [1, 2] for instance.

For linear state systems

$$\begin{cases} \dot{x} = F\,x + G\,u\,, \\ y = H\,x + E\,u\,, \end{cases} \tag{5.5}$$

the answer to the question is that algebraical observability of x with respect to (u, y) is *equivalent* to the classical Kalman definition of observability of system (5.5). The proof of this is as follows.

As is well known system (5.5) is observable in the classical sense iff $\mathrm{rk}_{\mathbb{R}} \mathcal{O}(F, H) = n$, where n is the number of components of the state, x, and where

$$\mathcal{O}(F, H) = \begin{pmatrix} H \\ H\,F \\ \vdots \\ H\,F^{n-1} \end{pmatrix}.$$

Now the following equalities

$$\begin{aligned} H\,x &= y - E\,u &&= z_0\,, \\ H\,F\,x &= \dot{z}_0 - H\,G\,u &&= z_1\,, \\ H\,F^2\,x &= \dot{z}_1 - H\,F\,G\,u &&= z_2\,, \\ &\vdots \\ H\,F^{n-1}\,x &= \dot{z}_{n-2} - H\,F^{n-2}\,G\,u &&= z_{n-1}\,, \end{aligned} \tag{5.6}$$

result from the equations of system (5.5). The reader will notice that they are written such that *only* supposedly differentiable quantities are differentiated. They may be rewritten as

$$\mathcal{O}(F, H)\;x = \begin{pmatrix} z_0 \\ z_1 \\ z_2 \\ \vdots \\ z_{n-1} \end{pmatrix}. \tag{5.7}$$

Therefore if system (5.5) is observable in the classical sense then it is so in the algebraic sense. The converse follows from Corollary 1 below.

The equivalence between the algebraic and the classical definitions of observability for the class of systems (5.5) was first mentioned in Sect. 5.1.2 of [7] but without a complete proof.

It is worth emphasizing the fact that the algebraic definition of systems applies without any change to so-called *implicit* or *descriptor* linear systems

$$\begin{cases} M\,\dot{x} = F\,x + G\,u\,, \\ \quad y = H\,x + E\,u\,, \end{cases}$$

where the matrix M is *singular*. Compare to [8] and references therein.

The largest class of systems where algebraic and classical observability may be compared is the class of rational state systems

$$\begin{cases} \dot{x}_i = \dfrac{p_i(u, x)}{q_i(u, x)}\;(1 \le i \le n)\,, \\ y_j = \dfrac{f_j(u, x)}{g_i(u, x)}\;(1 \le j \le p)\,, \end{cases} \tag{5.8}$$

where u stands for $u_1, u_2, \ldots, u_m$, and p_i, q_i, f_i and g_i are differential polynomials of *order zero* in x with coefficients in $\mathbf{k} = \mathbb{R}$, the algebraic observability (of x with respect to (u, y)) is equivalent to the *generic local observability* of the system as defined in [1]. This was first obtained in [9].

5.4 Partial Answers to Some Observation Problems

Among all benefits of the differential algebraic approach to observation problems, application of decision methods is perhaps the most appealing.

5.4.1 Computing

One of the most used estimation algorithms, the Kalman filter (or its extended form), is often applied in practice without prior asserting the validity of its conditions. The reason of this is that there is no systematic method for such a verification.

The differential algebraic theory of observability is constructive *in principle*.

Most of the decision methods which may be used are already described in [6]. This is the case for general polynomial systems

$$P_i(w, z, \zeta) = 0\,, \quad i = 1, 2, \ldots \tag{5.9}$$

with *constant* coefficients. For rational state systems (5.8) the observability test is formally similar to the Jacobian rank condition which is found in [1].

For polynomial state systems with *nonconstant* coefficients let $\mathbf{k}$ be an *ordinary* differential field (*not necessarily of constants*). Let $\mathcal{X}$ be

$$\begin{cases} \dot{x}_i = f_i(u, x) \ \ (1 \le i \le n)\,, \\ y_j = h_j(u, x) \ \ (1 \le j \le p)\,, \end{cases} \tag{5.10}$$

where the f_i's and h_j's are (nondifferential) *polynomials* in their arguments with coefficients in $\mathbf{k}$. Let

$$\begin{aligned} P_i(U, X, Y) &= X_i^{(1)} - f_i(U, X) \quad (1 \le i \le n); \\ P_{n+j}(U, X, Y) &= h_j(U, X) - Y_j \quad (1 \le j \le p) \end{aligned}$$

be the differential polynomials defining $\mathcal{X}$. Let $\sigma : \mathbf{k}\{U, Y\} \to \mathbf{k}\{u, y\}$ be the substitution map which sends U to u and Y to y, where $\mathbf{k}\{u, y\}$ is the differential $\mathbf{k}$-subalgebra of $\mathbf{k}\{u, x, y\}$ generated over $\mathbf{k}$ by u and y. Let P be in $\mathbf{k}\{U, X, Y\}$ and P^σ denote the element of $\mathbf{k}\{u, y\}\{X\}$ obtained by regarding P as a differential polynomial in X with coefficients in $\mathbf{k}\{U, Y\}$ and by applying σ to each of these

coefficients, and let $\mathbf{I}(\mathcal{X})^\sigma$ stand for the differential ideal of $\mathbf{k}\{u, y\}\{X\}$ consisting of P^σ ($P \in \mathbf{I}(\mathcal{X})$). The ideal of definition, $\mathfrak{a}$, of $\mathbf{k}\langle u, y\rangle(x)$ over $\mathbf{k}\langle u, y\rangle$ is equal to $\mathbf{I}(\mathcal{X})^\sigma \bigcap \mathbf{k}\langle u, y\rangle\big[(X_i)_{1\leq i\leq n}\big]$. Note that the set $\mathcal{A}$ consisting of the P_i ($1 \leq i \leq n$) form an *autoreduced set* with respect to any *ranking* of $\mathbf{k}\{U, X, Y\}$ such that U, Y and their derivatives all are lower than X.

Let us now inductively define some polynomials which will turn out to be generators of the ideal $\mathbf{I}(\mathcal{X})^\sigma \bigcap \mathbf{k}\langle u, y\rangle\big[(X_i)_{1\leq i\leq n}\big]$ of $\mathbf{k}\langle u, y\rangle\big[(X_i)_{1\leq i\leq n}\big]$.

Starting with

$$Q_i(U, X, Y) = P_{n+i}(U, X, Y) \quad (1 \leq i \leq p),$$

then let Q_{p+i} be the *remainder* of the derivative of Q_i ($1 \leq j \leq n$) with respect to the previously mentioned autoreduced set, $\mathcal{A}$. The polynomial Q_{p+i} is merely the derivative of Q_i in which $X_j^{(1)}$ is eliminated by substituting $P_j + f_j$ for $X_j^{(1)}$ ($1 \leq j \leq n$) (The linear combination of P_j ($1 \leq j \leq n$) which appears reduces to zero when the remainder is taken, so that it can be ignored.) Explicitly, Q_{p+i} is as follows

$$Q_{p+i} = \sum_{1\leq j\leq m} \frac{\partial Q_i}{\partial U_j} U_j^{(1)} + \sum_{1\leq j\leq n} \frac{\partial Q_i}{\partial X_j} f_j - Y_i^{(1)} + Q_{i\bullet} \quad (1 \leq i \leq p),$$

where the notations

$$P_\bullet \equiv P_{(1)}, \quad P_{(2)}, \quad \ldots \tag{5.11}$$

for a differential polynomial P stand for the differential polynomials obtained by replacing the coefficients of P by their respective derivatives respectively at order 1, 2, etc.

Note that this formula is nothing but a counterpart of Lie derivatives: Authors usually consider the functions h_j as free of u and the functions f_i and h_j as with *constant* coefficients so that in the left hand side of the latter equation the first sum as well as the last term are absent.

This construction of Q_{p+i} from Q_i is iterated in order to get Q_{2p+i} ($1 \leq i \leq p$) as the remainder of the derivative of Q_{p+i} ($1 \leq i \leq p$). And so on.

By their definition,

$$Q_i^\sigma \in \mathbf{I}(\mathcal{X})^\sigma \bigcap \mathbf{k}\langle u, y\rangle\Big[\big(X_j\big)_{1\leq j\leq n}\Big] \ (i \in \mathbb{N}).$$

Conversely, let P

$$P \in \mathbf{I}(\mathcal{X})^\sigma \bigcap \mathbf{k}\langle u, y\rangle\Big[\big(X_j\big)_{1\leq i\leq n}\Big].$$

As an element of $\mathbf{I}(\mathcal{X})$, P may easily be written in the form

$$P = \sum_{1 \le i \le n,\ j \in \mathbb{N}} A_{i,j}\, P_i^{(j)} + \sum_{i \in \mathbb{N}} B_i\, Q_i.$$

where B_i $(i \in \mathbb{N})$ are in

$$\mathbf{I}(\mathcal{X}) \bigcap \mathbf{k}\{U, Y\}\left[(X_i)_{1 \le i \le n}\right].$$

Since the differential ideal of $\mathbf{k}\{U, X, Y\}$ generated by P_i $(1 \le i \le n)$ has no nonzero element in common with $\mathbf{I}(\mathcal{X}) \bigcap \mathbf{k}\{U, Y\}\left[\left(X_j\right)_{1 \le i \le n}\right]$ (this results from an obvious degree argument), the first sum in the previous equality must be zero.

This ends the proof that Q_i^{σ} $(i \in \mathbb{N})$ form a basis of $\mathfrak{a}$.

Lemma 1 *A set of generators of the ideal of definition of* $\mathbf{k}\langle u, y\rangle(x)$ *over* $\mathbf{k}\langle u, y\rangle$ *is given by*

$$\begin{array}{l} Q_1(u, X, y), Q_2(u, X, y), \ldots, Q_p(u, X, y), \\ Q_{p+1}(u, X, y), Q_{p+2}(u, X, y), \ldots, Q_{2p}(u, X, y), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{array}$$

In addition, it comes from the Hilbert basis theorem that only finitely many Q_i *suffice to generate the ideal* $\mathfrak{a}$. *That is, there is some* μ *in* $\mathbb{N}$ *such that the first* μ *rows of the previous list of* $Q_i(u, X, y)$ *generate the ideal of definition of* $\mathbf{k}\langle u, y\rangle(x)$ *over* $\mathbf{k}\langle u, y\rangle$. *According to Theorem 16 of [6], the observability of* $\mathcal{X}$ *is equivalent to the fact that the* $\mathbf{k}\langle u, y\rangle(x)$*-matrix*

$$\left[\frac{\partial Q_i}{\partial X_j}(u, x, y)\right]_{\substack{1 \le i \le \mu\, p \\ 1 \le j \le n}}$$

is of rank n.

Now it is a basic fact that the above rank is equal to the rank of the first $n\ p$ rows.

Corollary 1 *If* $\mathcal{X}$ *possesses a state description as above, then* $\mathcal{X}$ *is observable if, and only if, the following* $\mathbf{k}\langle u, y\rangle(x)$*-matrix*

$$\left[\frac{\partial Q_i}{\partial X_j}(u, y, x)\right]_{\substack{1 \le i \le n\, p \\ 1 \le j \le n}}$$

(which is formally the counterpart of the matrix of Lie derivatives which appears in the Hermann–Krener observability Jacobian rank condition) is of rank n.

The main difference between this rank condition and the one in [1] is that the rank is not over $\mathbf{k}$ (which is usually $\mathbb{R}$) but over a much bigger field (and, here the rank condition is a necessary and sufficient condition).

For arbitrary systems, observability tests resort on decision methods such as characteristic set of the defining differential ideal of $\mathcal{X}$. See [6] for more details. Very promising, Thomas decomposition was also proposed as decision methods for the same tests, see [10, 11].

5.5 Regular Observability

The notion of regular observability refers to the classical one of *universal inputs* as thoroughly treated in [12]. Bad inputs (as opposed to universal ones) are supposed to occlude the functioning of online estimation schemes when they happened to be applied to a system. The present differential algebraic approach has brought a new light to this notion of singularity of the observability property. Here is an abstract of the result which may be found in [6] in more details.

Let $\mathcal{X}$ be a system with variables w, z, and ζ, and with coefficients in $\mathbf{k}$. It is a matter of fact that, when z is observable with respect to w then for special observations $\overline{w}$, $\pi^{-1}(\overline{w})$ may contain infinitely many elements, leading to a singularity of the generic notion of observability. Here π is the projection map of Sect. 5.2. An example of such situations is the following

$$\begin{cases} \dot{x}_1 = x_1 x_2 \,, \\ \dot{x}_2 = u + x_2 \,, \\ y = x_1 \,. \end{cases} \tag{5.12}$$

x is observable with respect to u, y since

$$x_1 = y \quad \text{and} \quad x_2 = \frac{\dot{y}}{y} \,.$$

But in practice, in any time interval where y is identically zero (or, merely, small), the observability of x_2 is singular in the sense that it is lost.

An observation $\overline{w} \in \mathcal{X}_w$ is said to be *singular* for the observation of z with respect to w if $\pi^{-1}(\overline{w})$ is *infinite*. Observations $\overline{w} \in \mathcal{X}_w$ which are not singular are called *regular*. The variable z is said to be *regularly observable* with respect to w if there is no singular observation for its observability with respect to w.

The best result obtained in this approach reads as follows.

Theorem 1 *Let $\mathcal{X}$ be a system with variables w, z, and ζ, and with coefficients in* $\mathbf{k}$. *The variable z is regularly observable with respect to w if z is* primitive *over* $\mathbf{k}\{w\}$.

Recall that an element ξ of $\mathbf{k}\{w, z\}$ is said to be *primitive* over $\mathbf{k}\{w\}$ if it is a zero of a polynomial

$$a_d\, \xi^d + a_{d-1}\, \xi^{d-1} + \cdots + a_0 = 0$$

such that

(a) the a_i's are in $\mathbf{k}\{w\}$,
(b) the perfect differential ideal $\{a_d(w), a_{d-1}(w), \ldots, a_0(w)\}$ of $\mathbf{k}\{w\}$ is the unit ideal.

5.5.1 Sensor Selection

Given a dynamic system with differential field extension

$$\mathbf{k}\langle u, z\rangle$$

with input u and latent variable z the sensor selection problem consists of the selection of sensors which endow the system with some properties. Among all such desirable properties is the basic one of observability. In this section partial answers to the following questions will be provided.

(a) What is the minimal number of sensors that make the dynamics observable?
(b) When the sensors are bound to measure state components, what is their minimum number?
(c) How may the *observability margin* be improved by selecting the sensors?

Let y denote an arbitrary output of the system. By definition, an output, y, is componentwise algebraic over $\mathbf{k}\langle u, z\rangle$. An output makes the dynamics observable if

$$\mathrm{d}^{\circ}_{\mathbf{k}\langle u,z,y\rangle}\mathbf{k}\langle u, y\rangle = 0\,,$$

that is, if each component of z is algebraic over $\mathbf{k}\langle u, y\rangle$. Clearly, $y = z$ is an output which makes the system observable. Let n denote the number of components of z, and

$$\mathbf{I} = \left\{p \in \mathbb{N} : \exists y_1, y_2, \ldots, y_p \in \mathbf{k}\,(u, z)\,,\ \mathrm{d}^{\circ}_{\mathbf{k}\langle u,z,y\rangle}\mathbf{k}\langle u, y\rangle = 0\right\}$$

The set $\mathbf{I}$ is the one of integers p such that there exists an output y with p components which makes the system observable.

It is a nonempty (since $n \in \mathbf{I}$) subset of $\mathbb{N}$. Therefore, $\mathbf{I}$ contains a smallest element which is precisely the minimal number of sensors which make the system observable.

The sensor selection problem characterizing this minimum number, p, of sensors is an open problem.

Later in this section it is shown that the minimum number of sensors is 1 for rational state systems, providing a partial answer to Question 1 above.

Next, about Question 2, what if the output y is chosen as a subset of the components of z, instead of vector rational function of u and z? A complete but trivial, inelegant, and computationally costly answer consists of performing the $2^n - 1$ observability tests!

Up to the knowledge of the author there is no partial contribution to Question 3.

Back to Question 1, here is the surprising answer for the class of rational state systems.

Theorem 2 *Let the state of a system $\mathcal{X}$ be given by*

$$\dot{x} = f(u, x) \tag{5.13}$$

with a vector rational function f of input u, state x, and with coefficients in a differential field $\mathbf{k}$. *Let m and n be the respective numbers of components of u and x. Let* $\mathbf{K}$ *be a differential extension field of* $\mathbf{k}$. *If* $\mathbf{K}$ *contains nonconstants then there always is a scalar output*

$$y = \sum_{i=1}^{n} \alpha_i \, x_i \tag{5.14}$$

with $\alpha_1, \alpha_2, \ldots, \alpha_n$ in $\mathbf{K}$, *which makes x observable with respect to (u, y). Moreover, for y as in (5.14) to make $\mathcal{X}$ observable it is sufficient that the associated α's be linearly independent over the subfield of constants of* $\mathbf{K}$.

Proof Let $y^{[n]}$ denote the vector

$$y^{[n]} = \begin{pmatrix} y \\ \dot{y} \\ \vdots \\ y^{(n-1)} \end{pmatrix}.$$

By Corollary 1, for the output (5.14) to make x observable with respect to (u, y) it is necessary, and sufficient, that the Jacobian matrix of $y^{[n]}$ with respect to x

$$\frac{\partial y^{[n]}}{\partial x'} = \begin{pmatrix} \dfrac{\partial y}{\partial x'} \\ \dfrac{\partial \dot{y}}{\partial x'} \\ \vdots \\ \dfrac{\partial y^{(n-1)}}{\partial x'} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial y}{\partial x_1} & \dfrac{\partial y}{\partial x_2} & \cdots & \dfrac{\partial y}{\partial x_n} \\ \dfrac{\partial \dot{y}}{\partial x_1} & \dfrac{\partial \dot{y}}{\partial x_2} & \cdots & \dfrac{\partial \dot{y}}{\partial x_n} \\ \vdots & & & \\ \dfrac{\partial y^{(n-1)}}{\partial x_1} & \dfrac{\partial y^{(n-1)}}{\partial x_2} & \cdots & \dfrac{\partial y^{(n-1)}}{\partial x_n} \end{pmatrix}$$

be of rank n over $\mathbf{K}\langle u, y\rangle(x)$ where the complete system is considered as with coefficients in $\mathbf{K} \supseteq \mathbf{k}$.

Now y may be written as

$$y = \alpha' x$$

where

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} .$$

The first row of $\partial y^{[n]}/\partial x'$ is merely

$$\frac{\partial y}{\partial x'} = \alpha' .$$

The second row of the same matrix is

$$\frac{\partial \dot{y}}{\partial x'} = \dot{\alpha}' + \alpha' \frac{\partial f(u, x)}{\partial x'} .$$

Therefore,

$$\mathrm{rk}_{\mathbf{k}\langle u, y\rangle(x)} \frac{\partial y^{[n]}}{\partial x'} = \mathrm{rk}_{\mathbf{k}\langle u, y\rangle(x)} \begin{pmatrix} \alpha' \\ \dot{\alpha}' \\ \dfrac{\partial \ddot{y}}{\partial x'} \\ \vdots \\ \dfrac{\partial y^{(n-1)}}{\partial x'} \end{pmatrix}$$

by substituting the linear combination

$$\frac{\partial \dot{y}}{\partial x'} - \frac{\partial y}{\partial x'} \frac{\partial f(u, x)}{\partial x'} = \dot{\alpha}'$$

over $\mathbf{k}\langle u, y\rangle(x)$ of the first two rows for the second row of $\partial y^{[n]}/\partial x'$.

More generally, by the Leibniz formula,

$$y^{(i)} = \alpha^{(i)'} x + \sum_{j=1}^{i} \binom{i}{j} \alpha^{(i-j)'} x^{(j)} ,$$

and by the fact that $x^{(j)}$ is in $\mathbf{k}\langle u\rangle(x)$ for all $j \geq 1$, it is clear that

$$\mathrm{rk}_{\mathbf{k}\langle u,y\rangle(x)}\frac{\partial y^{[n]}}{\partial x'}=\mathrm{rk}_{\mathbf{k}\langle u,y\rangle(x)}\begin{pmatrix}\alpha'\\ \dot{\alpha}'\\ \ddot{\alpha}'\\ \vdots\\ \alpha^{(n-1)}\end{pmatrix}$$

by an immediate induction on the row's index of the Jacobian matrix.

Now note that the matrix

$$\mathcal{W}(\alpha)=\begin{pmatrix}\alpha'\\ \dot{\alpha}'\\ \vdots\\ \alpha^{(n-1)}\end{pmatrix}$$

is *square*, and of order n, and does not involve neither u nor x. Therefore it is of rank n over $\mathbf{K}\langle u, y\rangle(x)$ if, and only, its determinant is nonzero.

Next note that the determinant of $\mathcal{W}(\alpha)$ is simply a differential polynomial in $\alpha_1, \alpha_2, \ldots, \alpha_n$ with coefficients in the field of constants of $\mathbf{K}$. Then the following theorem is used.

Theorem 3 *If G is a nonzero differential polynomial in n indeterminates with coefficients in a differential field containing nonconstant elements then G possesses a zero $(z_1, \ldots, z_n)$ over $\mathbf{K}$.*

For a proof see [13–16] for instance.

This terminates the proof of the first assertion in the theorem.

The second assertion follows from the following. The n elements $\alpha_1, \alpha_2, \ldots, \alpha_n$ of $\mathbf{K}$ are said to be linear dependent over constants if there is a nontrivial relation

$$c_1\,\alpha_1 + c_2\,\alpha_2 + \cdots + c_n\,\alpha_n = 0$$

with *constant* coefficients. It is a classical result that

Theorem 4 *$\alpha_1, \alpha_2, \ldots, \alpha_n$ are linearly dependent over constants if, and only if, the Wronskian matrix $\mathcal{W}(\alpha)$ is singular.*

For a proof see the same references [13–16] for instance. This ends the proof of the theorem.

5.6 Some of the Questions Without Partial Answers

The following is no way an exhaustive list of open problems. It is simply believed that the reader may be inspired to contribute to their solution. One of the most challenging open question is actually of a foundation level.

5.6.1 A Foundation Problem

Observation problems are basically encountered in engineering practices where *real problems* often refer to *real valued* parameters and variables. But as the reader has already noticed the present differential algebraic geometry approach has recourse to so-called *differential closures* of ground fields. Basically such fields are complex ones. For more details the reader may refer to Sect. 5.1 of [6].

5.6.2 Robustness

Keeping in mind that engineering observation problems deal with systems which may be inaccurately known. The most favorable lack of information is actually that of true values of parameters or coefficients: in other words, the form and orders and dimensions of the equations are exactly known, only actual parameter values are uncertain.

The question is then how observability and other observation problems assertions behave in the presence of parameter uncertainties?

For linear systems (5.5) a notion of *observability margin* may be defined characterizing the *distance* of a given system to unobservable ones. This generally uses matrix tools such as singular values.

For general systems the question is clearly related to the so-called field of *decisions methods for approximate systems* as tacked in [17, 18] and many other papers.

5.6.3 Decision Methods Problems

Among decision methods capable of dealing with real examples such as those one may find in biotechnology are also wanted. In this vein, there is a question with practical importance: Given an observable variable z with respect to w, what is the minimal order of derivatives of w involved in the observability of z with respect to w?

References

1. Hermann, R., Krener, A.J.: Nonlinear controllability and observability. IEEE Trans. Automat. Control **22**, 728–740 (1977)
2. Gauthier, J.P., Kupka, I.A.K.: Deterministic Observation Theory and Applications. Cambridge University Press, Cambridge, (2001)
3. Pommaret, J.F.: Géométrie différentielle algébrique et théorie du contrôle. C. R. Acad. Sci. Paris Sér. **I**(302), 547–550 (1986)

4. Fliess, M.: Quelques remarques sur les observateurs non linéaires. In: Proceedings Colloque GRETSI Traitement du Signal et des Images, GRETS, vol. I, pp. 169–172 (1987)
5. Diop, S., Fliess, M.: On nonlinear observability. In: Commault, C., Normand-Cyrot, D., Dion, J.M., Dugard, L., Fliess, M., Titli, A., Cohen, G., Benveniste, A., Landau, I.D. (eds) Proceedings of the European Control Conference, Paris, Hermès. pp. 152–157 (1991)
6. Diop, S.: From the geometry to the algebra of nonlinear observability. In: Anzaldo-Meneses, A., Bonnard, B., Gauthier, J.P., Monroy-Perez, F. (eds) Contemporary Trends in Nonlinear Geometric Control Theory and its Applications, pp. 305–345. World Scientific Publishing Company, Singapore (2002)
7. Pommaret, J.F.: Partial Differential Control Theory. Volume II: Control Systems. Springer Science+Business Media Dordrecht, Dordrecht (2001)
8. Bejarano, F.J., Floquet, T., Perruquetti, W., Zheng, G.: Observability and detectability analysis of singular linear systems with unknown inputs. In: Proceedings of the IEEE Conference on Decision and Control. pp. 4005–4010. IEEE Press, New York (2011)
9. Diop, S., Wang, Y.: Equivalence between algebraic observability and generic local observability. In: Proceedings of the IEEE Conference on Decision and Control, vol. 3, pp. 2864–2865. IEEE Press, New York (1993)
10. Bächler, T., Gerdt, V., Lange-Hegermann, M., Robertz, D.: Thomas decomposition of algebraic and differential systems. Lecture Notes in Computer Science, Vol. 6244, pp. 31–54. Springer, Berlin (2002)
11. Lange-Hegermann, M., Robertz, D.: Thomas decompositions of parametric nonlinear control systems. Technical report (2012)
12. Sussmann, H.J.: Single-input observability of continuous-time systems. Math. Syst Theory **12**, 371–393 (1979)
13. Ritt, J.F.: Differential Algebra. American Mathematical Society, Providence (1950)
14. Seidenberg, A.: Some basic theorems in differential algebra (characteristic p, arbitrary). Trans. Am. Math. Soc. **73**, 174–190 (1952)
15. Kolchin, E.R.: Differential Algebra and Algebraic Groups. Academic, New York (1973)
16. Kaplansky, I.: An Introduction to Differential Algebra, 2nd edn. Hermann, Paris (1976)
17. Chèze, G., Galligo, A.: From an approximate to an exact absolute polynomial factorization. J. Symbolic Comput. **41**, 682–696 (2006)
18. Kaltofen, E., Maye, J.P., Yang, Z., Zhi, L.: Approximate factorization of multivariate polynomials using singular value decomposition. J. Symbolic Comput. **43**, 359–376 (2008)

Chapter 6
On Symbolic Approaches to Integro-Differential Equations

François Boulier, François Lemaire, Markus Rosenkranz, Rosane Ushirobira and Nathalie Verdière

Abstract Recent progress in computer algebra has opened new opportunities for the parameter estimation problem in nonlinear control theory, by means of integro-differential input–output equations. This paper recalls the origin of integro-differential equations. It presents new opportunities in nonlinear control theory. Finally, it reviews related recent theoretical approaches on integro-differential algebras, illustrating what an integro-differential elimination method might be and what benefits the parameter estimation problem would gain from it.

Keywords Parameter estimation · Symbolic methods · Integro-differential equations · Integro-differential algebras

6.1 Introduction

Under the impulse of the founding papers of Fliess [25], a school of researchers developed an approach of nonlinear control theory formulated within the framework of Ritt and Kolchin differential algebra [34, 48]. This approach led to various

F. Boulier (✉) · F. Lemaire
University of Lille, CNRS, Centrale Lille, UMR 9189, CRIStAL, CFHP, Lille, France
e-mail: Francois.Boulier@univ-lille1.fr

F. Lemaire
e-mail: Francois.Lemaire@univ-lille1.fr

M. Rosenkranz
University of Kent, SMSAS, Canterbury, UK
e-mail: M.Rosenkranz@kent.ac.uk

R. Ushirobira
Inria Lille Nord Europe, Non-A, Villeneuve d'Ascq, France
e-mail: Rosane.Ushirobira@inria.fr

N. Verdière
University Le Havre, LMAH, Le Havre, France
e-mail: Nathalie.Verdiere@univ-lehavre.fr

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_6

constructive methods which were implemented on computer algebra software dedicated for differential algebra [10, 11]. In this context, the present paper is concerned with a parameter estimation method, which is connected to an algorithmic structural identifiability test, based on the computation of the so-called input–output equation of the parametric nonlinear dynamical system under investigation [20, 41]. Since numerical integration is often much less sensitive to noisy data than numerical differentiation, the parameter estimation step of this method provides more reliable estimates of parameters, by first transforming the input–output equation into an integro-differential equation. This idea has been tested on a range of examples [21, 22, 26, 40, 58, 59]. This important transformation of nonlinear differential equations to integro-differential ones, which so far has required some human skill, can now be achieved algorithmically [8, 13].

Integral equations have other advantages as compared to differential ones. First, they permit to handle non smooth functions, in particular, piecewise constant inputs [21, 22, 26]. Second, they may naturally depend on initial conditions: a feature which may be important for the parameter estimation problem. In both cases, it may be interesting to bypass differential elimination in order to compute the desired equation.

These results and this research were at least partially motivated by purely theoretical studies of the algebraic properties of integro-differential algebras and their operator rings [3–5, 27, 29, 45, 52]. In turn, they raise the fascinating and difficult task of extending the Ritt-Kolchin theory known as *differential algebra* to the broader theory *integro-differential algebra* since integral or integro-differential equations are not allowed within the framework of differential algebra. One important goal would be an elimination theory for integro-differential algebra. It would allow the computation of integro-differential input–output equations using a wider set of operations than in differential algebra, hence possibly faster computations as well as a greater variety of formulations for the input–output equations.

This paper is structured as follows. Section 6.2 recalls the origin of integro-differential equations. Here, the term "origin" carries two meanings: what are the first historic examples of integro-differential models? and what kind of modelling processes lead to such models? This section will prove interesting for readers who discover integro-differential equations and for algebraists who are not aware of the needs of modellers. Section 6.3 sketches the application to parameter estimation for nonlinear dynamical systems that motivated this new interest for integro-differential equations. This section will be interesting for applied researchers who are not aware of some key properties of Ritt and Kolchin differential algebra: properties that need not generalize to the integro-differential framework. Last, Sect. 6.4 reviews some attempts to design algebraic theories of integro-differential equations and some of the many issues that need to be addressed. It illustrates also, via two examples, what an integro-differential elimination method could be and what would be the benefit to the parameter estimation problem.

6.2 Origin of Integro-Differential Models

One of the simplest nonlinear integro-differential models studied in the literature is the Volterra-Kostitzin model [35, pp. 66–69], which may be used for describing the evolution of a population, in a closed environment, intoxicated by its own metabolic products (other applications of the same model are considered in Kostitzin's book). It is an integro-differential equation since the unknown function $y(t)$ appears both differentiated and under some integral sign.

$$\frac{\mathrm{d}y}{\mathrm{d}t}(t) = \varepsilon\, y(t) - k\, y(t)^2 - c\, y(t) \int_{t-T}^{t} K(t-\tau)\, y(\tau)\, \mathrm{d}\tau\,.$$

The independent variable t is time. The dependent variable $y(t)$ is the population, varying with time. The symbols ε, k, c and T denote parameters. The *kernel* (or *nucleus*) $K(t,\tau) = K(t-\tau)$ is the *residual action function*. For instance, it could be very similar to a "survival function" in population dynamics [31, p. 3]: a decreasing function, starting at $K(0) = 1$, equal to 0 outside the interval $[0, T]$. Then $K(t-\tau)$ would represent the "toxicity factor" of metabolic products which are the most toxic when produced, at $t = \tau$, become less toxic with the time, and have a negligible toxic effect at time $t = \tau + T$.

As we shall see later, nontrivial kernels introduce difficulties in the symbolic treatment of integro-differential equations. It is thus interesting to remark that a simplified version of the Volterra-Kostitzin model, with a trivial kernel $K(t,\tau) = 1$, was studied by Kostitzin himself (the model is then equivalent to a differential equation of order two). It was more recently reconsidered in [17] and [43, Chap. 4] and fitted against experimental data, in order to validate its pertinence.

6.2.1 Hereditary Theories

In integral or integro-differential models, integral terms depend on kernels of the very special form $K(t,\tau) = K(t-\tau)$. Such kernels permit to express "hereditary", "historical" or "plastic" effects, i.e. the idea that the evolution of the current state of the system being modelled, depends not only on the current state but also on its past. The original qualifier is "hereditary". The qualifier "historical" was suggested by Volterra [62, p. 300] to avoid any confusion with biological notions. The qualifier "plastic" (by opposition to "elastic") is used in structural mechanics [38, p. 59].

6.2.1.1 Historical Origin

It is interesting to remark that biology is one of the first scientific domains where hereditary modelling was considered to be promising. In [60, p. 295], Volterra claims

to have coined the expression "integro-differential equation". A few lines further, he refers to an article by Picard [44], which contains the following paragraph (p. 194), translated from French:

> But heredity plays especially a major role in life sciences and we do not know if we will ever be able to use the mathematical tool for the intimate study of biological phenomena, and if we will not always need to restrict ourselves to rough averages and frequency curves. We should not, however, reduce in advance our mathematical conception of the world, and we can dream of functional equations more complicated than the former [differential] ones because they will involve, in addition, integrals taken between a very distant past and the present time, integrals which will bring their part of heredity.

The study of hereditary models is strongly connected with the theory of functionals, i.e. functions z that depend on all the values that some other function $y(t)$ may take over some range $a \leqslant t \leqslant b$. This was investigated in detail by Volterra [63], who sketched hereditary formulations of magnetism, electricity, elasticity, … It is intimately related to the theory of the convolution product, which arises in distribution theory [24], and whose algebraic properties were much studied by Volterra, as a special case of the "composition" of two functions. See [63] for the theory and [16, pp. 293–294] for more on the history.

6.2.2 *Some Classical Integro-Differential Models*

6.2.2.1 A Predator-Prey Model

The following integro-differential system [62, pp. 328–329] models two populations: one of them feeds on the other. It enhances a classical Volterra model of population dynamics. Volterra models here the fact that the increase of population does not depend only on the current amount of food available but also on the food which was available in the past.

$$\begin{aligned} \frac{\mathrm{d}y_1}{\mathrm{d}t}(t) &= y_1(t)\left(\varepsilon_1 - \gamma_1\, y_2(t) - \int_0^t f_1(t-\tau)\, y_2(\tau)\,\mathrm{d}\tau\right), \\ \frac{\mathrm{d}y_2}{\mathrm{d}t}(t) &= y_2(t)\left(-\varepsilon_2 + \gamma_2\, y_1(t) + \int_0^t f_2(t-\tau)\, y_1(\tau)\,\mathrm{d}\tau\right). \end{aligned} \tag{6.1}$$

6.2.2.2 Elastic Torsion of a Wire

This example is borrowed from [63, Chap. V, pp. 147–149]. To a first approximation the connection between the moment m of the torsional couple and the corresponding angle of torsion ω is given, in the case of static equilibrium, by the linear relation $\omega = k\, m$ where k is a constant depending on the characteristics of the wire. Let $m(\tau)$ denote the torsional moment acting on the wire at time τ. In order to find the angle of torsion $\omega(t)$ at time t we must add to the right-hand side of $\omega(t) = k\, m(t)$ a corrective

term depending on all the values of $m(\tau)$ for τ prior to t, and therefore a *functional* of $m(\tau)$. Assume the hereditary effects modelled by this functional is linear. One then obtains the following relation between $\omega(t)$ and $m(t)$:

$$\omega(t) = k\, m(t) + \int_{-\infty}^{t} f(t,\tau)\, m(\tau)\, \mathrm{d}\tau\,. \tag{6.2}$$

Solving this Volterra integral equation [63, Chap. II, P. 44] with respect to $m(t)$, denoting the reciprocal kernel of $\frac{1}{k} f(t,\tau)$ by $\varphi(t,\tau)$ and assuming the hereditary effects prior to $t = 0$ are negligible, we get

$$m(t) = \frac{1}{k}\,\omega(t) + \int_{0}^{t} \varphi(t,\tau)\,\omega(\tau)\,\mathrm{d}\tau\,. \tag{6.3}$$

Let us pass now from the static to the dynamic case and try to study the oscillations of the wire. For this, suppose that the angular velocity and acceleration are no longer negligible. The equation of motion of the wire is obtained from (6.3) by means of d'Alembert's principle, by substituting

$$m(t) - \mu\,\frac{\mathrm{d}^2\omega}{\mathrm{d}t^2}(t) \quad (\mu \,\text{constant})$$

for $m(t)$. We then get an integro-differential equation giving $\omega(t)$ in terms of $m(t)$:

$$m(t) - \mu\,\frac{\mathrm{d}^2\omega}{\mathrm{d}t^2}(t) = h\,\omega(t) + \int_{0}^{t} \varphi(t,\tau)\,\omega(\tau)\,\mathrm{d}\tau\,. \tag{6.4}$$

6.2.2.3 Propagation of a Nervous Impulse

This model is borrowed from [46, Chap. XXXV, Eq. (32), p. 426]. See also [31, Chap. 1, p. 5]. This nonpolynomial integral model is interesting because it has a trivial kernel and is equivalent to a polynomial system of integro-differential equations.

It is currently widely admitted that nervous impulses are propagated as follows in neurons: differences of ionic concentrations between the inside and the outside of axons make their membranes polarized. The occurrence of an electric current in the neighborhood of some region of interest opens ionic channels, causing changes of ionic concentrations, hence an electric current in the region itself. The nervous influx is obtained by repeating this phenomenon along the whole axon. The details of the ionic activities, *at a fixed position* of the axon, are described by the famous Hodgkin–Huxley nonlinear differential model [33, Chap. 5, pp. 205–206].

The following model is probably older than the Hodgkin–Huxley model [30]. It is built on quite similar biological hypotheses and is concerned by the distance $u(t)$ traveled by an influx along a nerve (an axon, possibly). The parameter I represents an electric current suddenly established in the neighborhood region at $t = 0$. The

parameter h is a concentration threshold above which a nerve excitation is triggered [46, Eq. (6), p. 379]. It is assumed that the ionic concentration and the electric current satisfy a linear differential equation of first order, depending on two parameters K and k [46, Eq. (18), p. 423]. The time t_1 at which the region of interest releases an influx is then a function of k, K, h and I. The parameter α is a constant depending on physical properties of the nerve: radius, specific resistances of the core of the nerve and its surrounding sheaths [46, Eq. (10), p. 421]. The integral term of the model comes here from the fact that the distance is an integral of the speed [46, Eq. (21), p. 424]. Since it is not motivated by any hereditary consideration, the absence of any nontrivial kernel is not surprising:

$$h\, e^{\alpha\, u(t)+k\, t} = \frac{h\, K\, I}{K\, I - k\, h} + K\, I \int_{t_1}^{t} e^{\alpha\, u(\tau)+k\, \tau}\, \mathrm{d}\tau\,. \tag{6.5}$$

Let now $v(t)$ be the exponential. The nonpolynomial integral equation (6.5) can be encoded by the following polynomial integro-differential system:

$$\begin{aligned} \frac{\mathrm{d}v}{\mathrm{d}t}(t) &= \left(\alpha\, \frac{\mathrm{d}u}{\mathrm{d}t}(t) + k\right) v(t)\,, \\ h\, v(t) &= \frac{h\, K\, I}{K\, I - k\, h} + K\, I \int_{t_1}^{t} v(\tau)\, \mathrm{d}\tau\,. \end{aligned} \tag{6.6}$$

Of course, the Eq. (6.5) can also easily be transformed to differential form, by simple differentiations. However, this is no longer true when considering a time-varying (possibly non-smooth) current $I(t)$.

6.3 Integro-Differential Equations for Parameter Estimation

In this section, we present the application to parameters estimation for nonlinear dynamical systems that motivated our interest for integro-differential algebra. Together with the application, we introduce key concepts of differential algebra. In the next section, we will discuss issues raised by their generalization to integro-differential algebra.

6.3.1 Statement of the Estimation Problem

The academic two-compartment model depicted in Fig. 6.1 is a close variant of [59, (1), p. 517] endowed with an input $u(t)$. Compartment 1 represents the blood system and compartment 2 represents some organ. Both compartments are supposed

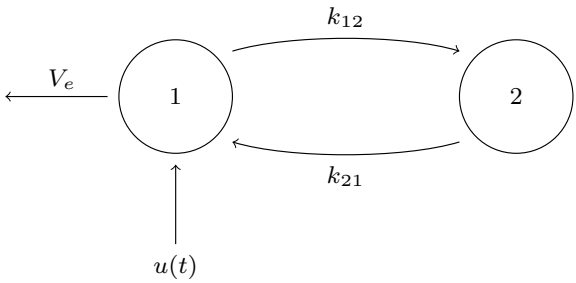


Fig. 6.1 A two-compartment model featuring three parameters.

to have unit volumes. The function $u(t)$, which has the dimension of a flow, represents a medical drug, injected in compartment 1. The drug diffuses between the two compartments, following linear laws: the proportionality constants are named k_{12} and k_{21}. The drug exits compartment 1, following a law of Michaelis-Menten type. Such a law indicates a hidden enzymatic reaction. In general, it depends on two constants V_e and k_e. For the sake of simplicity, it is assumed that $k_e = 1$. The state variables in this system are $x_1(t)$ and $x_2(t)$. They represent the concentrations of drug in each compartment. This information is sufficient to write the two first equations of the mathematical model (6.7). The last equation of (6.7) states that the output, denoted $y(t)$, is equal to $x_1(t)$. This means that only $x_1(t)$ is observed: some numerical data are available for $x_1(t)$ but not for $x_2(t)$. The problem addressed here then consists in estimating the three parameters k_{12}, k_{21} and V_e from these data and the knowledge of $u(t)$.

$$\begin{aligned} \dot{x}_1(t) &= -k_{12}\, x_1(t) + k_{21}\, x_2(t) - \frac{V_e\, x_1(t)}{1 + x_1(t)} + u(t)\,, \\ \dot{x}_2(t) &= k_{12}\, x_1(t) - k_{21}\, x_2(t)\,, \\ y(t) &= x_1(t)\,. \end{aligned} \tag{6.7}$$

6.3.2 The Algebraic Setting

6.3.2.1 On the Solutions

Ritt and Kolchin *differential algebra* provides an algebraic framework for polynomial differential systems. Differential systems involving rational fractions, such as (6.7), are easily handled. Other kinds of nonlinearities are not directly covered by the theory but many nonpolynomial systems can be transformed into polynomial ones, using techniques similar to the one we used for the propagation model of nervous impulses.

Differential algebra imposes, however, another restriction, which is more important to us, since it reduces its applicability to control theory: the solutions of the systems under study are supposed to belong to integral domains (e.g. an equation such as $u(t)\,v(t) = 0$ would imply that $u(t) = 0$ or $v(t) = 0$) and must be differentiable infinitely many times. The input $u(t)$ of (6.7) must then be smooth. One cannot study the case of a piecewise constant function $u(t)$ without leaving the realm of differential algebra.

6.3.2.2 Differential Polynomial Ideal

Subtract right-hand sides from left-hand sides of (6.7). Multiply the first equation by its denominator and state that this latter is nonzero. One obtains a system of three differential polynomial equations and one inequation:

$$p_1 = p_2 = p_3 = 0\,, \quad 1 + x_1 \neq 0\,. \tag{6.8}$$

The left-hand sides of (6.8) belong to the *differential polynomial ring*

$$\mathscr{R} = \mathbb{Q}(k_{12}, k_{21}, V_e)\{u, y, x_1, x_2\}\,.$$

The three symbols y, x_1, x_2 are *differential indeterminates*. To this system, one associates a *differential ideal*

$$\mathfrak{A} = [p_1, p_2, p_3] : (1 + x_1)\,.$$

Technically, the ideal $\mathfrak{A}$ is defined as the ideal of $\mathscr{R}$ generated by the three differential polynomials and their derivatives up to any order, *saturated* by the multiplicative family generated by $1 + x_1$. This means that if any differential polynomial of the form $(1 + x_1)\,g$ belongs to $\mathfrak{A}$, then g itself belongs to $\mathfrak{A}$. It can be proved that $\mathfrak{A}$ is a *prime* (hence *radical*) differential ideal.

6.3.2.3 Theorem of Zeros

As already pointed out, in differential algebra, solutions are sought in differential rings that are free of zero-divisors. This restriction has some drawbacks for applications in control theory. Algebraically, it has a big advantage: the differential ideal $\mathfrak{A}$ is then the set of all differential polynomials that annihilate over the whole solution set of (6.7). In particular, a Theorem of Zeros [48, Chap. I, 16] holds in differential algebra:

Theorem 1 *A differential polynomial g annihilates over all the solutions of a system of differential polynomial equations $f_1 = f_2 = \cdots = f_n = 0$ if, and only if, a power of g belongs to the differential ideal generated by $f_1, f_2, \ldots, f_n$.*

The formulation above is not completely precise since the algebraic structure $\mathscr{S}$ which is supposed to contain the solutions is not given. Many precise variants could be given: $\mathscr{S}$ could be some *differential field*, an algebra of formal power series, a field of meromorphic functions … See [9] for more details on this question.

6.3.2.4 Elimination Theory

An *elimination theory* has been available in differential algebra from its very beginning [48, 57]. Some algorithms such as [39, diffgrob], [14, 15, RosenfeldGroebner], [47, rif], [2, 36, 49, Thomas algorithm] are implemented in computer algebra systems. In the sequel, we concentrate on RosenfeldGroebner and its most recent implementation in the MAPLE package [10, DifferentialAlgebra]. The first argument of such an algorithm is a system of differential polynomial equations and inequations. The second argument is a *ranking*, i.e. a total ordering on the set

$$\{k_{12}, k_{21}, V_e\} \cup \{w^{(r)} \mid r \geqslant 0, w \text{ differential indeterminate}\}\,.$$

For an example such as (6.8), the output of the software is a *regular differential chain*—a notion slightly more general than Ritt's *characteristic set* [7, Definition 3.1]—of the differential ideal $\mathfrak{A}$ w.r.t. the ranking. Regular differential chains are finite sets of differential polynomials. By choosing a suitable ranking such as

$$(\text{the derivatives of } x_1, x_2) \gg (\text{the derivatives of } y, u) \gg (\text{the parameters})\,,$$

one can directly read in the output of the software the differential polynomial of $\mathfrak{A}$ which has the *lowest rank* w.r.t. the ranking. This differential polynomial is the so-called *differential input–output equation* of (6.7). It depends only on y, u, their derivatives, and the parameters to be estimated.

6.3.3 The Input–Output Equation of the Problem

A pretty-printed form of the differential input–output equation of (6.7) is:

$$\begin{aligned} -\theta_1\, u(t) + \theta_2\, \frac{y(t)}{y(t)+1} + \theta_3\, \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{y(t)^2}{y(t)+1}\right) \\ - \theta_4\, \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{y(t)+1}\right) = \dot{u}(t) - \ddot{y}(t)\,, \end{aligned} \tag{6.9}$$

where the θ_i stand for the following *blocks of parameters*:

$$\theta_1 = k_{21}\,, \quad \theta_2 = k_{21}\, V_e\,, \quad \theta_3 = k_{12} + k_{21}\,, \quad \theta_4 = k_{12} + k_{21} + V_e\,. \tag{6.10}$$

6.3.3.1 Structural Identifiability Study

The *structural identifiability study* is a preliminary study of the input–output equation. It can be viewed as a theoretical parameter estimation process, where the observed function $y(t)$, the input $u(t)$, and their derivatives up to any order are supposed to be as "generic" as possible and perfectly known.

A huge amount of literature is devoted to this question. See [1, 23, 37, 41, 42, 56, 59] and the references therein.

In our example, the structural identifiability study leads in a straightforward way to the desired conclusion of the global structural identifiability for this model. The essential argument is as follows:

(a) Evaluating (6.9) at (at least) four different values of t, it is possible to build an invertible linear system whose unknowns are the blocks of parameters θ_i.
(b) Knowing the blocks of parameters θ_i, it is easy to recover the values of the model parameters k_{12}, k_{21}, V_e, by solving the polynomial system (6.10).

It is worth noticing that step 1 of this argument eventually relies on the assumption that (6.9) is the equation of minimal order and degree constraining $y(t), u(t)$ and the parameters to be estimated. Ultimately, this argument relies on the Theorem of Zeros.

Indeed, on some other example, a non-minimal input–output equation could be artificially obtained by adding a polynomial of the form $\theta\, m$ (θ any parameter block, m belonging to the differential ideal $\mathfrak{A}$) to the minimal equation. At step 1, such a polynomial m would always evaluate to zero, yielding a linear system that would always be singular. The whole argument would then collapse.

6.3.3.2 Integro-Differential Form of the Input–Output Equation

A parameter estimation method can be designed by implementing "numerically" the steps 1 and 2 above, using the available numerical data for the observed function $y(t)$ and the input $u(t)$. If the data is noisy (but not only then), a straightforward implementation is however very likely to produce useless results since the second derivative of $y(t)$ then needs to be estimated numerically. In order to obtain more accurate results, it is desirable to convert the differential input–output equation (6.9) to integral form since numerical integration is less sensitive to noise than numerical differentiation. There are different ways to do it. Some methods are given in [12]. One possibility consists in applying twice the integration operator on (6.9). On our example, the result is the nonlinear Volterra integral equation (6.11). This formula still involves a derivative of the output, but evaluated at $t = a$. Viewing this derivative as an extra parameter θ_5, to be estimated, Eq. (6.11) does not involve any derivative of the output:

$$
\begin{aligned}
&- \theta_1 \int_a^t \int_a^{\tau_1} u(\tau_2)\, \mathrm{d}\tau_2\, \mathrm{d}\tau_1 \\
&+ \theta_2 \int_a^t \int_a^{\tau_1} \frac{y(\tau_2)}{y(\tau_2)+1}\, \mathrm{d}\tau_2\, \mathrm{d}\tau_1 \\
&+ \theta_3 \left(\int_a^t \frac{y(\tau)^2}{y(\tau)+1}\, \mathrm{d}\tau - \frac{y(a)^2}{y(a)+1}\,(t-a) \right) \\
&- \theta_4 \left(\int_a^t \frac{1}{y(\tau)+1}\, \mathrm{d}\tau - \frac{1}{y(a)+1}\,(t-a) \right) \\
&- \dot{y}(a)\,(t-a) \\
&= \int_a^t u(\tau)\, \mathrm{d}\tau - u(a)\,(t-a) - y(t) + y(a)\,.
\end{aligned}
\tag{6.11}
$$

In general, the resulting formula is an integro-differential equation. The method outlined here always produces formulas with trivial kernels. It can be completely automated, thanks to recent progresses in computer algebra [8, 12, 13].

6.3.3.3 Actual Parameter Estimation

An integro-differential input–output equation such as (6.11) is used to build an overdetermined[1] linear system, whose unknowns are the parameter blocks θ_i. Its solutions, obtained by linear least squares, may be used as a first guess for nonlinear least squares such as the Levenberg–Marquardt method or Linear Matrix Inequalities. See [19, 20, 41, 59].

Recovering the initial model parameters from refined estimates of the parameter blocks θ_i is usually a difficult problem for which no satisfactory general solution is known. An important difficulty is raised by possible algebraic relations between blocks of parameters. In our example, we have such a relation:

$$\theta_1\,(\theta_3 - \theta_4) + \theta_2 = 0\,. \tag{6.12}$$

6.3.4 *Algorithmic Transformation to Integro-Differential Form*

Transforming (6.9) into the integral equation (6.11) is very easy because the former equation has the special form

$$\sum \theta_i \frac{\mathrm{d}}{\mathrm{d}t} F_i\,,$$

where the θ_i are constant expressions (their derivatives are zero) and the F_i are order zero fractions. However, the input–output equations returned by differential

[1] A square linear system is sufficient for the identifiability study, which is of theoretical nature. On real, possibly noisy, data, it is preferable to evaluate the input–output equations at much more values of t, and thereby obtain an overdetermined linear system.

elimination algorithms do not always have this shape. Instead, they have the form of polynomials in the derivatives of the differential indeterminates, which implies that the $\frac{\mathrm{d}}{\mathrm{d}t} F_i$ expressions are expanded. A first algorithm for converting this raw form into (6.9) is published in [8], with a flaw fixed in [12]. An enhanced version, with a canonical output, is published in [13].

6.4 Towards Algebraic Theories

The systematic treatment of integral operators in algebra—usually under the name of Rota-Baxter operators—has been inaugurated with Glen Baxter's seminal paper [6]. While originally viewed in a probability context, Rota-Baxter operators have soon found interest in broader areas of algebra, especially through Gian-Carlo Rota's well-known papers [54, 55]. For a modern survey on Rota-Baxter algebra we refer to the monograph [28].

The notion of Rota-Baxter operator was combined with differential algebra structures in [50, 53] for creating an algebraic framework that allows a constructive treatment of boundary problems for linear ordinary differential equations. In particular, the Green's operator (resolvent operator), which maps the forcing function to the solution of the boundary problem, is expressed as a Fredholm integral operator that belongs to a suitable operator ring.

Let us explain this in some more detail. We start from an *integro-differential algebra* $(\mathcal{F}, \partial, \int)$, meaning an algebra over some field K with two K-linear operators $\partial, \int : \mathcal{F} \to \mathcal{F}$ that are supposed to capture differentiation and integration in an algebraic context. Hence the derivation ∂ is required to satisfy the Leibniz axiom (= product rule for differentiation) while $\int$ must satisfy the Rota-Baxter axiom (= integration by parts); moreover we stipulate $\partial \circ \int = 1_{\mathcal{F}}$ for tying the two notions together, just as the fundamental theorem of calculus does in the case $\mathcal{F} = C^\infty(\mathbb{R})$; note that this is a special case of the generalized Leibniz integral rule (6.14). It turns out that the other composition is not quite the identity but $\int \circ \partial = 1_{\mathcal{F}} - E$, where $E : \mathcal{F} \to \mathcal{F}$ is a multiplicative linear map that may be thought as the evaluation at the initialization point of the integral operator $\int$. Indeed, this is what happens in the most important example $\mathcal{F} = C^\infty(\mathbb{R})$ where $\partial f(x) = df/dx$ and $\int f(x) = \int_a^x f(\xi)\, d\xi$ for some initialization point $a \in \mathbb{R}$; consequently, here $E : \mathcal{F} \to \mathbb{R}$ is the evaluation $f(x) \mapsto f(a)$. Of course $C^\infty(\mathbb{R})$ contains many other integro-differential algebras, for example the polynomials $\mathbb{R}[x]$ or the analytic functions $C^\omega(\mathbb{R})$, and various intermediate algebras like the exponential polynomials (real or complex linear combinations of $x^k e^{\lambda x}$ for any $k \in \mathbb{N}$ and $\lambda \in \mathbb{R}$).

Each integro-differential algebra $(\mathcal{F}, \partial, \int)$ now gives rise to an operator ring $\mathcal{F}[\partial, \int]$ that contains both differential operators such as $a(x)\,\partial^2 + b(x)\,\partial + c(x)$ for coefficient functions $a(x), b(x), c(x) \in \mathcal{F}$ and integral operators such as $x^2 e^x \int e^{-2x}$, as well as evaluation operators E_a for various points $a \in \mathbb{R}$. Note that the integral operator $x^2 e^x \int e^{-2x}$ is here understood as a non-commutative operator composition, acting as $f(x) \mapsto \int_a^x x^2 e^{x-2\xi} f(\xi)$. We refer to $\mathcal{F}[\partial, \int]$ as the *integro-differential*

operator ring over $\mathcal{F}$. One can prove that every boundary problem over $\mathcal{F}$ has a Green's operator $G \in \mathcal{F}[\partial, \int]$, which can be determined algorithmically if one has a fundamental system of solutions for the underlying homogeneous differential equation [53, Thm. 26].

The treatment of *linear partial differential equations* is considerably more difficult. Of course one will replace $(\mathcal{F}, \partial, \int)$ by a structure $(\mathcal{F}, \partial_x, \partial_y, \int^x, \int^y)$, where $\mathcal{F}$ is an algebraic structure describing multivariate (for simplicity here: bivariate) functions with two derivations ∂_x, ∂_y and two Rota-Baxter operators $\int^x, \int^y$. However, this would not lead us very far since even very simple examples like $u_x - u_y = f$ cannot be solved in terms of these basic building blocks. One crucial missing piece is a structure for *substitutions*. For treating linear partial differential equations, it is sufficient to allow only linear substitutions like $u(x, y) \mapsto u(ax + by, cx + dy)$ for $a, b, c, d \in \mathbb{R}$. The resulting algebraic structure, a so-called *Rota-Baxter hierarchy*, is surprisingly complex and has been described in [51]. In fact, the current setup omits the derivations ∂_x, ∂_y for keeping the complications to a minimum (derivations are comparatively easy to add since their algebraic relations are far less complex than those connected with the Rota-Baxter operators).

Similar to the case of plain integro-differential algebras, every Rota-Baxter hierarchy comes with a *multivariate operator ring* that we could provisionally denote by $\mathcal{F}[\int^x, \int^y]$ or by $\mathcal{F}[\partial_x, \partial_y, \int^x, \int^y]$ if the derivations are added in. The main innovation from the ordinary case is that substitution operators $M^* = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)^*$, acting as described above for a matrix $M \in \mathbb{R}^{2\times 2}$, are also part of the basic building blocks. Their interaction with the Rota-Baxter operators is complicated, but normal forms have been deduced [51, Thm 4.10], for the case of arbitrarily many variables.

Describing the nature of those operator relations would lead us too far afield for the present paper. It will suffice to mention just one special case of Axiom (7) of [51, Def. 2.3], namely

$$\int^x M^* \int^x = M^* \int^x \int^y - \int^x M^* \int^y,$$

where $M \in \mathbb{R}^{2\times 2}$ is the substitution matrix with $a = c = 1, b = d = 0$, acting by $u(x, y) \mapsto u(x, x)$. Written in the usual notation, this is exactly *Dirichlet's rule* to be mentioned in Sect. 6.4.1.3. However, the great advantage of the operator-ring framework is that there is a normal form (namely the right-hand side). In fact, we are confident that these normal forms are in fact canonical (meaning every simplification can only lead to a single normal form), which is equivalent to the algebraic statement that the chosen identities are a non-commutative Gröbner basis for the ideal of operator relations. This is work in progress, more than half of the proof is completed but there the derivations are very long.

Up to now we have spoken about linear differential and integral equations, both ordinary and partial (of course this includes also the mixed integro-differential cases). For passing to nonlinear equations, one can pass to the ring of *integro-differential polynomials*. In the univariate case, this has been introduced in [52]. Essentially the same structure—but extended to the partial case as well as differential fractions—was subsequently treated in [8, 13]. The basic idea in all case is that nested integrals

like $\int u^2 u'^3 \int u' u''^2 \int u'' (u''')^5$ or $\int u'^2 \int u'' \int u u''^4$, or linear combinations of these, are in canonical form only if the highest derivative of u appears nonlinearly in each of the nested integrands. Hence the first example above is canonical, but the second is not.

While the algebraic investigation of the linear case (univariate and multivariate integro-differential operator rings) are now gradually maturing to some extent, the nonlinear case of *integro-differential polynomials/fractions* is wide open. In the last two sections we will only address two prominent issues that appear to pose considerable difficulties and, by the same token, a highly interesting arena of algebraic research.

6.4.1 Computational Issues

6.4.1.1 The Generalized Leibniz Integral Rule

Though there does exist applications of integro-differential equations which only need trivial kernels, this is certainly not the case of equations arising from hereditary modelling, which have the form:

$$\int_a^t K(t,\tau)\, f(\tau)\, \mathrm{d}\tau\,.$$

Over such expressions, the formula

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_a^t = \text{the identity operator} \tag{6.13}$$

does not hold anymore. Instead, one must apply the general form of Leibniz integral rule, which gives:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_a^t K(t,\tau)\, f(\tau)\, \mathrm{d}\tau = \int_a^t \frac{\mathrm{d}K}{\mathrm{d}t}(t,\tau)\, f(\tau)\, \mathrm{d}\tau + K(t,t)\, f(t)\,. \tag{6.14}$$

However, this formula raises a computational problem, in the case of singular kernels. Let us quote Volterra [63, Chap. II, p. 54]:

> It is rather curious to note that it was precisely these singular cases that were the first in order of time to arise; the first integral equation considered goes back to Abel and is as follows:
>
> $$\sqrt{2g}\, z(t) = \int_0^t \frac{y(\tau)}{(t-\tau)^{\frac{1}{2}}}\, \mathrm{d}\tau\,, \tag{6.15}$$
>
> and the kernel $(t-\tau)^{-\frac{1}{2}}$ becomes infinite at $\tau = t$.

Differentiating (6.15) causes a division by zero. In particular, we see that, contrarily to what happens in differential algebra, differentiation of integro-differential expressions is not always defined.

6.4.1.2 Integration by Change of Variable

The change of variable formula is:

$$\int_a^b F(\varphi(t))\,\frac{\mathrm{d}\varphi}{\mathrm{d}t}(t)\,\mathrm{d}t = \int_{\varphi(a)}^{\varphi(b)} F(t)\,\mathrm{d}t\,.$$

Kostitzin applies it [35, p. 68] for studying the Volterra-Kostitzin model in the case of a trivial kernel. Having proven that

$$\frac{\mathrm{d}y}{\mathrm{d}t}(t) = F(y(t))\,,$$

he deduces that

$$t = \int_{y(0)}^{y(t)} \frac{1}{F(\tau)}\,\mathrm{d}\tau\,.$$

This identity shows that, in an integro-differential algebra theory in which integration operators with general bounds would be allowed, the equality test between two expressions might be a difficult problem.

6.4.1.3 Dirichlet's Rule

Volterra attributes this formula to Dirichlet in [61, p. 36] and often uses it:

$$\int_a^t \int_a^\tau F(\tau_2,\tau)\,\mathrm{d}\tau_2\,\mathrm{d}\tau = \int_a^t \int_{\tau_2}^t F(\tau_2,\tau)\,\mathrm{d}\tau\,\mathrm{d}\tau_2\,.$$

In the parameter estimation problem described in Sect. 6.3, this formula could have been used in order to produce another form of the integral equation (6.11), which would have involved a non trivial kernel. Indeed:

$$\int_a^t \int_a^\tau F(\tau_2)\,\mathrm{d}\tau_2\,\mathrm{d}\tau = \int_a^t (t-\tau)\,F(\tau)\,\mathrm{d}\tau\,.$$

This formula illustrates another difficulty that arises when testing equality between two integro-differential expressions. Recent progresses on this issue are given in [51].

6.4.2 On Generalizations of the Theorem of Zeros

As pointed out in Sect. 6.3, the Theorem of Zeros is implicitly used in the structural identifiability test based on the input–output equation. This section shows that it might not generalize in the integro-differential framework. Observe a similar difficulty occurs in other theories such as difference algebra [18].

Let us define an integro-differential ring $\mathscr{R} = \mathbb{Q}\{u\}$ as the smallest ring containing the integro-differential indeterminate u, the rational numbers, stable under derivation and integration. For simplicity, let us restrict ourselves to the case of a derivation δ being the left inverse of the integration $\int$, i.e. an abstract form of (6.13). Let us denote $x = \int 1$.

Given any $p \in \mathscr{R}$, let us define the integro-differential ideal generated by p as the smallest ideal of the ring $\mathscr{R}$, containing p, stable under derivation and integration. Let us denote it $[p]$. Consider now

$$p = u - \int u\,, \tag{6.16}$$

which is meant to be an abstract form of the left-hand side of the integral equation

$$u(t) - \int_0^t u(\tau)\,\mathrm{d}\tau = 0\,.$$

This equation admits $u(t) = 0$ for unique solution. However, we prove below that $u^m \notin [p]$ for any non-negative integer m, i.e. that, in the algebraic framework sketched above, the Theorem of Zeros does not hold.

Proposition 1 *Take $p \in \mathscr{R}$. Then, any element q of $[p]$ can be written as $q = \sum_{i=1}^{s} a_i\, M_i$ where $a_i \in \mathbb{Q}$ and each M_i has the form*

$$M_i = m_{i,0} \int m_{i,1} \int \cdots \int m_{i,k-1} \int m_{i,k}(\delta^{b_i} p) \int m_{i,k+1} \int \cdots \int m_{i,t_i}\,, \tag{6.17}$$

where b_i is a non-negative integer, the $m_{i,j}$ are monomials in x, and u and its derivatives.

Proof Admitted.

Let us denote by $w(M_i)$ the *weight in u* of any M_i of the form of (6.17), with $w(M_i) = 1 + \sum_{j=0}^{t_i} \deg(m_{i,j}, [u, \delta u, \ldots])$, where $\deg(m_{i,j}, [u, \delta u, \ldots])$ denotes the total degree of $m_{i,j}$ in the variables u, δu, …

Lemma 1 *Take $p = u - \int u$ and consider some M_i in the form of (6.17). Then replacing u by $\alpha\, u$ in M_i yields $\alpha^{w(M_i)} M_i$, for any $\alpha \in \mathbb{Q}$. Moreover, replacing u by e^x in M_i yields 0 if $b_i > 0$, or a polynomial in x and e^x whose degree in e^x is at most $w(M_i) - 1$ otherwise.*

Proof Immediate.

Proposition 2 *Take* $p = u - \int u$. *Then* $u^m \notin [p]$ *for any non-negative integer* m.

Proof Assume that $u^m \in [p]$ for some m. Let us prove that this yields a contradiction. By Proposition 1, we have $u^m = \sum_{i=1}^{s} a_i\, M_i$ where the a_i are in $\mathbb{Q}$ and the M_i have the form of (6.17). The monomial u^m is homogeneous of degree m. By an homogeneity argument and by Lemma 1, all M_i have the same weights $w(M_i) = m$ in u.

Substituting $u = e^x$ in u^m yields e^{mx}. However, substituting $u = e^x$ in any M_i either yields 0 if $b_i > 0$, or a polynomial in x and e^x whose degree in e^x is at most $m - 1$ by Lemma 1. This yields a contradiction since e^x, e^{2x}, …, e^{mx} are linearly independent over $\mathbb{Q}[x]$.

6.4.3 On Derivation-Free Elimination

The most challenging theoretical issue would consist in developing an elimination theory for integro-differential equations that would permit to bypass differential algebra methods. To our knowledge, such a derivation-free elimination theory has not been considered yet. To illustrate what it could ideally do, we show that Eq. (6.11) can be obtained from Eq. (6.7) without performing any differentiation. Let us slightly rewrite Eq. (6.7) as

$$\begin{aligned} f_1(t) &:= -\dot{y}(t) - k_{12}\, y(t) + k_{21}\, x_2(t) - V_e\, \frac{y(t)}{1 + y(t)} + u(t)\,, \\ f_2(t) &:= -\dot{x}_2(t) + k_{12}\, y(t) - k_{21}\, x_2(t)\,. \end{aligned}$$

Since $f_1(t)$ and $f_2(t)$ are identically zero, we obtain the equation

$$0 = k_{21} \int_a^t \int_a^\tau f_1(\tau_2) + f_2(\tau_2)\, \mathrm{d}\tau_2\, \mathrm{d}\tau + \int_a^t f_1(\tau)\, \mathrm{d}\tau\,.$$

Simplifying the previous equation yields

$$\begin{aligned} k_{21} \int_a^t \int_a^\tau u(\tau_2)\, \mathrm{d}\tau_2\, \mathrm{d}\tau - k_{21}\, V_e \int_a^t \int_a^\tau \frac{y(\tau_2)}{1 + y(\tau_2)}\, \mathrm{d}\tau_2\, \mathrm{d}\tau \\ - (k_{21} + k_{12}) \int_a^t y(\tau)\, \mathrm{d}\tau - V_e \int_a^t \frac{y(\tau)}{1 + y(\tau)}\, \mathrm{d}\tau + k_{21}\, (y(a) + x_2(a))\, (t - a) \\ = y(t) - y(a) - \int_a^t u(\tau)\, \mathrm{d}\tau\,. \end{aligned}$$

From $f_1(a) = 0$, one has

$$k_{21}\, x_2(a) = \dot{y}(a) + k_{12}\, y(a) + V_e\, \frac{y(a)}{1 + y(a)} - u(a)\,.$$

Replace $k_{21}\,x_2(a)$ by its value in the last equation. Simple computations show that

$$\begin{aligned}
&k_{21}\left(\int_a^t\int_a^\tau u(\tau_2)\,\mathrm{d}\tau_2\,\mathrm{d}\tau + y(a)\,(t-a) - \int_a^t y(\tau)\,\mathrm{d}\tau\right)\\
-\,k_{21}\,V_e\int_a^t\int_a^\tau \frac{y(\tau_2)}{1+y(\tau_2)}\,\mathrm{d}\tau_2\,\mathrm{d}\tau &- V_e\left(\int_a^t \frac{y(\tau)}{1+y(\tau)}\,\mathrm{d}\tau - \frac{y(a)}{1+y(a)}\right)\\
&+k_{12}\left(y(a)\,(t-a) - \int_a^t y(\tau)\,\mathrm{d}\tau\right) + \dot{y}(a)\,(t-a)\\
&= y(t) - y(a) - \int_a^t u(\tau)\,\mathrm{d}\tau + u(a)\,(t-a)\,.
\end{aligned}$$

This last equation is equivalent (up to the sign) to (6.11), by using the properties $y^2/(1+y) = y - 1 + 1/(1+y)$ and $1/(y+1) = 1 - y/(y+1)$.

6.4.4 On Alternative Input–Output Equations

A structural identifiability study, based on the differential input–output equation, was sketched in Sect. 6.3.3.1. Since this equation is computed in the strict framework of differential algebra, it does not feature any initial value of any non-observed variable. In some cases, however, the initial conditions of some non-observed variables are known, and their knowledge is necessary to prove the structural identifiability of the model. A first approach to overcome this difficulty is presented in [21, 22]: the integration of the input–output equations followed by some further manipulations yields expressions involving the initial conditions of the non-observed variables, permitting to prove the structural identifiability. We show in this section that integro-differential elimination offers another approach since it permits to compute integral input–output equations featuring naturally these important initial values. The system Σ under study is inspired from [32]:

$$\dot{x}_1(t) = \theta_1\,x_2(t) + u(t)\,, \tag{6.18}$$
$$\dot{x}_2(t) = \theta_2\,x_1(t)\,x_2(t) + \theta_3\,x_2(t) + u(t)\,, \tag{6.19}$$

together with the assumptions

- $x_1(0) = x_{10}$ is known,
- $x_2(t)$ and $u(t)$ are observed on some interval $[0, t_0]$,
- $x_1(t)$ is *not* observed on $]0, t_0]$.

The parameters to be estimated are θ_1, θ_2 and θ_3. Differential elimination methods permit to compute the following differential input–output equation:

$$\ddot{x}_2(t)\,x_2(t) - \dot{x}_2^2(t) + \dot{x}_2(t)\,u(t) - x_2(t)\,\dot{u}(t) - \theta_1\,\theta_2\,x_2^3(t) - \theta_2\,u(t)\,x_2^2(t)\,.$$

The structural identifiability study sketched in Sect. 6.3.3.1 would conclude to the non-identifiability of Σ since θ_3 does not even appear in this equation. Converting this equation to integro-differential form would obviously not change this conclusion. Now, it is interesting to observe that one can obtain an expression depending on θ_3 by evaluating (6.19) at $t = 0$, provided that $x_2(0) \neq 0$. Another expression can be obtained if $x_2(0) = 0$ and $\dot{x}_2(0) \neq 0$: differentiating (6.19) and rewriting the term $\dot{x_1}(t)$ using (6.18) yields

$$\ddot{x}_2(t) = \theta_2 \Big((\theta_1 x_2(t) + u(t)) x_2(t) + x_1(t) \dot{x}_2(t) \Big) + \theta_3 \dot{x}_2(t) + \dot{u}(t) . \tag{6.20}$$

Evaluating this expression at $t = 0$ provides an expression for θ_3 as a function of $x_1(0)$, $x_2(0)$, $\dot{x_2}(0)$, $\ddot{x}_2(0)$, $u(0)$, $\dot{u}(0)$, θ_1 and θ_2. More generally, a formula can be obtained provided that some derivative of $x_2(t)$ does not vanish at $t = 0$.

Let us now compute an integral input–output equation, using the integration operator from the beginning. First put Σ in an integral form:

$$x_1(t) = x_{10} + \int_0^t \theta_1 x_2(\tau) + u(\tau) \, \mathrm{d}\tau , \tag{6.21}$$

$$x_2(t) = x_{20} + \int_0^t \theta_2 x_1(\tau) x_2(\tau) + \theta_3 x_2(\tau) + u(\tau) \, \mathrm{d}\tau . \tag{6.22}$$

Using (6.21) for replacing $x_1(t)$ by its value in (6.22) yields

$$x_2(t) = x_{20} + \int_0^t \theta_2 \left(x_{10} + \int_0^\tau \theta_1 x_2(\tau_2) + u(\tau_2) \, \mathrm{d}\tau_2 \right) x_2(\tau) + \theta_3 x_2(\tau) + u(\tau) \, \mathrm{d}\tau . \tag{6.23}$$

Expanding (6.23) yields

$$x_2(t) = I_0(t) + (x_{10} \theta_2 + \theta_3) I_1(t) + \theta_2 I_2(t) + \theta_1 \theta_2 I_3(t) \tag{6.24}$$

with

$$I_0(t) = x_{20} + \int_0^t u(\tau) \, \mathrm{d}\tau , \qquad I_1(t) = \int_0^t x_2(\tau) \, \mathrm{d}\tau ,$$

$$I_2(t) = \int_0^t x_2(\tau) \int_0^\tau u(\tau_2) \, \mathrm{d}\tau_2 \, \mathrm{d}\tau , \qquad I_3(t) = \int_0^t x_2(\tau) \int_0^\tau x_2(\tau_2) \, \mathrm{d}\tau_2 \, \mathrm{d}\tau .$$

Let us now follow the structural identifiability study sketched in Sect. 6.3.3.1 on (6.24). The linear system considered at step 1 is invertible since the three terms $I_1(t)$, $I_2(t)$ and $I_3(t)$ are linearly independent.[2] Thus the structural identifiability

[2] Indeed, if the three terms were linearly dependent, there would exist three constants A, B, C such that $A I_1(t) + B I_2(t) + C I_3(t) = 0$ modulo the prime differential ideal $\mathfrak{A}$ generated by Σ. Differentiate this equation. Divide it by $x_2(t)$. Differentiate again. One gets $B u(t) + C x_2(t) = 0$.

study would infer the global structural identifiability of Σ. The input–output equation obtained by integro-differential elimination is therefore not equivalent to the one obtained by plain differential elimination.

References

1. Audoly, S., Bellu, G., D'Angio, L., Saccomani, M.P., Cobelli, C.: Global identifiability of nonlinear models of biological systems. IEEE Trans. Biomed. Eng. **48**(1), 55–65 (2001)
2. Bächler, T., Gerdt, V., Lange-Hegermann, M., Robertz, D.: Algorithmic Thomas decomposition of algebraic and differential systems. J. Symb. Comput. **47**(10), 1233–1266 (2012)
3. Bavula, V.V.: The algebra of integro-differential operators on a polynomial algebra. J. Lond. Math. Soc. **83**(2), 517–543 (2011)
4. Bavula, V.V.: The algebra of integro-differential operators on an affine line and its modules. J. Pure Appl. Algebra **17**(3), 495–529 (2013)
5. Bavula, V.V.: The algebra of polynomial integro-differential operators is a holonomic bimodule over the subalgebra of polynomial differential operators. Algebras Represent. Theory **17**(1), 275–288 (2014)
6. Baxter, G.: An operator identity. Pac. J. Math. **8**, 649–663 (1958)
7. Boulier, F., Lemaire, F.: A normal form algorithm for regular differential chains. Math. Comput. Sci. **4**(2), 185–201 (2010). https://doi.org/10.1007/s11786-010-0060-3
8. Boulier, F., Lemaire, F., Regensburger, G., Rosenkranz, M.: On the integration of differential fractions. In: Proceedings of the 38th International Symposium on Symbolic and Algebraic Computation, ISSAC'13, pp. 101–108, New York, NY, USA, 2013. ACM
9. Boulier, F., Lemaire, F.: A computer scientist point of view on Hilbert's differential theorem of zeros. (preprint) (2007). http://hal.archives-ouvertes.fr/hal-00170091
10. Boulier, F., Cheb-Terrab, E.: Differential Algebra. Package of MapleSoft MAPLE standard library since MAPLE 14 (2008)
11. Boulier, F., Hubert, É.: diffalg. Package of MapleSoft MAPLE standard library from MAPLE V to MAPLE 13 (1996)
12. Boulier, F., Korporal, A., Lemaire, F., Perruquetti, W., Poteaux, A., Ushirobira, R.: An algorithm for converting nonlinear differential equations to integral equations with an application to parameter estimation from noisy data. In: LNCS 8660: Proceedings of Computer Algebra and Scientific Computing (CASC) 2014, Warsaw, Poland, pp. 28–43 (2014)
13. Boulier, F., Lallemand, J., Lemaire, F., Regensburger, G., Rosenkranz, M.: Additive normal forms and integration of differential fractions. J. Symb. Comput. **77**, 16–38 (2016)
14. Boulier, F., Lazard, D., Ollivier, F., Petitot, M.: Representation for the radical of a finitely generated differential ideal. In: ISSAC'95: Proceedings of the 1995 international symposium on Symbolic and algebraic computation, pp. 158–166. ACM Press, New York (1995). http://hal.archives-ouvertes.fr/hal-00138020
15. Boulier, F., Lazard, D., Ollivier, F., Petitot, M.: Computing representations for radicals of finitely generated differential ideals. Appl. Algebra Eng. Commun. Comput. **20**(1), 73–121 (2009). (1997 Techrep. IT306 of the LIFL)
16. Bourbaki, N.: Éléments d'Histoire des Mathématiques, Collection Histoire de la Pensée, Vol. iv, 2nd edn. Hermann (1969)
17. Chassé, J.L., Legay, J.M., Pavé, A.: Le modèle de Volterra-Kostitzin en dynamique des populations. Ajustement et interprétation des paramètres. Ann. Zool. Ecol. Anim. **9**(3) (1977)
18. Cohn, R.: Difference Algebra. Interscience Publishers (1965)

However, $B\,u(t) + C\,x_2(t)$ is not reduced to zero by Σ, viewed as a regular differential chain (or a characteristic set) of $\mathfrak{A}$. This contradiction proves the linear independence of the three terms.

19. Denis-Vidal, L., Cherfi, Z., Talon, V., Brahmi, E.H.: Parameter identifiability and parameter estimation of a diesel engine combustion model. J. Appl. Math. Phys. **2**, 131–137 (2014)
20. Denis-Vidal, L., Joly-Blanchard, G., Noiret, C.: System identifiability (symbolic computation) and parameter estimation (numerical computation). Numer. Algorithms **34**, 282–292 (2003)
21. Denis-Vidal, L., Joly-Blanchard, G., Verdière, N.: Identifiability of ordinary or delayed nonlinear models: a distribution approach. In: Proceedings of the IEEE, Munich (2006)
22. Denis-Vidal, L., Joly-Blanchard, G., Verdière, N.: Identifiability and estimation of nonlinear models: a distribution framework. In: Proceedings of the ECC, Greece, Koos (2007)
23. Diop, S., Fliess, M.: Nonlinear observability, identifiability, and persistent trajectories. In: Proceedings of 30th CDC, Brighton, pp. 714–719 (1991)
24. Duistermaat, J.J., Kolk, J.A.C.: Distributions: Theory and Applications. Birkhäuser, Basel (2010)
25. Fliess, M.: Automatique et corps différentiels. Forum Math. **1**, 227–238 (1989)
26. Fliess, M., Mboup, M., Mounier, H., Sira-Ramirez, H.: Questioning some paradigms of signal processing via concrete examples. In: Silva-Navarro, G., Sira-Ramirez, H. (eds.) Algebraic Methods in Flatness, Signal Processing and State Estimation, Algebraic Methods in Flatness, Signal Processing and State Estimation, Mexico, pp. 1–21 (2003). Editorial Lagares
27. Gao, X., Guo, L.: Constructions of free commutative integro-differential algebras. In: Barkatou, M., Cluzeau, T., Regensburger, G., Rosenkranz, M. (eds.) Algebraic and Algorithmic Aspects of Differential and Integral Operators. LNCS, vol. 8372, pp. 1–22 (2014)
28. Guo, L.: An Introduction to Rota-Baxter Algebras. International Press (2012)
29. Guo, L., Regensburger, G., Rosenkranz, M.: On integro-differential algebras. J. Pure Appl. Algebra **218**(3), 456–473 (2014)
30. Hodgkin, A.L., Huxley, A.F.: A Quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiolog. **117**, 500–544 (1952)
31. Jerri, A.J.: Introduction to Integral Equations with Applications. Monographs and Textbooks in Pure and Applied Mathematics, vol. 93. Marcel Dekker Inc. (1985)
32. Joly-Blanchard, G., Denis-Vidal, L.: Some remarks about an identifiability result of nonlinear systems. Automatica **34**(9), 1151–1152 (1998)
33. Keener, J., Sneyd, J.: Mathematical Physiology I: Cellular Physiology. Interdisciplinary Applied Mathematics, vol. 8/I, 2nd edn. Springer, Berlin (2010)
34. Kolchin, E.R.: Differential Algebra and Algebraic Groups. Academic, New York (1973)
35. Kostitzin, V.A.: Biologie Mathématique. Armand Colin (1937) (avec une préface de Vito Volterra)
36. Lange-Hegermann, M., Robertz, D.: Thomas decompositions of parametric nonlinear control systems. IFAC Proc. Vol. **46**(2), 296–301 (2013)
37. Ljung, L., Glad, S.T.: On global identifiability for arbitrary model parametrisations. Automatica **30**, 265–276 (1994)
38. Lubliner, J.: Plasticity Theory. Dover (2008)
39. Mansfield, E.L.: Differential Gröbner Bases. Ph.D. thesis, University of Sydney, Australia (1991)
40. Mboup, M., Join, C., Fliess, M.: Numerical differentiation with annihilator in noisy environment. Numer. Algorithm **50**(4), 439–467 (2009)
41. Noiret, C.: Utilisation du calcul formel pour l'identifiabilité de modèles paramétriques et nouveaux algorithmes en estimation de paramètres. Ph.D. thesis, Université de Technologie de Compiègne (2000)
42. Ollivier, F.: Le problème de l'identifiabilité structurelle globale : approche théorique, méthodes effectives et bornes de complexité. Ph.D. thesis, École Polytechnique, Palaiseau, France (1990)
43. Pavé, A.: Modeling Living Systems, from Cell to Ecosystem. ISTE/Wiley, New York (2012)
44. Picard, É.: La mathématique dans ses rapports avec la physique. In: Actes du IVème congrès international des mathématiques, Rome, Italie, 1908. Gauthier-Villars. 10 avril (1908)
45. Quadrat, A., Regensburger, G.: Polynomial solutions and annihilators of ordinary integro-differential operators. IFAC Proc. Vol. **46**(2), 308–313 (2013)

46. Rashevsky, N.: Mathematical Biophysics: Physico-Mathematical Foundations of Biology, vol. 1, 3rd edn.. Dover Publications Inc. (1960)
47. Reid, G.J., Wittkopf, A.D., Boulton, A.: Reduction of systems of nonlinear partial differential equations to simplified involutive forms. Eur. J. Appl. Math. **7**(6), 635–666 (1996)
48. Ritt, J.F.: Differential Algebra. American Mathematical Society Colloquium Publications, vol. 33. American Mathematical Society, New York (1950)
49. Robertz, D.: Formal Algorithmic Elimination for PDEs. Lecture Notes in Mathematics, vol. 2121. Springer, Berlin (2014)
50. Rosenkranz, M.: A new symbolic method for solving linear two-point boundary value problems on the level of operators. J. Symb. Comput. **39**(2), 171–199 (2005)
51. Rosenkranz, M., Gao, M., Guo, L.: An algebraic study of multivariable integration and linear substitution. Technical Report (2015). arXiv:1503.01694
52. Rosenkranz, M., Regensburger, G.: Integro-differential polynomials and operators. In: Jeffrey, D. (ed.) ISSAC'08: Proceedings of the 2008 International Symposium on Symbolic and Algebraic Computation. ACM Press (2008)
53. Rosenkranz, M., Regensburger, G.: Solving and factoring boundary problems for linear ordinary differential equations in differential algebras. J. Symb. Comput. **43**(8), 515–544 (2008)
54. Rota, G.-C.: Baxter algebras and combinatorial identities (I, II). Bull. Am. Math. Soc. **75**, 325–334 (1969)
55. Rota, G.-C.: Baxter operators, an introduction. In: Gian-Carlo Rota on Combinatorics, Introductory papers and Commentaries. Birkhäuser, Boston (1995)
56. Sedoglavic, A.: A probabilistic algorithm to test local algebraic observability in polynomial time. J. Symb. Comp. **33**(5), 735–755 (2002)
57. Seidenberg, A.: An elimination theory for differential algebra. Univ. California Publ. Math. (New Series) **3**, 31–65 (1956)
58. Verdière, N., Denis-Vidal, L., Joly-Blanchard, G.: A new method for estimating derivatives based on a distribution approach. Numer. Algorithms **61**, 163–186 (2012)
59. Verdière, Nathalie, Denis-Vidal, L., Joly-Blanchard, G., Domurado, D.: Identifiability and estimation of pharmacokinetic parameters for the ligands of the macrophage mannose receptor. Int. J. Appl. Math. Comput. Sci. **15**(4), 517–526 (2005)
60. Volterra, V.: Sur les équations intégro-différentielles et leurs applications. Acta Mathematica **35**(1), 295–356 (1912)
61. Volterra, V.: Leçons sur les équations intégrales et les équations intégro-différentielles. Gauthier-Villars, Paris, : Leçons professées à la faculté des sciences de Rome, publiées par M. Tomassetti et F.-S, Zarlatti (1913)
62. Volterra, V.: Applications des mathématiques à la biologie. L'enseignement mathématique. Leçon faite le 17 juin 1937, dans la série des Conférences internationales des Sciences mathématiques (1937)
63. Volterra, V.: Theory of Functionals and of Integral and Integro-Differential Equations. Dover Publications Inc. (1959). With a biography and a bibliography by Sir Edmund Whittaker

Chapter 7
Algebraic Estimation in Partial Derivatives Systems: Parameters and Differentiation Problems

Rosane Ushirobira, Anja Korporal and Wilfrid Perruquetti

Abstract Two goals are sought in this paper: namely, to provide a succinct overview on algebraic techniques for numerical differentiation and parameter estimation for linear systems and to present novel algebraic methods in the case of several variables. The state-of-art in the introduction is followed by a brief description of the methodology in the subsequent sections. Our new algebraic methods are illustrated by two examples in the multidimensional case. Some algebraic preliminaries are given in the appendix.

Keywords Parameter estimation · Numerical differentiation · Partial derivatives systems · Algebraic methods

7.1 Introduction

Many challenging questions in signal processing and control involve the estimation of derivatives of measured time signals, usually in noisy environment. This important issue is known as a numerical differentiation. Several approaches were proposed on this subject, based on different frameworks in applied mathematics and engineering. In control theory, designing a differentiator is an important problem, with many applications [2, 6, 17]. Some classical solutions are based on the least-squares polynomial interpolation and provide good offline results for this matter, see for example [15]. On another groundwork, just to mention a few works, numerical differentiators

R. Ushirobira (✉) · A. Korporal
Inria, Non-A team, 40 avenue Halley, 59650 Villeneuve d'Ascq, France
e-mail: Rosane.Ushirobira@inria.fr

A. Korporal
e-mail: Anja.Korporal@inria.fr

W. Perruquetti
École Centrale de Lille & CRIStAL (UMR CNRS 9189) & Université Lille Nord de France & Non-A, Inria, Villeneuve d'Ascq, France
e-mail: Wilfrid.Perruquetti@inria.fr

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_7

defined on an observer design basis were proposed in [4, 6, 13, 14, 41] and digital filter processing techniques used in [1, 3, 32, 37]. A high-order sliding mode based differentiator is designed in [8] by developing the results from [16] and it provides very satisfactory estimations despite some chattering in the response. We may also remark that the homogeneous finite-time-differentiator defined in [29] presents no chattering, but it is more sensitive to the signal amplitude.

The problem of estimating parameters in linear systems appears in the mathematical modeling of a physical phenomena. Differential equations in the considered model may contain parameters that are simply difficult to determine through data collecting, perhaps due to noisy measurements. This essential problem has attracted the attention of researchers in many fields. For instance, parameter estimation is a central subject in statics inference and several procedures can be applied to this problem, such as the maximum likelihood. Also, parameter estimation problems are also often related to optimization techniques.

There are countless works on numerical differentiation and parameter estimation. Among the recent advances on these issues, a promising solution is provided by differential algebra and operational calculus tools. This algebraic branch was initiated in the works by Fliess and Sira-Ramírez [10]. A clear description of the procedure, containing many useful explanations, can be found in [9, 19]. In despite of the innovative character of this framework, this algebraic approach remains quite underused. Nevertheless, some works do apply these ideas, see for instance [22, 42, 43] and for more practical developments, see for example [24, 27, 50]. For more details, the reader may refer to a quite extensive survey recently published in [39].

Algebraic methods within the numerical differentiation context were first applied to the univariable numerical differentiation by Mboup et al. in [20] where the authors use Jacobi projections to construct estimators for the derivatives. As described by the authors, the key idea of the method in this latter is to consider the nth-derivative of a smooth signal at a point τ as a single parameter to be estimated from a noisy observation of the signal. From that, a pointwise derivative can be estimated by varying τ. A truncated Taylor series expansion of the signal is the starting point in this technique, and the computations are then made in the operational domain. A slight drawback in the approximation by a truncated Taylor polynomial model may be its ephemeral character. To reduce this fast transient behavior, an improvement of the technique was proposed in [33]. A through study of these algebraic estimators, with emphasis on the error analysis, can be found in [18]. In that work, estimators based on fractional derivatives were introduced. An interesting computer architecture to accelerate the computation of the aforesaid algebraic derivative estimator was implemented in [28] using reconfigurable logic and implemented in an FPGA (field-programmable gate array).

In the multidimensional case, the estimation of derivatives of a noisy signal concern also many problems in engineering. For instance, in economy issues, in addition to the fields of signal processing or control. To tackle this numerical differential problem, most likely more problematic than the unidimensional case, several techniques were developed. The most commonly used is the finite differences method. The instability of possible solutions to these problems arise from the presence of noise due to the differentiation.

The use of algebraic tools for multidimensional numerical differentiation was addressed by Riachy et al. in [34–36]. Their inspiration comes from the original ideas in [10] and from the solutions proposed by [20].

The algebraic method developed in this paper is motivated by the parameter estimation methods elaborated in [45–47]. In those works, Weyl algebra based tools grant the estimation of amplitudes, signal and frequencies of a sinusoidal signal, providing faster estimates than known methods [5, 12, 48] (simulated examples in [12] provide fast estimates, however in more than a fraction of the period). The main advantage of our method is to give closed formulas for derivatives and parameter estimates. Furthermore, algebraic estimation techniques strongly rely on differential elimination. So, a number of different estimators (i.e. appropriate differential operators providing estimates) can be devised for a given estimation problem (this is well illustrated through a change-point detection problem in [21]). Hence, it appears that the quality of an estimator varies markedly with the order of the selected differential operators used in the elimination. The Weyl Algebra point of view introduced here within the algebraic context allows to characterize and to select the minimal order operators associated to any given estimation problem. Finally, let us stress that all algebraic approaches mentioned above in this Introduction share a very useful characteristic: obtained estimates are integrals of the noisy measured signal, so these integrals act as time varying filters.

Section 7.2 starts with a general introduction of the procedure of algebraic estimation, followed by the presentation of two estimation problems: numerical differentiation and parameter identification. In Sect. 7.4, the algebraic estimation of derivatives is illustrated through a significant example. To expose our method on a multidimensional parameter estimation problem, a particular partial differential equation was examined. It is the example of the heat conduction on a thin rod that is discussed and treated by algebraic estimations (this type of equation was considered in [38], also based on algebraic techniques). Proposed solutions to this problem in the algebraic framework are given in Sect. 7.5. The Appendix contains generalities on algebraic structures, as well as useful properties for the algebraic methods.

7.2 Problem Formulation

As mentioned in the Introduction, numerous engineering problems concern the estimation of state variables or parameters. In this section, we describe briefly how algebraic methods proceed, in general, to this estimation.

Most of the time, the mathematical modeling of physical phenomena provide a description of the aforementioned practical problems through a differential equations framework. States or parameters to be identified appear in the terms of these differential equations.

Roughly speaking, for such a given differential equations system, algebraic methods observe typically the following sequence of steps:

(a) Passage to the operational domain through the Laplace transform or by using Mikusinski operational calculus [25, 26, 49]: thanks to this step, differential equations are converted into algebraic ones, consequently allowing algebraic concepts to be applied. The resulting algebraic expressions depend on the Laplace variable s.
(b) Computations with the algebraic equations using structural properties: in this part, the aim is to apply algebraic tools on the equations in order to find expressions and closed formulas for the parameters or derivatives estimations.
(c) Return to the time domain and identification: the algebraic expressions in the Laplace variable s found in the previous step are converted into the time domain through the inverse Laplace transform. Possibly a time dependent system on the parameters must be solved.

It is notably in Step 2. That the advantages among different algebraic approaches can be seen. Indeed, in the differential elimination necessary in this process, structural properties of differential algebra are useful. Most of these procedures result in estimates given by integrals (rather than derivatives) of the noisy measured signal and these integrals will then provide noise attenuation. Thanks to special forms for the *annihilators* (differential operators involved in the differential elimination) developed in the appendix, the identification process presented in this paper will result in faster and less noise sensitive estimates. Therefore, these particular annihilators allow a better choice of suitable differential operators allowing the elimination of problematic parameters, or yield a more convenient matrix representation that will ease the solution of a system.

7.2.1 Derivative Estimation Problem

Throughout this paper, $\mathbb{K}$ denotes a field of characteristic zero (usually $\mathbb{R}$ or $\mathbb{C}$ in many applications). Let $\mathbf{x} = (x_1, \ldots, x_m)$ be an element in $\mathbb{K}^m$ ($m \in \mathbb{N}$). An m-tuple $N \in \mathbb{N}^m$ will be written as $N = (N_1, \ldots, N_m)$. We consider the partial order $\preceq$ on $\mathbb{N}^m$ defined by $N \preceq M$ if $N_i \leq M_i$, for all $1 \leq i \leq m$.

Let $f : U \subset \mathbb{R}^m \to \mathbb{R}$ be a multivariate signal where U is some neighborhood of 0. For a given $I = (i_1, \ldots, i_m) \in \mathbb{N}^m$, we denote $|I| := i_1 + \cdots + i_m$, $I! := i_1! \cdots i_m!$, $\mathbf{x}^I = x_1^{i_1} \ldots x_m^{i_m}$ and $\frac{\partial^I}{\partial \mathbf{x}^I} = \frac{\partial^{i_1}}{\partial x_1^{i_1}} \ldots \frac{\partial^{i_m}}{\partial x_m^{i_m}}$.

In practical problems, the available signal f is usually corrupted by a noise. Denote by f_ϖ the noisy multivariate signal

$$f_\varpi(\mathbf{x}) = f(\mathbf{x}) + \varpi(\mathbf{x}),$$

where $\varpi(\mathbf{x})$ is an additive noise. Assume that f admits a Taylor series expansion at 0 and write:

$$f(\mathbf{x}) = \sum_{I \in \mathbb{N}^m} \frac{a_I}{I!} \mathbf{x}^I, \quad \text{where } a_I = \frac{\partial^I f}{\partial \mathbf{x}^I}(0).$$

For $N = (N_1, \ldots, N_m) \in \mathbb{N}^m$, the truncated Taylor series f_N at order N is given by:

$$f_N(\mathbf{x}) = \sum_{I \preceq N} \frac{a_I}{I!} \mathbf{x}^I. \tag{7.1}$$

The multivariate Laplace transform of a function $g : \mathbb{R}^m \to \mathbb{R}$ is given by:

$$G(\mathbf{s}) := \mathscr{L}(g)(\mathbf{s}) = \int_{\mathbb{R}_+^m} g(\mathbf{t})\, \mathrm{e}^{-\mathbf{s}\,{}^T\mathbf{t}}\, d\mathbf{t}, \tag{7.2}$$

where $\mathbf{s} = (s_1, \ldots, s_m)$ is the Laplace (multi)variable and ${}^T\mathbf{t}$ denotes the transpose of $\mathbf{t} \in \mathbb{R}^m$. That implies, for instance:

$$\mathscr{L}\left(\frac{\mathbf{x}^I}{I!}\right) = \frac{1}{\mathbf{s}^{I+\mathbf{1}}}$$

where $\mathbf{s}^I = s_1^{i_1} \ldots s_m^{i_m}$. To realize $f_N(\mathbf{x})$ in the operational domain, we apply the Laplace transform (7.2) on (7.1). It results:

$$F_N(\mathbf{s}) = \sum_{I \preceq N} \frac{a_I}{\mathbf{s}^{I+1}}. \tag{7.3}$$

For $\mathbf{x}, \mathbf{t} \in \mathbb{K}^m$, we use the notation:

$$\int_{\mathbf{0}}^{\mathbf{x}} g(\mathbf{t}) d\mathbf{t} = \int_0^{x_1} \ldots \int_0^{x_m} g\,(t_1, \ldots, t_m)\, dt_1 \ldots dt_m.$$

Recall that for a multivariate function g and its Laplace transform G, the inverse Laplace transform satisfies

$$\mathscr{L}^{-1}\left(\frac{1}{\mathbf{s}^I} \frac{\partial^J G}{\partial \mathbf{s}^J}\right) = \frac{1}{(I-\mathbf{1})!} \int_{\mathbf{0}}^{\mathbf{x}} (\mathbf{x}-\tau)^{I-\mathbf{1}}\, (-\tau)^J g(\tau) d\tau \tag{7.4}$$

where $\boldsymbol{\emptyset} = (\tau_1, \ldots, \tau_m)$ and $\mathbf{1} = (1, \ldots, 1) \in \mathbb{K}^m$. For the sake of simplicity, we set:

$$v_{I,J} = v_{I,J}(\tau) = (\mathbf{x}-\tau)^I\, (-\tau)^J, \tag{7.5}$$

and a shorter notation can be used:

$$\mathscr{L}^{-1}\left(\frac{1}{\mathbf{s}^I} \frac{\partial^J G}{\partial \mathbf{s}^J}\right) = \frac{1}{(I-\mathbf{1})!} \int_{\mathbf{0}}^{\mathbf{x}} v_{I-1,J}(\tau) g(\tau) d\tau. \tag{7.6}$$

As we have seen in the introduction, a remarkable work on numerical differentiation by algebraic methods was written by Mboup et al. [20]. To illustrate their approach, we consider an example in the one-dimensional case. Consider the

approximating polynomial function of degree N of a real-valued signal $f(t)$, analytic on some time interval:

$$f(t) = \sum_{i=0}^{N} \frac{a_i}{i!} t^i, \tag{7.7}$$

originated from its Taylor series expansion, hence

$$a_i = f^{(i)}(0), \ \forall 0 \leq i \leq N.$$

The goal is to estimate the derivatives of the signal $f(t)$, that means, the coefficients a_i in (7.7). Often, $f(t)$ will be assumed to be the measured signal from a signal $x(t)$ with some negligible noise, so we may consider only $f(t)$. For example, the estimation of the first derivative of $f(t)$ can be obtained in the following way: from the degree one polynomial

$$f(t) = a_0 + a_1 t,$$

we obtain the operational domain expression given by the action of the Laplace transform. That yields:

$$Y(s) = \frac{a_0}{s} + \frac{a_1}{s^2}.$$

In [20], a *minimal* annihilator Π is proposed to eliminate the term a_0. It consists of a suitable differential operator, in this case:

$$\Pi = \frac{1}{s^2} \frac{d}{ds} s.$$

(meaning that the expression is multiplied by s, then taking the derivative with respect to s and finally multiplying by $\frac{1}{s^2}$). The term a_0 is eliminated after the action of Π. The time domain representation obtained thanks to the inverse Laplace transform provides an estimate $\widehat{\dot{f}}(t)$ of the first derivative $a_1 = \dot{f}(0)$:

$$\widehat{\dot{f}}(t) = \frac{6}{T^3} \int_0^T (T - 2\tau) \ f(t - \tau) d\tau.$$

(in practice, f is replaced by its measure). The idea presented here is to individually estimate each derivative a_J for $J \preceq N$. To formalize our procedure, we consider the following sets:

$$\Theta = \{a_I \mid I \preceq N\}, \ \Theta_{\text{est}} = \{a_J\} \text{ and } \Theta_{\overline{\text{est}}} = \Theta \setminus \Theta_{\text{est}}.$$

The definition of Θ, Θ_{est} and $\Theta_{\overline{\text{est}}}$ is clear: Θ contains all the parameters, Θ_{est} contains the parameters to be estimated and $\Theta_{\overline{\text{est}}}$ the remaining ones. The relation ($\mathscr{R}$) below follows from (7.3):

$$\mathscr{R} : P(\mathbf{s}) F_N(\mathbf{s}) + Q(\mathbf{s}) + \overline{Q}(\mathbf{s}) = 0 \tag{7.8}$$

where

$$P(\mathbf{s}) = \mathbf{s}^N,\ Q(\mathbf{s}) = -a_J \mathbf{s}^{N-J-1} \in \mathbb{K}_{\Theta_{\text{est}}}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right] \text{ and}$$

$$\overline{Q}(\mathbf{s}) = -\sum_{I \leq N, I \neq J} a_I \mathbf{s}^{N-I-1} \in \mathbb{K}_{\Theta_{\overline{\text{est}}}}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right] \tag{7.9}$$

By $\mathbb{K}_{\Theta_{\text{est}}}$ and $\mathbb{K}_{\Theta_{\overline{\text{est}}}}$, we denote respectively the algebraic extensions $\mathbb{K}_{\Theta_{\text{est}}} = \mathbb{K}(\Theta_{\text{est}})$ and $\mathbb{K}_{\Theta_{\overline{\text{est}}}} = \mathbb{K}(\Theta_{\overline{\text{est}}})$.

Based on the relation $\mathscr{R}$ (7.8), three polynomials P, Q and $\overline{Q}$ are defined taking into account the coefficients to be identified: P is the polynomial multiplying $F_N(\mathbf{s})$, Q contains the coefficient to be estimated, while $\overline{Q}$ is formed by all the remaining terms. To obtain an equation containing only known terms and a_J, the polynomial $\overline{Q}$ must be eliminated. That will provide a formula for the estimate of a_J.

To annihilate $\overline{Q}$, some particular differential operators must be chosen to act on ($\mathscr{R}$). These operators are called *annihilators*. Such algebraic estimators for a_J will be constructed by using structural properties of the Weyl algebra (see the Appendix).

Let us stress that if Π is an annihilator estimating a_J, the partial derivative of f at any other point $\mathbf{p} \in \mathbb{K}^m$ can be obtained by computing $\Pi\ (\mathscr{L}\ (f(\mathbf{x} + \mathbf{p})))$.

7.2.2 Parameter Estimation

An example of parametric identification was given in the seminal paper by Fliess and Sira-Ramírez [10] and it concerns a first order input-output system:

$$\dot{y}(t) = ay(t) + u(t) + \gamma_0$$

where a is a parameter to be identified and γ_0 is a constant perturbation. In the operational domain, thanks to the Laplace transform, the above equation becomes:

$$sY(s) - y(0) = aY(s) + U(s) + \frac{\gamma_0}{s}$$

where s is the Laplace variable, $Y(s)$ and $U(s)$ denote the Laplace transform of $y(t)$ and $u(t)$ respectively. The action of the differential operator $\frac{1}{s^2}\frac{d^2}{ds^2}s$ on this expression yields:

$$\left(\frac{1}{s}\frac{d}{ds}Y(s) + \frac{2}{s^2}\frac{d^2}{ds^2}Y(s)\right) a = \frac{d^2}{ds^2}Y(s) + \frac{4}{s}\frac{d}{ds}Y(s) + \frac{2}{s^2}Y(s) - \left(\frac{1}{s}\frac{d^2}{ds^2}U(s) + \frac{2}{s^2}\frac{d}{ds}U(s)\right).$$

Operational calculus rules yield the following estimation for a:

$$a = \frac{\int_0^1 \left(\left(2\,\nu^2 - 3\,\nu + 1\right)\nu\, t\, u\,(t\nu) + \left(-6\,\nu^2 + 6\,\nu - 1\right) y\,(t\nu)\right)\, d\nu}{t \int_0^1 \nu\, \left(2\,\nu^2 - 3\,\nu + 1\right)\, y\,(t\nu)\, d\nu}$$

The parameter identification for a partial differential equation can be thought in a similar way. To illustrate this, a simple example of the one-dimensional heat equation is studied in the sequel. A similar algebraic approach was studied, for instance, in [38] for this same problem and in [11] for the parameter identification of a linear model of the planar motion of a heavy rope. The unidimensional Laplace transform was used in both these examples providing operational functions a as the solutions of an initial value problem. In this work, the Laplace transform in two variables is used to convert the partial differential equation into the operational domain representation.

7.3 Annihilators via the Weyl Algebra

In the previous section, we have indicated that our aim is to annihilate the polynomial $\overline{Q}$ in the relation $\mathscr{R}$ (7.8), containing undesired parameters, see (7.9). That will be done by the action of *annihilators*: these are differential operators (or polynomials in the variables $\frac{\partial^l}{\partial \mathbf{s}^l}$) with polynomial coefficients (or rational functions) in the Laplace variables $s_1, \ldots, s_m$. A practical realization of differential operators acting on polynomial variables is the Weyl algebra. So, this algebra appears naturally in this context and its structural properties will be quite useful in the choice of the annihilators.

This algebraic viewpoint is inspired by the work of Fliess et al. [9, 10, 19]. Details about the algebraic notions defined in the sequel can be found in the appendix and in [23, 31] as well.

Next, we keep the notation defined in the appendix (see Sect. 7.7) to define the differential operators *annihilators*. They will help to construct algebraic estimators, either of derivatives or partial derivatives, or also of parameters.

Recall that A_m denotes the Weyl algebra $\mathsf{A}_m = \mathbb{K}[\mathbf{s}]\left[\frac{\partial}{\partial \mathbf{s}}\right]$ and $\mathsf{B}_m = \mathbb{K}\,(\mathbf{s})\left[\frac{\partial}{\partial \mathbf{s}}\right]$, respectively the polynomial rings in $\frac{\partial}{\partial \mathbf{s}}$ with coefficients in the polynomial ring $\mathbb{K}[\mathbf{s}]$ and in the fraction field $\mathbb{K}\,(\mathbf{s})$.

Definition 1 Let $R \in \mathbb{K}_{\Theta_{\text{est}}}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right]$. A *$R$-annihilator with respect to* B_m is an element of $\text{Ann}_{\mathsf{B}_m}(R) = \{F \in \mathsf{B}_m \mid F\,(R) = 0\}$.

Consider $m \geq 2$. Let us remark that $\text{Ann}_{\mathsf{B}_m}(R)$ is a left ideal of B_m. Therefore, by Stafford's theorem (Theorem 1, Appendix), $\text{Ann}_{\mathsf{B}_m}(R)$ is generated by two generators Π_1 and $\Pi_2 \in \mathsf{B}_m$:

$$\text{Ann}_{\mathsf{B}_m}(R) = \mathsf{B}_m \Pi_1 + \mathsf{B}_m \Pi_2.$$

We call the annihilators Π_1 and Π_2 *minimal R-annihilators with respect to* B_m. The attribute *minimal* comes from the order of the differential operators. Notice that

$\mathrm{Ann}_{\mathbf{B}_m}(R)$ contains annihilators in finite integral form, i.e. operators with coefficients in $\mathbb{K}\left[\frac{1}{\mathbf{s}}\right]$.

Let us stress that thanks to the above Stafford's theorem, only two generators for the ideal are needed for a given $m \geq 2$.

Lemma 1 *Consider* $R(\mathbf{s}) = \alpha \mathbf{s}^N = \alpha s_1^{n_1} \ldots s_m^{n_m}$, $N = (n_1, \ldots, n_m) \in \mathbb{Z}^m$ *with* $\alpha \in \mathbb{K}_{\Theta_{\overline{est}}}$. *A minimal* R*-annihilator is given by*

$$s_i \frac{\partial}{\partial s_i} - n_i, \ \forall 1 \leq i \leq m.$$

Recall that the *degree* of a monomial $\mathbf{s}^I \in \mathbb{K}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right]$ is $|I|$. The *total degree* of a polynomial in $\mathbf{s}$ is the maximum degree of its monomials.

Remark 1 Consider $R \in \mathbb{K}_{\Theta_{\overline{\text{est}}}}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right]$ with a monomial s^I of maximal degree. So R has total degree $|I|$. Let $i_k = \max\{i_j \mid j = 1, \ldots, m\}$. If $|I| > 0$, then $\frac{\partial^{i_k+1}}{\partial s_k^{i_k+1}}$ is clearly an R-annihilator.

Now, recall that the polynomial to be annihilated in this differentiation problem is (see (7.9)):

$$\overline{Q}(\mathbf{s}) = -\sum_{I \preceq N, I \neq J} a_I \mathbf{s}^{N-I-1} \in \mathbb{K}_{\Theta_{\overline{\text{est}}}}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right].$$

By the previous remark, it results immediately:

Lemma 2 *The differential operators* $\frac{\partial^{n_k}}{\partial s_k^{n_k}}$ *are* $\overline{Q}$*-annihilators, for all* $1 \leq k \leq m$.

To construct an alternative annihilator, an algorithm is sketched below:

Algorithm 4

Input: A polynomial $R = \sum_{\substack{I \in \mathbb{Z}^m \\ \text{finite}}} b_I \mathbf{s}^I$ in $\mathbb{K}_{\Theta_{\overline{\text{est}}}}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right]$ of total degree $d \in \mathbb{N}$

Output: An R-annihilator

(a) Set $\Pi = 1 \in D$.
(b) Choose a monomial of degree d in R, say $b_J \mathbf{s}^J$ with $J = (j_1, \ldots, j_m)$ (so $|J| = d$).
(c) Choose $j_k = \min\{j_\ell > 0 \mid \ell = 1, \ldots, m\}$.
(d) Apply $\pi = s_k \frac{\partial}{\partial s_k} - j_k$ (see Lemma 1) on R.
(e) (a) If $\pi(R) = 0$, then return π and stop the algorithm.
    (b) If $\pi(R) \neq 0$, then set $\Pi = \pi \circ \Pi$ and return to step (2) with $R \leftarrow \pi(R)$.

Example 1 Consider $m = 2$ and the polynomial $R(s_1, s_2) = a_{00} s_1^2 s_2 + a_{01} s_1^2 + a_{10} s_1 s_2 + a_{20} s_2 \in \mathbb{K}_{\Theta_{\overline{\text{est}}}}[s_1, s_2]$. A R-annihilator constructed with the above algorithm is $\left(s_1 \frac{\partial}{\partial s_1} - 2\right) \circ \left(s_2 \frac{\partial}{\partial s_2} - 1\right)$.

The concept of an estimator must be defined in order to take into account the remaining terms in the relation $(\mathscr{R})$, after the action of an annihilator. Notice that the parameters to be estimated might appear in the set of coefficients of both polynomials Q and P, but they might as well be present exclusively in one of the two. Therefore a $\overline{Q}$-annihilator must not eliminate all terms in Θ_{est}, as formalized in the definition below:

Definition 2 An *estimator* $\pi \in \mathsf{B}$ is a $\overline{Q}$-annihilator satisfying

$$\text{coeffs}\,(\pi(\mathscr{R})) \cap \mathbb{K}_{\Theta_{\text{est}}} \neq \emptyset,$$

where coeffs (R) denotes the set of coefficients of a polynomial $R \in \mathbb{K}_{\Theta}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right]$.

It is implied by this definition that the criterion on the coefficients must be considered in the choice of annihilators in the Algorithm 4.

It is important to stress that in some cases, it may be interesting to adopt another way of proceeding. For instance, several different annihilators can be constructed for each $\overline{Q}$. Then, Stafford's theorem can be applied to provide two minimal generators by using the package `Stafford` [30, 31]. The final step is to observe the criterion in Definition 2 in these generators to obtain an estimator.

7.4 Derivative Estimation

To illustrate the algebraic method for numerical differentiation, we present here the estimation for a derivative in the two-dimensional case. Hence, the Eq. (7.1) is considered for $m = 2$. For this example, we assume that the parameter to be estimated is $a_{21} = \frac{\partial^3 f}{\partial x_1^2 \partial x_2}(0, 0)$. Based on (7.1), a truncated Taylor series at $N = (2, 1)$ will then be used:

$$f(x_1, x_2) = a_{00} + a_{10}x_1 + a_{01}x_2 + a_{20}x_1^2 + a_{11}x_1x_2 + a_{21}x_1^2x_2.$$

The coefficient to be estimated is a_{21} in the truncated Taylor series above, so we may distinguish the following polynomials P, Q and $\overline{Q} \in \mathbb{K}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right]$ in the relation $(\mathscr{R})$ (see (7.8)):

$$\begin{aligned}
P(s_1, s_2) &= s_1^2 s_2, \\
Q(s_1, s_2) &= -a_{21} s_1^{-1} s_2^{-1} \in \mathbb{K}_{\Theta_{\text{est}}}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right] \text{ and} \\
\overline{Q}(s_1, s_2) &= -\sum_{\substack{(i,j) \preceq N \\ (i,j) \neq (2,1)}} a_{ij} s_1^{1-i} s_2^{-j} \in \mathbb{K}_{\Theta_{\overline{\text{est}}}}\left[\mathbf{s}, \frac{1}{\mathbf{s}}\right].
\end{aligned}$$

The first step of the estimation is to determine minimal $\overline{Q}$-annihilators. To begin, Lemma 1 helps to find two $\overline{Q}$-annihilators $\frac{\partial^2}{\partial s_1^2}\left(s_1\frac{\partial}{\partial s_1}+1\right)$ and $\frac{\partial}{\partial s_2}\left(s_2\frac{\partial}{\partial s_2}+1\right)$. However, they are not estimators since they clearly eliminate Q as well and Q is the only term in $\mathscr{R}$ with coefficients in $\mathbb{K}_{\Theta_{\text{est}}}$ (see Definition 2).

We then follow the Algorithm 4 to determine an alternative $\overline{Q}$-annihilator that may also be a estimator. In the case of a_{21}, we obtain:

$$\Pi = \frac{1}{s_1^2 s_2}\frac{\partial^2}{\partial s_1 s_2}\left(s_1\frac{\partial}{\partial s_1}-1\right).$$

Let us remark that for other coefficients $a_{k\ell}$, some annihilators are proposed in [36] and in [44] as well. Using Remark 4 in the Appendix, it can be shown that the annihilator Π is a minimal annihilator.

The action of Π on the relation $\mathscr{R}$ with P, Q and $\overline{Q}$ defined above, provides the following expression:

$$2\frac{F(s_1,s_2)}{{s_1}^3{s_2}^2}+4\frac{\frac{\partial}{\partial s_1}F(s_1,s_2)}{{s_1}^2{s_2}^2}+2\frac{\frac{\partial}{\partial s_2}F(s_1,s_2)}{{s_1}^3 s_2}+4\frac{\frac{\partial^2}{\partial s_2\partial s_1}F(s_1,s_2)}{{s_1}^2 s_2}+\frac{\frac{\partial^2}{\partial {s_1}^2}F(s_1,s_2)}{s_1{s_2}^2}$$
$$+\frac{\frac{\partial^3}{\partial s_2\partial {s_1}^2}F(s_1,s_2)}{s_1 s_2}+2\frac{a_{2,1}}{{s_1}^6{s_2}^4}=0.$$

Isolating the term with a_{21} and applying the inverse Laplace transform (7.6) provides the consequent estimate:

$$a_{21}=-\frac{360}{x_1^5x_2^3}\int_{\mathbf{0}}^{(x_1,x_2)}\left(v_{2,1,0,0}+v_{0,1,2,0}-4v_{1,1,0,0}-4v_{1,0,0,1}+v_{2,0,0,1}+v_{0,0,2,1}\right)f(\tau)d\tau,$$

where we use the notation (7.5):

$$v_{I,J}=(x_1-\tau)^{i_1}(x_2-\eta)^{i_2}(-\tau)^{j_1}(-\eta)^{j_2},$$

for all $I=(i_1,i_2)$, $J=(j_1,j_2)\in\mathbb{N}^2$.

7.5 Parameter Estimation

In the previous section, we examined the case of numerical differentiation where annihilators were used to eliminate the undesired terms of the truncated Taylor series seen in the operational domain. Moreover, as we have seen in the introduction, a similar procedure may provide estimates for parameters in an ordinary differential equation.

In this section, we present a parameter identification problem for a two-dimensional partial differential equation. The following classical example was studied in [38, 44],

for instance. Consider the problem of the heat conduction in a thin rod of length 1. Let $w : (z, t) \mapsto w(z, t)$ be the function representing the temperature at position z at time t. The partial differential equation describing this problem is given by:

$$\frac{\partial^2}{\partial z^2} w(z, t) - \beta \frac{\partial}{\partial t} w(z, t) - \alpha w(z, t) = 0 \tag{7.10}$$

The rod is assumed to be perfectly isolated at $z = 0$, so $\frac{\partial}{\partial z} w(0, t) = 0$. The condition at $z = 1$ is not of interest to us and we assume that the initial temperature is 0. In addition, we suppose that the temperature $w(z, t)$ at any time t and position z at the rod can be measured and used in the parameter estimation. To simplify the notation, we write $q_0 : t \mapsto w(0, t)$ and $f : t \mapsto w(z, 0)$.

The goal is to identify the parameters α and β. The algebraic method used in the previous subsection is applied here. Using the notation $r : t \mapsto \frac{\partial}{\partial z} w(0, t)$, the Laplace transform (7.2) is employed to realize the partial differential equation (7.10) as an algebraic equation in the Laplace variable $\mathbf{s} = (s_1, s_2)$ (s_1 corresponds to z and s_2 corresponds to t):

$$\left(s_1^2 - \beta s_2 - \alpha\right) W(s_1, s_2) + \beta F(s_1) - s_1 Q_0(s_2) - R(s_2) = 0, \tag{7.11}$$

where $W(\mathbf{s})$, $F(s_1)$, $Q_0(s_2)$ and $R(s_2)$ denote the Laplace transforms of $w(z, t)$ with respect to z and t, of $f(z)$ with respect to z, and of $q_0(t)$ and $r(t)$ with respect to t respectively. Since by hypothesis, $R \equiv 0$ and $F \equiv 0$, the Eq. (7.11) leads to:

$$\left(-\beta s_2 + {s_1}^2 - \alpha\right) W(s_1, s_2) - s_1 Q_0(s_2) = 0. \tag{7.12}$$

Here the set of parameters Θ_{est} to be estimated is

$$\Theta_{\text{est}} = \{\alpha, \beta\},$$

while $\Theta_{\overline{\text{est}}} = \emptyset$. Following the procedure described at the beginning of Sect. 7.2, a system on the indeterminates α and β will be determined by acting suitable annihilators on (7.12). So a two-steps procedure will be applied. In the first step, we rewrite (7.12) in the form of a $\mathscr{R}$-relation (see (7.8)):

$$\mathscr{R} : P(\mathbf{s}) W(\mathbf{s}) + Q(\mathbf{s}) + \overline{Q}(\mathbf{s}) = 0, \tag{7.13}$$

where

$$P(\mathbf{s}) = s_1^2 - \beta s_2 - \alpha, \;\; Q(\mathbf{s}) = -s_1 Q(s_2) \text{ and } \overline{Q}(\mathbf{s}) = 0. \tag{7.14}$$

As mentioned before, the problem of annihilating $\overline{Q}$ is tackled by finding suitable $\overline{Q}$-annihilators that will lead to a system in Θ_{est}. Notice in this example that the parameters to be estimated also appear in the coefficients of the polynomial P.

Here, the polynomial $\overline{Q}$ to be annihilated in (7.14) is 0, so we may consider the above Eq. (7.14). In order to apply the inverse Laplace transform (7.4), we divide this equation by $s_1^3 s_2^2$. Using the notation $v_{I,J} = (z-\tau)^{i_1}\,(t-\eta)^{i_2}\,(-\tau)^{j_1}(-\eta)^{j_2}$, for all $I = (i_1, i_2)$, $J = (j_1, j_2) \in \mathbb{N}^2$, we obtain in the spatial domain:

$$A_{11}\alpha + A_{12}\beta = B_1$$

where

$$\begin{cases} A_{11} = -\dfrac{1}{2}\displaystyle\int_0^{(z,t)} v_{2,1,0,0}\; w\,(\tau,\eta)\;\mathrm{d}\tau\,\mathrm{d}\eta, \\ A_{12} = -\dfrac{1}{2}\displaystyle\int_0^{(z,t)} v_{2,0,0,0}\; w\,(\tau,\eta)\;\mathrm{d}\tau\,\mathrm{d}\eta, \\ B_1 = -\displaystyle\int_0^{(z,t)} v_{0,1,0,0}\,(w\,(\tau,\eta) - q_0\,(\eta))\;\mathrm{d}\tau\;\mathrm{d}\eta. \end{cases}$$

In the second step, we will try to eliminate the term with the polynomial Q_0: in this case, the polynomial $\overline{Q}$ to be annihilated in (7.14) is $\overline{Q}(\mathbf{s}) = -s_1 Q_0\,(s_2)$ while $Q(\mathbf{s}) = \beta\,F\,(s_1)$. We propose the $\overline{Q}$-annihilator $\pi = \frac{\partial^2}{\partial s_1^2}$. Applying π on the relation (7.13) gives:

$$-\alpha\frac{\partial^2}{\partial {s_1}^2}W + \left(\frac{\mathrm{d}^2}{\mathrm{d}{s_1}^2}F\,(s_1) - s_2\frac{\partial^2}{\partial {s_1}^2}W\right)\beta + {s_1}^2\frac{\partial^2}{\partial {s_1}^2}W + 4\,s_1\frac{\partial}{\partial s_1}W + 2\,W = 0. \tag{7.15}$$

After dividing the above equation by a suitable monomial in $\mathbf{s}$, namely $s_1^2 s_2^2$, we obtain:

$$\begin{cases} A_{21} = -\dfrac{1}{2}\displaystyle\int_0^{(z,t)} v_{2,1,0,2}\; w\,(\tau,\eta)\;\mathrm{d}\tau\,\mathrm{d}\eta \\ A_{22} = -\dfrac{1}{2}\displaystyle\int_0^{(z,t)} v_{2,0,0,2} w\,(\tau,\eta)\;\mathrm{d}\tau\,\mathrm{d}\eta, \\ B_2 = \displaystyle\int_0^{(z,t)} \left(4v_{1,1,0,0} - v_{2,1,0,0} - v_{0,1,2,0}\right)\; w\,(\tau,\eta)\;\mathrm{d}\tau\,\mathrm{d}\eta. \end{cases}$$

A system on Θ_{est} results from the actions of $\overline{Q}$-annihilators:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}.$$

Solving the system provides the estimates of α and β.

Remark 2 In [44], other annihilators were proposed since the statement of the problem and its initial and boundary conditions were different. Again, recall that a very special property of the two-dimensional Weyl algebra is Stafford's theorem (see Theorem 1). This important result allows the computation of two minimal annihilators. Moreover the package `Stafford` [30, 31] uses a highly efficient algorithm to calculate these differential operators.

7.6 Conclusion

In this paper, we provided a short preview on algebraic estimation for derivatives and for parameters in linear systems. Advantages and possible drawbacks of this algebraic framework were evoked in a brief state-of-art. More detailed problem statements were given in the subsequent sections, followed by proposed solutions within the algebraic context. The algebraic properties in the appendix, notably concerning the Weyl algebra, support these solutions. Furthermore, we illustrate our algebraic method with two typical examples: in the case of two-dimensional numerical differentiation, while in the case of parameter estimation for partial derivatives systems, the thin rod example is studied. An essential point deserves to be emphasized: the algebraically obtained estimated are based on integrals of measured signals. These particular integrals play the role of time-varying filters. Furthermore, closed formulas for derivatives and parameters estimates that obtained with our method, via the Weyl algebra tools, are presented in this paper.

7.7 Appendix

We recall below some basic definitions and properties of the Weyl algebra.

Definition 3 Let $m \in \mathbb{N} \setminus \{0\}$. The *Weyl algebra* $\mathsf{A}_m(\mathbb{K})$ (or A_m) is the $\mathbb{K}$-algebra with generators $p_1, q_1, \ldots, p_m, q_m$ and relations

$$[p_i, q_j] = \delta_{ij}, \ [p_i, p_j] = [q_i, q_j] = 0, \forall\ 1 \leq i, j \leq m$$

where $[\cdot, \cdot]$ is the commutator defined by $[u, v] := uv - vu$, for all $u, v \in \mathsf{A}_m(\mathbb{K})$.

The Weyl algebra A_m can be realized as the algebra of A polynomial differential operators on the polynomial ring $\mathbb{K}[\mathbf{s}]$ by setting:

$$p_i = \frac{\partial}{\partial s_i} \text{ and } q_i = s_i \times \cdot\ , \forall\ 1 \leq i \leq m,$$

where $\times$ denotes the multiplication map. That implies that A_m can be written as $\mathsf{A}_m = \mathbb{K}[\mathbf{q}][\mathbf{p}] = \mathbb{K}[\mathbf{s}]\left[\frac{\partial}{\partial \mathbf{s}}\right]$. The algebra of differential operators $\mathsf{B}_m(\mathbb{K})$ (or B_m) on $\mathbb{K}[\mathbf{s}]$ with coefficients in the rational functions field $\mathbb{K}(\mathbf{s})$ is naturally related to $\mathsf{A}_m(\mathbb{K})$. In this case, we can write $\mathsf{B}_m := \mathbb{K}(\mathbf{q})[\mathbf{p}] = \mathbb{K}(\mathbf{s})\left[\dfrac{\partial}{\partial \mathbf{s}}\right]$.

A $\mathbb{K}$-basis for A_m is given by $\left\{\mathbf{q}^I\mathbf{p}^J \mid I, J \in \mathbb{N}^m\right\}$ where $\mathbf{q} = q_1^{i_1}\dots q_m^{i_m}$ and $\mathbf{p} = p_1^{i_1}\dots p_m^{i_m}$. An operator $F \in \mathsf{A}_m$ can be written in a *canonical form*,

$$F = \sum_{I,J} \lambda_{IJ}\mathbf{q}^I\mathbf{p}^J \quad \text{with} \quad \lambda_{IJ} \in \mathbb{K}.$$

Similarly, an element $F \in \mathsf{B}_m$ can be written as

$$F = \sum_I g_I(\mathbf{s})\frac{\partial^I}{\partial s^I}, \quad \text{where} \quad g_I(\mathbf{s}) \in \mathbb{K}(\mathbf{s}).$$

The *order* of F is defined as $\mathrm{ord}(F) = \max\{[I| \mid g_I(\mathbf{s}) \neq 0\}$. This definition holds for the Weyl algebra A_m as well, since $\mathsf{A}_m \subset \mathsf{B}_m$. Some useful properties of A_m and B_m are given by the following propositions (see for instance [7]):

Proposition 1 *The algebra A_m is a domain. Moreover, A_m is a simple algebra (*i.e. *it contains no nontrivial ideals) and also a left Noetherian ring (*i.e. *every left ideal is finitely generated).*

These properties are shared by B_m. In addition, A_m is neither a principal right domain, nor a principal left domain. Nevertheless this is true for B_1:

Proposition 2 *B_1 admits a left division algorithm, that is, if $F, G \in \mathsf{B}_1$, then there exists $Q, R \in \mathsf{B}_1$ such that $F = QG + R$ and $\mathrm{ord}(R) < \mathrm{ord}(G)$. Consequently, B_1 is a principal left domain.*

Alas, this proposition does not hold for B_m for $m \geq 2$. But an important theorem by T. Stafford (see [40]) provides an remarkable property on the number of generators of a left ideal in the Weyl algebra. Namely, Stafford proved that every left ideal of D ($D = \mathsf{B}_m$ or B_m) can be generated by two elements in D:

Theorem 1 (Stafford) *Let $\mathfrak{a}$ be a left ideal of D generated by three elements F_1, F_2 and $F_3 \in D$. Then, there exist G_1 and $G_2 \in D$ such that*

$$\mathfrak{a} = D(F_1 + G_1F_3) + D(F_2 + G_2F_3).$$

An effective implementation in `Maple`, named `Stafford`, of this important theorem can be found in the work of Quadrat and Robertz [30].

Remark 3 It is important to notice that the principality of B_1 was largely used in the initial works on algebraic methods applied to univariate numerical differentiation, such as [20] or parameter estimation in ordinary differential equations, see for

instance [47]. In the multivariate case, the principality holds no longer, therefore the importance of Stafford's theorem.

To close this part, we remark a useful identity:

Remark 4 For arbitrary $N, M \in \mathbb{N}^r$, we have

$$\frac{\partial^N}{\partial \mathbf{s}^N}\frac{1}{\mathbf{s}^M} = \sum_{0 \preceq J \preceq N} \binom{N}{J} (-1)^{|N-J|} \frac{M^{\overline{N-J}}}{\mathbf{s}^{M+N-J}} \frac{\partial^J}{\partial \mathbf{s}^J},$$

where $\binom{N}{J} = \binom{n_1}{j_1} \dots \binom{n_r}{j_r}$, $M^{\overline{N}} = m_1^{\overline{n_1}} \dots m_r^{\overline{n_r}}$ and $m_i^{\overline{n_i}}$ denotes the rising factorial ($m_i^{\overline{n_i}} = m_i(m_i+1)\dots(m_i+n_i-1)$).

References

1. Al-Alaoui, M.: A class of second-order integrators and low-pass differentiators. IEEE Trans. Circuits Syst. I: Fundam. Theory Appl. **42**(4), 220–223 (1995)
2. Carlsson, B., Ahln, A., Sternad, M.: Optimal differentiation based on stochastic signal models. IEEE Trans. Signal Process. **39**, 341–353 (1991)
3. Chen, C.-K., Lee, J.-H.: Design of high-order digital differentiators using L_1 error criteria. IEEE Trans. Circuits Syst. II: Analog. Digit. Signal Process. **42**(4), 287–291 (1995)
4. Chitour, Y.: Time-varying high-gain observers for numerical differentiation. IEEE Trans. Autom. Control **47**(9), 1565–1569 (2002)
5. Coluccio, L., Eisinberg, A., Fedele, G.: A property of the elementary symmetric functions on the frequencies of sinusoidal signals. Signal Process. **89**(5), 765–777 (2009)
6. Dabroom, A.M., Khalil, H.K.: Discrete-time implementation of high-gain observers for numerical differentiation. Int. J. Control. **72**(17), 1523–1537 (1999)
7. Dixmier, J.: Algèbres enveloppantes. Gauthier-Villars (1974)
8. Efimov, D., Fridman, L.: A hybrid robust non-homogeneous finite-time differentiator. IEEE Trans. Autom. Control **56**(5), 1213–1219 (2011)
9. Fliess, M., Mboup, M., Mounier, H., Sira-Ramírez, H.: Questioning some paradigms of signal processing via concrete examples. In: Sira-Ramírez, G.S.-N.H. (ed.) Algebraic Methods in Flatness, Signal Processing and State Estimation. Editiorial Lagares, pp. 1–21 (2003)
10. Fliess, M., Sira-Ramírez, H.: An algebraic framework for linear identification. ESAIM Control Optim. Calc. Variat. **9**, 151–168 (2003)
11. Gehring, N., Knppel, T., Rudolph, J., Woittennek, F.: Algebraic identification of heavy rope parameters. In: Proceedings of the 16th IFAC Symposium on System Identification, pp. 161–166 (2012)
12. Hou, M.: Parameter identification of sinusoids. IEEE Trans. Autom. Control **57**(2), 467–472 (2012)
13. Ibrir, S.: On observer design for nonlinear systems. Intern. J. Syst. Sci. **37**(15), 1097–1109 (2006)
14. Ibrir, S.: Linear time-derivative trackers. Automatica **40**(3), 397–405 (2004)
15. Ibrir, S., Diop, S.: A numerical procedure for filtering and efficient high-order signal differentiation. Int. J. Appl. Math. Comput. Sci. **14**(2), 201–208 (2004)
16. Levant, A.: Robust exact differentiation via sliding mode technique*. Automatica **34**(3), 379–384 (1998)

17. Levant, A.: Higher-order sliding modes, differentiation and output-feedback control. Int. J. Control. **76**(9–10), 924–941 (2003)
18. Liu, D.: Analyse d'erreurs d'estimateurs des dérivées de signaux bruités et applications. Ph.D. Dissertation, University of Lille 1, France (2011)
19. Mboup, M.: Parameter estimation for signals described by differential equations. Appl. Anal. **88**, 29–52 (2009)
20. Mboup, M., Join, C., Fliess, M.: Numerical differentiation with annihilators in noisy environement. Numer. Algorithms **50**, 439–467 (2009)
21. Mboup, M.: Parameter estimation via differential algebra and operational culculus. Research Report (2007)
22. Mboup, M.: Neural spike detection and localisation via Volterra filtering. In: 22nd IEEE Workshop on Machine Learning for Signal Processing. Santander (2012)
23. McConnell, J., Robson, J.: *Noncommutative Noetherian Rings*. A. M. Soc., Ed. Hermann (2000)
24. Menhour, L., d'Andrea Novel, B., Boussard, C., Fliess, M., Mounier, H.: Algebraic nonlinear estimation and flatness-based lateral/longitudinal control for automotive vehicles. In: Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference, pp. 463–468 (2011)
25. Mikusiński, J., Boehme, T.: Operational Calculus. Series International Series of Monographs in Pure and Applied Mathematics. Pergamon Press (1987)
26. Mikusiński, J.: Operational Calculus. Pergamon, Oxford (1959)
27. Morales, R., Somolinos, J., Sira-Ramírez, H.: Control of a DC motor using algebraic derivative estimation with real time experiments. Measurement **47**, 401–417 (2014)
28. Morales, R., Rincón, F., Gazzano, J.D., Lopez, J.C.: Real-time algebraic derivative estimations using a novel low-cost architecture based on reconfigurable logic. Sensors **14**(5), 9349–9368 (2014)
29. Perruquetti, W., Floquet, T., Moulay, E.: Finite-time observers: application to secure communication. IEEE Trans. Autom. Control **53**(1), 356–360 (2008)
30. Quadrat, A., Robertz, D.: Computation of bases of free modules over the Weyl algebras. J. Symbolic Comput. **42**, 1113–1141 (2007)
31. Quadrat, A.: An Introduction to Constructive Algebraic Analysis and its Applications. INRIA, Research Report RR-7354. (2010). https://hal.inria.fr/inria-00506104
32. Rader, C., Jackson, L.: Approximating noncausal IIR digital filters having arbitrary poles, including new Hilbert transformer designs, via forward/backward block recursion. IEEE Trans. Circuits Syst. I: Regul. Pap. **53**(12), 2779–2787 (2006)
33. Reger, J., Mai, P., Sira-Ramirez, H.: Robust algebraic state estimation of chaotic systems. In: Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control, 2006 IEEE, pp. 326–331 (2006)
34. Riachy, S., Bachalany, Y., Mboup, M., Richard, J.-P.: An algebraic method for multi-dimensional derivative estimation. In: MED'08, 16th IEEE Mediterranean Conference on Control and Automation. Ajaccio (2008)
35. Riachy, S., Bachalany, Y., Mboup, M., Richard, J.-P.: Différenciation numérique multivariable I : estimateurs algébriques et structure. In: Sixième Conférence Internationale Francophone d'Automatique Nancy, France, 2-4 juin 2010. Nancy (2010)
36. Riachy, S., Mboup, M., Richard, J.-P.: Multivariate numerical differentiation. J. Comput. Appl. Math. **236**(6), 1069–1089 (2011)
37. Roberts, R.A., Mullis, C.T.: Digital Signal Processing. Addison-Wesley, Reading (1987)
38. Rudolph, J., Woittennek, F.: An algebraic approach to parameter identification in linear infinite dimensional systems. In: Proceedings of the 16th Mediterranean Conference on Control and Automation, pp. 332–337 (2008)
39. Sira-Ramírez, H., Rodríguez, C.G., Romero, J.C., Juárez, A.L.: Algebraic Identification and Estimation Methods in Feedback Control Systems, 1st edn. Wiley Publishing, Hoboken (2014)
40. Stafford, J.T.: Module structure of Weyl algebras. J. Lond. Math. Soc. **18**(3), 429–442 (1978)
41. Su, Y., Zheng, C., Mueller, P., Duan, B.: A simple improved velocity estimation for low-speed regions based on position measurements only. IEEE Trans. Control. Syst. Technol. **14**(5), 937–942 (2006)

42. Trapero, J., Sira-Ramírez, H., Battle, V.: A fast on-line frequency estimator of lightly damped vibrations in flexible structures. J. Sound Vib. **307**, 365–378 (2007)
43. Trapero-Arenas, J., Mboup, M., Pereira-Gonzalez, E., Feliu, V.: On-line frequency and damping estimation in a single-link flexible manipulator based on algebraic identification. In: 16th Mediterranean Conference on Control and Automation, pp. 338–343 (2008)
44. Ushirobira, R., Korporal, A., Perruquetti, W.: On an algebraic method for derivatives estimation and parameter estimation for partial derivatives systems. In: Mathematical Theory of Networks and Systems. Netherlands (2014)
45. Ushirobira, R., Perruquetti, W., Mboup, M., Fliess, M.: Estimation algébrique des paramètres intrinsèques d'un signal sinusoïdal biaisé en environnement bruité. In: Proceedings of the Gretsi. Bordeaux (2011)
46. Ushirobira, R., Perruquetti, W., Mboup, M., Fliess, M.: Algebraic parameter estimation of a biased sinusoidal waveform signal from noisy data. In: Sysid 2012, 16th IFAC Symposium on System Identification. Brussels (2012)
47. Ushirobira, R., Perruquetti, W., Mboup, M., Fliess, M.: Algebraic parameter estimation of a multi-sinusoidal waveform signal from noisy data. In: European Control Conference. Zurich (2013)
48. Xia, X.: Global frequency estimation using adaptive identifiers. IEEE Trans. Autom. Control. **47**(7), 1188–1193 (2002)
49. Yosida, K.: Operational Calculus: A Theory of Hyperfunctions. Springer, New York (1984)
50. Yu, L., Barbot, J.-P., Floquet, T.: Tire/Road contact condition identification using algebraic numerical differentiation. In IFAC Symposium on System Identification, SYSID, Saint-Malo, France (2009)

Part III
Algebraic Geometry Methods for Systems and Control Theory

Chapter 8
Symbolic Methods for Solving Algebraic Systems of Equations and Applications for Testing the Structural Stability

Yacine Bouzidi and Fabrice Rouillier

Abstract In this work, we provide an overview of the classical symbolic techniques for solving algebraic systems of equations and show the interest of such techniques in the study of some problems in dynamical system theory, namely testing the structural stability of multidimensional systems.

Keywords Algebraic systems · Real solving · Symbolic methods · Certified computations · Structural stability · Multidimensional systems

8.1 Introduction

In this work, we address the problem of solving algebraic system of equations of the form

$$\begin{cases} f_1(x_1, \ldots, x_n) = 0, \\ f_2(x_1, \ldots, x_n) = 0, \\ \qquad \vdots \\ f_s(x_1, \ldots, x_n) = 0, \end{cases} \tag{8.1}$$

where $f_1, f_2, \ldots, f_s$ are polynomials in the variables $x_1, \ldots, x_n$ with coefficients in the field of rational numbers $\mathbb{Q}$.

Before going further, a first and important question that shall be asked is: *what does solving algebraic systems mean?* Actually, answering to this question clearly and in all generality is not an easy task. The answer depends often on various parameters among which the nature of the solutions, the context of the computations as well as the field of applications for which the computations are performed.

Y. Bouzidi (✉)
Inria, Lille-Nord Europe, Villeneuve-d'Ascq, France
e-mail: Yacine.bouzidi@inria.fr

F. Rouillier
Inria, Paris, France
e-mail: Fabrice.rouillier@inria.fr

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_8

In the case of univariate polynomial equations, i.e. equations of the form $f(x) = 0$ where f is a polynomial with arbitrary coefficients, since the work of Abel in the 19th century, it is known that there is no general algebraic formulas for the solutions (solutions in radicals) when the degree of f is higher than four. An usual way to obtain a representation of the solutions is then via numerical approximations. Several methods exist for getting such approximations. One can mention for example the classical Newton–Raphson method for approximating a root (see [1] and references therein), or the bisection methods based on inclusion/exclusion criteria (Sturm's theorem, Descartes' rule of sign...) for approximating all the roots (see [2] and references therein). In addition, in many applications, one would like to perform exact computations with the resulting roots, e.g., checking the vanishing of an algebraic expression, computing it sign, etc. A suitable representation that allows such kind of computations consists in a polynomial that vanishes on the root and an isolating interval that contains this root and no other roots of the polynomial. Such an interval can then be refined to obtain an approximation of the root up to an arbitrary precision.

When it comes to systems of polynomial equations in several variables, an important aspect that governs the study of the solutions concerns their nature. More precisely, two type of systems can be distinguished. Those which admit a finite number of solutions in the algebraic closure of $\mathbb{R}$, i.e. $\mathbb{C}$, and those admitting an infinite number of solutions in $\mathbb{C}$.

For systems that admit a finite number of solutions, similarly as for univariate polynomial equations, one generally aims at finding numerical approximations of all the solutions which now are given as vectors of intervals. Two family of methods emerge, those which start from the initial polynomial system and compute numerical approximations of the solutions using for example multivariate variants of Newton–Raphson methods, interval evaluation, inclusion/exclusion criteria, homotopy continuation, etc. (see [3, 4] and references therein), and those which first focus on the computation of a formal expression of the solutions such as a univariate parametrization, a Gröbner basis or triangular sets and then compute numerical approximations of the solutions from these expressions. Such formal expressions ease in general the computation of numerical approximation of the solutions by reducing the problem to that of computing approximations of the roots of a univariate polynomial. It is worth mentioning that while the former methods (purely numerical methods) search for the solutions locally (in a given region of the solutions' space) and require regularity assumption on the input system in order to return an exact result (e.g., the system need to be squarefree, i.e., devoid from multiple solution), the methods based on the computation of formal expressions of the solutions provide a description for all the solutions of the system and do not made any assumption on the input.

Finally, for systems with an infinite number of solutions, the question of solving becomes rather vague and the specification of the output difficult to establish. In many applications, a frequently asked question concerns the existence of real solutions of a given system. More generally, a central problem for systems with infinite number of solutions is the computation of one real point in each connected component.

In this chapter we review some classical techniques for solving systems of polynomial equations focusing our attention on the exact symbolic methods, that is,

methods providing an exact and complete description of the solutions. In addition, in order to motivate the use of such methods in the context of dynamical systems theory, we present an application of the latter to the problem of testing the stability of multidimensional systems (e.g. [5]) which we give the general statement below.

Structural stability of multidimensional systems. Let consider a single-input single-output (SISO) multidimensional discrete linear system, described within the frequency domain by a transfer function

$$G(z_1, \ldots, z_n) = \frac{N(z_1, \ldots, z_n)}{D(z_1, \ldots, z_n)}, \tag{8.2}$$

where N and D are polynomials in the complex variables $z_1, \ldots, z_n$ with rational coefficients with $\gcd(N, D) = 1$. This system is said to be *structurally stable* if the denominator of its transfer function is devoid from zeros in the complex unit polydisc $\mathbb{D}^n := \prod_{k=1}^{n} \{z_k \in \mathbb{C} || z_k| \leq 1\}$, or in other words:

$$D(z_1, \ldots, z_n) \neq 0 \text{ for } |z_1| \leq 1, \ldots, |z_n| \leq 1. \tag{8.3}$$

In order to check the above condition, a first step consists in rewriting it under algebraic form (conditions that involve only algebraic systems of equations). The resulting conditions are then processed by means of solving systems algorithms. As we will see further in the text, depending on the dimension of the multidimensional system, the resulting algebraic systems admits, either a finite number of zeros (for one or two dimensional systems) or an infinite number of zeros (for n-dimensional systems with $n \geqslant 3$). In each case, dedicated solving algorithms are used for testing the resulting conditions.

The chapter is organized as follow. We first recall in Sect. 8.2 the basic mathematical material behind the problem of solving symbolically systems of polynomial equations. In Sect. 8.3, we present some basic results about the roots of univariate polynomials. In Sect. 8.4 we provide a short introduction to Gröbner basis, a key tool in the study of systems of polynomial equations. Section 8.5 is devoted to the problem of solving systems with finitely many solutions called *zero-dimensional* systems. Finally, we address in Sect. 8.6 the problem of solving algebraic systems with an infinite number of solutions. At the end of each Section, we illustrate the use of the presented techniques on the problem of testing the structural stability of multidimensional systems.

8.2 Preliminaries

In the sequel, we will borrow some elements from algebraic geometry and commutative algebra to address problem (8.1). This problem consists in studying the zero-sets of polynomial systems. Geometrically, such sets correspond to algebraic

varieties such as curves, surfaces or object of higher dimension. The good algebraic framework to study these kind of object is the theory of polynomial ideals. After defining the concepts of ideal and variety, we recall a classical result about the correspondence between them which is at the core of the solving systems theory. This correspondence allows one to translate any question about the zeros of a system into a question about ideals, so that it can be answered using symbolic algorithms.

Given a set of polynomials $f_1, \ldots, f_s$ in $\mathbb{K}[x_1, \ldots, x_n]$, one can construct other polynomials as linear polynomial combinations of the latter. This leads to the following definition.

Definition 1 (*Ideal*) The set of polynomials of the form $\sum_{i=1}^{s} g_i f_i$, with $g_i \in \mathbb{K}[x_1, \ldots, x_n]$, is called the ideal generated by $f_1, \ldots, f_s$ and denoted $\langle f_1, \ldots, f_s \rangle$

The ideal $\langle f_1, \ldots, f_s \rangle$ contains $f_1, \ldots, f_s$ and is a stable subset under addition and multiplication by elements in $\mathbb{K}[x_1, \ldots, x_n]$. Actually, it is the smallest subset of $\mathbb{K}[x_1, \ldots, x_n]$ that satisfies this property. Another important property is that every ideal in $\mathbb{K}[x_1, \ldots, x_n]$ is generated by a finite number of polynomials. This property stems from the fact that $\mathbb{K}[x_1, \ldots, x_n]$ is *noetherian*. Another important consequence of the *noetherianity* of $\mathbb{K}[x_1, \ldots, x_n]$ is that every ascending chain of ideals $I_1 \subsetneq I_2 \subsetneq \cdots I_k \subsetneq \cdots$ in $\mathbb{K}[x_1, \ldots, x_n]$ stabilizes. From the computation point of view, this last property is crucial since it guarantees the termination of algorithms involving polynomial ideals in $\mathbb{K}[x_1, \ldots, x_n]$.

The geometrical objects we are going to study are defined as the zero-sets of systems of polynomial equations called algebraic varieties. In Sect. 8.6.1, we further introduce the notion of semi-algebraic set that consists in the points of an algebraic variety which satisfy certain inequalities.

Definition 2 (*Algebraic variety*) Let $f_1, \ldots, f_s$ be polynomials in $\mathbb{K}[x_1, \ldots, x_n]$. Then, the set

$$\mathcal{V}(f_1, \ldots, f_s) = \left\{(a_1, \ldots, a_n) \in \mathbb{K}^n \mid f_i(a_1, \ldots, a_n) = 0, \ \forall\, i \in \{1, \ldots, s\}\right\}$$

is called the algebraic variety defined by $f_1, \ldots, f_s$.

Hence, the algebraic variety defined by a set of polynomials $f_1, \ldots, f_s$ is the subset of the affine space $\mathbb{K}^n$ that forms the zeros of the polynomial system $\{f_1 = \cdots = f_s = 0\}$. This variety is also defined as the zero set of the the ideal $\langle f_1, \ldots, f_s \rangle$. In the sequel, we often consider $\mathbb{K} = \mathbb{Q}$ and study two kind of varieties: the complex variety $\mathcal{V}_{\mathbb{C}}$, i.e., the set of complex zeros of a given ideal, and the real variety $\mathcal{V}_{\mathbb{R}}$, i.e., the set of its real zeros.

Example 1 Consider the polynomial $f(x, y) = x^4 - x^2 + y^2 \in \mathbb{Q}[x, y]$. The variety $\mathcal{V}_{\mathbb{R}}(f)$ corresponds to the points of $\mathbb{R}^2$ that satisfy the equation $f(x, y) = 0$ (see Fig. 8.1).

There exists an important correspondence between the algebraic concept of ideal and the geometric concept of variety. To understand this correspondence let start with the following definition.

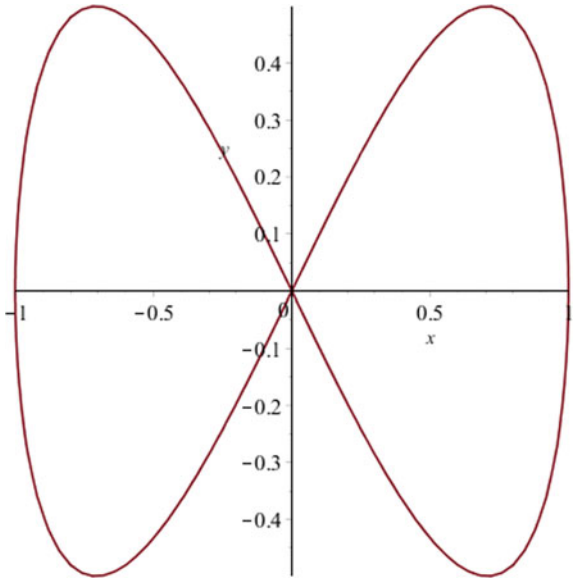


Fig. 8.1 The real variety associated to $x^4 - x^2 + y^2$

Definition 3 Let V be an algebraic variety of $\mathbb{K}^n$. Define the set:

$$\mathcal{I}(V) = \{f \in \mathbb{K}[x_1, \ldots, x_n] \mid f(a_1, \ldots, a_n) : (a_1, \ldots, a_n) \in V\}.$$

The set $\mathcal{I}(V)$ is an ideal of $\mathbb{K}[x_1, \ldots, x_n]$. It is called the ideal of V.

Given an algebraic variety V, we can easily notice that the variety corresponding to the ideal of V is V itself, i.e., $\mathcal{V}(\mathcal{I}(V)) = V$. However, the reciprocal, i.e., $\mathcal{I}(\mathcal{V}(I)) = I$ is not always true as illustrated by the following example.

Example 2 Let consider the ideal $\langle (x-y)^2 \rangle \subset \mathbb{C}[x, y]$. $\mathcal{V}_{\mathbb{C}}(I)$ is the complex line given by the equation $x = y$ whose the corresponding ideal is $\langle x - y \rangle$, i.e., $\mathcal{I}(\mathcal{V}_{\mathbb{C}}(\langle (x-y)^2 \rangle)) \neq \langle (x-y)^2 \rangle$

In fact, the previous example shows that the correspondence between ideals and varieties is in general not one-to-one, different ideals can lead to the same variety. However, when $\mathbb{K}$ is an algebraically closed field, a fundamental result establishes a bijection between the set of varieties and the set of the so-called *radical* ideals.

Theorem 1 ([6, Sect. 4.1]) *If $\mathbb{K}$ is algebraically closed, then for any $I \subset \mathbb{K}[x_1, \ldots, x_n]$*

$$\mathcal{I}(V(I)) = \sqrt{I},$$

where $\sqrt{I} = \{g \in \mathbb{K}[x_1 \ldots, x_n] \mid \exists\, e \in \mathbb{N}, g^e \in I\}$ is called the radical of I.

The previous theorem, known as the *Hilbert's Nullstellensatz theorem*, is the analogous of the fundamental theorem of algebra that relates a univariate polynomial to the set of its roots. It is at the core of the theory of solving algebraic systems of

polynomials with coefficients in an algebraically closed field. In particular, it allows to translate any question about the solutions of an algebraic system of equations to a question about the radical ideal generated by this system.

Finally, when manipulating systems of algebraic equations, we are often interested in describing the nature of the corresponding zero-sets (algebraic varieties). For instance, the latter can consist in a finite number of points (e.g. the roots of a univariate polynomial), or an infinite number of points (e.g. the circle defined by the ideal $\langle x^2 + y^2 - 1) \rangle$). Intuitively, a convenient way to describe the nature of an algebraic variety is to consider the *degree of freedom* of an arbitrary point moving on it. In the case of a finite number of points, there is no way to move from a point to another point while remaining on the variety, the degree of freedom is thus zero. In the case of the variety defined by $\langle x^2 + y^2 - 1 \rangle$, one can only move along the circle $x^2 + y^2 - 1 = 0$, the degree of freedom is equal to one. This notion of *degree of freedom* bears the name of *dimension* of an algebraic variety. It may be defined in various equivalent ways. The following definition gives an intuitive description of it. For more details on the dimension of an algebraic variety and how the latter can be computed, the reader may refer to [6, Sect. 9].

Definition 4 (*Dimension*) Let $V \subset \mathbb{K}^n$ be an algebraic variety. The dimension of V is the largest positive integer d such that there exists an algebraic variety $W \subset \mathbb{K}^d$ so that the projection

$$\begin{aligned} \mathbb{C}^n &\to \mathbb{C}^d \\ (x_1, \ldots, x_n) &\mapsto (x_{i_1}, \ldots, x_{i_d}), \end{aligned}$$

where $\{i_1, \ldots, i_d\}$ is a subset of $\{1, \ldots, n\}$, is surjective onto $\mathbb{C}^d / W$.

8.3 The Univariate Case

In this section, we start by recalling some classical tools and algorithms for the study of the roots of univariate polynomials. Beside the fact that such a material is a basic building block in solving systems problems, which are generally reduced to univariate ones (see Sect. 8.5.2), some of the presented results play also an important role in many algorithms that compute with multivariate polynomials considered as univariate polynomials with coefficients in polynomial rings (see Sect. 8.6.1).

8.3.1 GCD, Resultant, Subresultants

Definition 5 (*Generalized remainder sequence*) Let $\mathbb{D}$ be a domain, $\mathbb{F}$ its fraction field and $f, g \in \mathbb{D}[x]$ with degree$(f) >$ degree(g). Consider the sequence $(\rho_i, r_i, q_i, s_i, t_i)_{i=0\ldots l}$ with $\rho_i \in \mathbb{F}^\star$, $r_i, q_i, s_i, t_i \in \mathbb{F}[x]$ such that:

- $\rho_0 r_0 = f \; s_0 = \rho_0^{-1} \; t_0 = 0$, and $\rho_1 r_1 = g \; s_1 = 0 \; t_1 = \rho_1$,
- for $i \geqslant 1$, $r_{i-1} = q_i r_i + \rho_{i+1} r_{i+1}$, degree$(r_{i+1}) <$ degree(r_i),

- $l \in \mathbb{N}, r_l \neq 0 \wedge r_{l+1} = 0$,
- $s_{i+1} := (s_{i-1} - q_i s_i)/\rho_{i+1}$, $t_{i+1} := (t_{i-1} - q_i t_i)/\rho_{i+1}$.

It is important to point out that l, as well as the degree sequence does not depend on the choice of the ρ_i. In addition, we have:

- When $\rho_0 = \ldots = \rho_l = 1$, $(r_i)_{i=0\ldots l}$ is the **classical remainder sequence**.
- When $\rho_0 = 1, \rho_1 = 1, \rho_i = (-1)^{i+1}$, $(r_i)_{i=0\ldots l}$ is the so called **signed Euclidean remainder sequence** which corresponds for $g = f'$ to the famous **Sturm sequence** (see Proposition 5).
- When the ρ_i's are recursively set as $\rho_i = \mathrm{lc}(q_i r_i - r_{i-1})$, where lc(.) denotes the leading coefficient, the $(r_i)_{i=0\ldots l}$ is the so called **monic remainder sequence**.

In all these cases, r_l is a GCD of f, g and $\forall i = 0 \ldots l, r_i = s_i f + t_i g$.

An important remark is that when f, g are in $\mathbb{D}[x]$, the polynomials r_i appearing in the above remainder sequences belong to $\mathbb{F}[x]$. In particular, if $\mathbb{D} = \mathbb{K}[y_1, \ldots, y_n]$ (i.e., the coefficients of f and g are polynomials in $y_1, \ldots, y_n$), then the sequence of r_i will have coefficients in $\mathbb{K}(y_1, \ldots, y_n)$ (i.e., rational fraction in $y_1, \ldots, y_n$). This fact prevents the remainders r_i from being specialized at any values of $y_1, \ldots, y_n$. More precisely, there exist $\alpha_1, \ldots, \alpha_n$ such that the i-th remainder of $f(\alpha_1, \ldots, \alpha_n, x)$ and $g(\alpha_1, \ldots, \alpha_n, x)$ is not equal to the specialization of the ith remainder of $f(y_1, \ldots, y_n, x)$ and $g(y_1, \ldots, y_n, x)$. Such "bad" specializations correspond to the values of $y_1, \ldots, y_n$ that cancel the denominators of some coefficients appearing in the computation of r_i.

One way to overcome this specialization issue is to keep computations in the polynomial ring of coefficients. This can be done using the notion of subresultant sequence, which we define now.

Let $f = \sum_{i=0}^{n} a_i x^i$ and $g = \sum_{i=0}^{m} b_i x^i$ with the convention that $f_i = g_i = 0$ if $i \leqslant 0$ and denote by $(r_i)_{i=0\ldots l}$ the monic remainder sequence of f and g as defined above. We introduce the following $(n + m - 2k)(n + m - k)$ matrix formed by the coefficients of f and g

$$S_k = \begin{pmatrix} a_n & a_{n-1} & \cdots & \cdots & \cdots & a_0 & & \\ & a_n & a_{n-1} & \cdots & \cdots & \cdots & a_0 & \\ & & \ddots & & & & & \ddots \\ & & & a_n & a_{n-1} & \cdots & \cdots & \cdots & a_0 \\ b_m & b_{m-1} & \cdots & \cdots & \cdots & b_0 & & \\ & b_m & b_{m-1} & \cdots & \cdots & \cdots & b_0 & \\ & & \ddots & & & & & \ddots \\ & & & b_m & b_{m-1} & \cdots & \cdots & \cdots & b_0 \end{pmatrix},$$

and we set $\sigma_k = \det(S_k)$. Note that S_0 is the well known *Sylvester matrix*.

For each $i = 0 \dots l$, we denote by r_i the i-th monic remainder of f and g of degree d_i and we denote by $s_i, t_i \in \mathbb{D}[x]$ of degree respectively strictly less than $m - d_i - 1$ and $n - d_i - 1$, the unique solution of the system of linear equations

$$S_{d_i}^T (s_i, t_i)^T = (0, \dots, 0, 1)^T,$$

and thus, it turns out that $\sigma_{d_i} r_i = \sigma_{d_i} s_i f + \sigma_{d_i} t_i g$ when $r_i \neq 0$ is a non-zero remainder in the monic remainder sequence and $\sigma_{d_i} = 0$ otherwise.

Definition 6 (*Subresultants*) Let $f, g \in \mathbb{D}[x]$ with $\text{degree}(f) = n \geqslant m = \text{degree}(g)$.

- The sequence $(\sigma_i)_{i=0\dots m}$ is called the principal subresultant sequence associated to the couple (f, g).
- The sequence $(\text{Sres}_i(f, g) = \sigma_i r_{d_i})_{i=0\dots m}$ is called the polynomial subresultant sequence associated to (f, g). Sres_i is the polynomial subresultant of degree i.
- The polynomial subresultant of degree 0, Sres_0 is called the resultant of f and g, it belongs to the ideal generated by f and g.

The subresultant sequence has properties that are comparable to those of the classical remainder sequences.

Proposition 1 *Let* $f, g \in \mathbb{D}[x]$, $f = a_0 + \dots + a_n x^n, g = b_0 + \dots + b_m x^m$ *and denote by* $\mathbb{F}$ *the fraction field of* $\mathbb{D}$. *The following properties are equivalent:*

- f, g *have a common root in* $\overline{\mathbb{F}}$, *the algebraic closure of* $\mathbb{F}$, *or* $a_n = b_m = 0$.
- f, g *have a non constant common factor in* $\mathbb{F}[x]$, *or* $a_m = b_n = 0$. *If* f, g *have a non constant common factor, then their* gcd *is proportional to the non-zero polynomial subresultant of minimal index.*
- $\exists s, t \in \mathbb{F}[x]$ *with* $\text{degree}(s) < m$ *and* $\text{degree}(t) < n$ *such that* $sf + tg = 0$.
- $\sigma_0 = \text{Resultant}(f, g, x) = 0$.

In addition, as mentioned above, the subresultant sequence is well specialized.

Proposition 2 *Let* $\mathbb{D}$ *and* $\mathbb{D}'$ *be unique factorization domains and* $\phi : \mathbb{D} \to \mathbb{D}'$ *be a morphism. Let* $f, g \in \mathbb{D}[x]$ *and suppose that* $deg(\phi(f)) = deg(f) > deg(g) = deg(\phi(g))$. *Then* $\phi(Sres_i(f, g)) = Sres_i(\phi(f), \phi(g)), \forall i = 0 \dots deg(g)$.

8.3.2 Real Roots of Univariate Polynomials with Real Coefficients

Let $P = \sum_{i=0}^{n} a_i x^i \in \mathbb{R}[x]$ be a polynomial with real coefficients. We can easily bound the module of its (complex) roots as well as the distance between two roots.

Proposition 3 [7, Prop. 10.9],[8, Thm. 1] *If* α *is a complex root of* P *and if* $a_n = 1$, *then* $|\alpha| < 1 + \max_{i=0}^{n}(|a_i|)$.

If P *has no multiple roots and if* $sep(P)$ *denotes the distance between two roots of* P, *then* $sep(P) \geqslant \sqrt{\frac{3}{n^n+2}} \cdot \frac{1}{||P||_2^{n-1}}$ *with* $||P||_2 = \sqrt{\sum_{i=0}^{d} a_i^2}$.

The above bounds give a straightforward exact algorithm to isolate the real roots of $P \in \mathbb{Q}[x]$.

Naive univariate isolation:

compute $\overline{P} := \frac{P}{\mathrm{Gcd}(P, \partial P/\partial x)}$, the squarefree-part of P by Euclid's algorithm;
compute $M = 1 + \max_{i=0}^{n}(|\frac{a_i}{a_n}|)$;
compute any $m < \sqrt{\frac{3}{n^n+2}} \cdot \frac{1}{||P||_2^{n-1}}$;
compute the sign sequence $\mathrm{sign}\left(\overline{P}\right)(-M + k\,m)$, $0 \leqslant i \leqslant \frac{2M}{m}$, and return the intervals in the form $(-M + k\,m, -M + (k+1)m)$ such that $\mathrm{sign}\left(\overline{P}\right)(-M + k\,m)\mathrm{sign}\left(\overline{P}\right)(-M + (k+1)m) < 0$ as well as the rational numbers $-M + k\,m$ such that $\mathrm{sign}\left(\overline{P}\right)(-M + k\,m) = 0$.

However, such a simple algorithm would have an exponential behavior with respect to the degree n and the computation time would explode very quickly when increasing this degree. Alternatively, modern algorithms (and implementations) avoid the brutal partitioning of the interval $(-M, M)$ and use the so-called bisection strategies. The latter consist in iteratively subdividing the initial interval until getting isolating intervals around the roots. At each step, an interval of the form $I_{c,k} = \left(\frac{c}{2^k}, \frac{c+1}{2^k}\right)$ with $0 \leqslant c < 2^k$ is "visited", and some oracle is used to determine whether the polynomial has 0, 1 or more than one root in $I_{c,k}$ with respect to the following general principle:

General bisection strategy:

List = (0, 1);
while *List* $\neq \emptyset$ **do**
  Remove $\left(\frac{c}{2^k}, \frac{c+1}{2^k}\right)$ from List;
  If P has one root in $\left(\frac{c}{2^k}, \frac{c+1}{2^k}\right)$ add $\left(\frac{c}{2^k}, \frac{c+1}{2^k}\right)$ to the result;
  If P has more than one root in $\left(\frac{c}{2^k}, \frac{c+1}{2^k}\right)$, add $\left(\frac{2c}{2^{k+1}}, \frac{2c+1}{2^{k+1}}\right)$ and $\left(\frac{2c+1}{2^{k+1}}, \frac{2c+2}{2^{k+1}}\right)$ to List;
end

Hence, given an oracle for counting the number of real roots inside an interval (or at least deciding if there is 0, 1 or more than 1 real root), the above bisection strategy yields an algorithm for isolating the real roots of a univariate polynomial. A well known Oracle for that purpose is based on the so-called Sturm sequence.

Definition 7 Let $P \in \mathbb{R}[x]$. A Sturm sequence associated with P on a given interval $(a, b) \in \mathbb{R}$ is a sequence $f_0(x), \ldots f_s(x) \in \mathbb{R}[x]$ such that:

- $f_0 = P$;
- f_s has no real root in (a, b);
- for $0 < i < s$, if $\alpha \in (a, b)$ is such that $f_i(\alpha) = 0$, then $f_{i-1}(\alpha) f_{i+1}(\alpha) < 0$;
- if $\alpha \in [a, b]$ is such that $f_0(\alpha) = 0$, then we have

$$\begin{cases} f_0 f_1(\alpha - \epsilon) < 0, \\ f_0 f_1(\alpha + \epsilon) > 0, \end{cases}$$

for any ϵ sufficiently small.

Proposition 4 *Let $P \in \mathbb{R}[x]$ and $f_0(x), \dots f_s(x)$ a Sturm sequence for P on (a, b). $V(a_1, \dots, a_s)$ denotes the number of sign changes in the sequence $a_1, \dots, a_s$ after removing zeros and $V_{stu}(P(c)) = V(f_0(c), \dots, f_s(c))$, then*

$$V_{stu}(P(b)) - V_{stu}(P(a))$$

equals the number of real roots of P in (a, b).

A key point is that computing a Sturm sequence for a polynomial amounts essentially to computing a remainder sequence of this polynomial and its derivative.

Proposition 5 [7, Thm. 2.50] *Using Definition 5, the remainder sequence $(r_i)_{i=0,\dots,l}$ obtained when taking $\rho_0 = 1, \rho_1 = 1, \rho_i = (-1)^{i+1}$, $f = P, g = P'$ is a Sturm sequence for P on any interval (a, b).*

As for the classical remainder sequence, note that the Sturm sequence does not behave well under specialization, but, as for the classical remainder sequence, it suffices to multiply all the polynomials by the corresponding subresultants in order to solve the problem of specialization: if $(\mathrm{Stu}_i)_{i=0,\dots,l}$ is the Sturm sequence, then $(\sigma_{n_i}\mathrm{Stu}_i)_{i=0,\dots,l}$ specializes well. Note that this new sequence, known as Sturm–Habicht sequence or signed subresultant sequence (see [7]), is not formally a Sturm sequence anymore but Proposition 4 can be adapted to get a well suited sign change counting for computing the roots of P in (a, b) using Sturm–Habicht sequences (see [7]).

The currently fastest implementations for the isolation of the real roots of univariate polynomials are not using Sturm (or Sturm–Habicht) sequences anymore but an Oracle based on Descartes' rule of signs:

Proposition 6 [7, Thm. 2.33] *Let $P = \sum_{i=0}^{n} a_n x^n \in \mathbb{R}[x]$ be a squarefree polynomial. The number of strictly positive real roots of P is dominated by $Var(P) = V(a_0, \dots, a_n)$ and equals $Var(P)$ modulo 2.*

In particular, if $\mathrm{Var}(P) = 0$, P has no positive roots, and if $\mathrm{Var}(P) = 1$, then P has exactly one positive root. This result can be adapted for inspecting the number of roots in an interval of the form $\left(\frac{c}{2^k}, \frac{c+1}{2^k}\right)$.

Corollary 1 *Let $P = \sum_{i=0}^{n} a_i x^i \in \mathbb{R}[x]$ be a squarefree polynomial and define the polynomials $P_{k,c} = 2^{kn} P\left(\frac{x+c}{2^k}\right)$, $R(P_{k,c}(x)) = x^n P_{k,c}(\frac{1}{x})$ and:*

$$T_1(R(P_{k,c})) = R(P_{k,c}(x + 1)).$$

The number of strictly positive real roots of P in $\left(\frac{c}{2^k}, \frac{c+1}{2^k}\right)$ is dominated by $Var(T_1(R(P_{k,c})))$ and equals $Var(T_1(R(P_{k,c})))$ modulo 2.

The formula in Corollary 1 is nowadays used in the general bisection algorithm in order to decide if a polynomial has 0, 1 or more than one root in $\left(\frac{c}{2^k}, \frac{c+1}{2^k}\right)$. Unlike the Sturm-based strategy, Descartes rule of signs does not provide the exact number

of roots but only a bound. However, the resulting algorithm still works since it has been shown (see [9]) that when the intervals $\left(\frac{c}{2^k}, \frac{c+1}{2^k}\right)$ are sufficiently small, then Descartes' rule of signs always return 0 or 1 and so allows to conclude.

8.4 Gröbner Bases

Computing modulo ideals in the univariate ring $\mathbb{Q}[x]$ reduces to a simple Euclidean division. Indeed, given an ideal $I = (f_1, \ldots, f_n) \subset \mathbb{Q}[x]$ and a polynomial $p \in \mathbb{Q}[x]$, computing the reduction of p modulo I amounts to compute the remainder of the Euclidean division of p by the greatest common divisor of $\{f_1, \ldots, f_n\}$. When it comes to the multivariate polynomial ring $\mathbb{Q}[x_1, \ldots, x_n]$, computing the reduction of $p \in \mathbb{Q}[x_1, \ldots, x_n]$ modulo an ideal $I \subset \mathbb{Q}[x_1, \ldots, x_n]$ consists in obtaining a canonical representation of p in $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$. A Gröbner basis of an ideal I is a computable set of generators of I that allows to perform this operation.

In order to define Gröbner bases, a first step is to extend the usual Euclidean division, from polynomials in $\mathbb{Q}[x]$, to polynomials in $\mathbb{Q}[x_1, \ldots, x_n]$. To do so, as for the Euclidean division in $\mathbb{Q}[x]$, one has to associate to each polynomial in $\mathbb{Q}[x_1, \ldots, x_n]$ a leading term with respect to which the reduction is made. This requires the introduction of the notion of *admissible ordering* on monomials in $\mathbb{Q}[x_1, \ldots, x_n]$. In the following, we denote by x^α the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ where $(\alpha = (\alpha_1, \ldots, \alpha_n))$.

Definition 8 An admissible monomial ordering in $\mathbb{Q}[x_1, \ldots, x_n]$ is a binary relation $<$ defined on the set of monomials x^α or equivalently on the set of $\alpha \in \mathbb{Z}^n_{\geqslant 0}$ such that:

- $<$ is a total ordering relation.
- For any α, β and $\gamma \in \mathbb{Z}^n_{\geqslant 0}$, $\alpha < \beta \Longrightarrow \alpha + \gamma < \beta + \gamma$.
- For any $\alpha, \beta \in \mathbb{Z}^n_{\geqslant 0}$, $\alpha < \alpha + \beta$.

These conditions imply Noetherianity, which means that every strictly decreasing sequence of monomials is finite.

In the following, we will mainly use the two following orderings and some others which we will define later.

- *Lexicographic order (Lex)*:

$$\begin{aligned} & x_1^{\alpha_1} \cdots x_n^{\alpha_n} <_{\text{Lex}} x_1^{\beta_1} \cdots x_n^{\beta_n} \\ & \Leftrightarrow \exists i_0 \leq n, \ \begin{cases} \alpha_i = \beta_i, & \text{for } i = 1, \ldots, i_0 - 1, \\ \alpha_{i_0} < \beta_{i_0}. \end{cases} \end{aligned} \tag{8.4}$$

- *Degree reverse lexicographic order (DRL)*:

$$x_1^{\alpha_1} \cdots x_n^{\alpha_n} <_{\text{DRL}} x_1^{\beta_1} \cdots x_n^{\beta_n}$$

$$\Leftrightarrow \begin{cases} \sum_k^n \alpha_k < \sum_k^n \beta_k \\ \qquad \text{or} \\ \sum_k^n \alpha_k = \sum_k^n \beta_k \text{ and } x_1^{-\alpha_n} \cdot \ldots \cdot x_n^{-\alpha_1} <_{\text{Lex}} x_1^{-\beta_n} \cdot \ldots \cdot x_n^{-\beta_1}. \end{cases} \tag{8.5}$$

We also need the following notation.

Definition 9 Let $p = \sum_\alpha a_\alpha x^\alpha \in \mathbb{Q}[x_1, \ldots, x_n]$ and let $<$ be a monomial ordering. Then, we have:

- The multidegree of p is $\text{multideg}(p) = \max_<(\alpha \in \mathbb{Z}^n_{\geqslant 0} : a_\alpha \neq 0)$.
- The leading coefficient of p is $\text{LC}_<(p) = a_{\text{multideg}(p)} \in \mathbb{Q}$.
- The leading monomial of p is $\text{LM}_<(p) = x^{\text{multideg}(p)}$.
- The leading term of p is $\text{LT}_<(p) = \text{LC}(p)\,\text{LM}(p)$.

Given any admissible monomial ordering $<$, one can easily extend the classical Euclidean division to *reduce* a polynomial $p \in \mathbb{Q}[x_1, \ldots, x_n]$ by a set of polynomials F, performing the reduction with respect to each polynomial of F until getting an expression which cannot be further reduced (see [6] for details). This yields the following result.

Theorem 2 *Let $F = \{f_1, \ldots, f_n\}$ be a set of polynomials in $\mathbb{Q}[x_1 \ldots, x_n]$. For any $p \in \mathbb{Q}[x_1, \ldots, x_n]$, there exists $q_1, \ldots, q_n, r \in \mathbb{Q}[x_1, \ldots, x_n]$ such that*

$$p = q_1\, f_1 + \cdots + q_n\, f_n + r,$$

and none of the monomials of r is divisible by a leading term of $f_1, \ldots, f_n$.

The above reduction is denoted by `Reduce`$(p, F, <)$ (reduction of the polynomial p with respect to F). The polynomial r is the output of the function `Reduce`$(p, F, <)$ and is called the remainder of the reduction of p by F. Unlike the univariate case, this remainder polynomial now depends on the order in which the reductions by the polynomials of F are performed, and thus, the reduction is not canonical. In order to remedy this situation, the notion of *Gröbner basis* of an ideal has been introduced by Buchberger. Roughly speaking, a Gröbner basis G of an ideal I is a set of polynomials that generates the ideal and for which the function `Reduce`$(p, G, <)$ is canonical. In that case, the aforementioned remainder is referred to as the *normal form* of p with respect to G. The following definition of Gröbner basis is purely mathematical.

Definition 10 A set of polynomials G is a Gröbner basis of an ideal I with respect to a monomial ordering $<$ if for all $f \in I$ there exists $g \in G$ such that $\text{LM}_<(g)$ divides $\text{LM}_<(f)$.

Theorem 3 [6, Sect. 2.6] *Let I be an ideal in $\mathbb{Q}[x_1, \ldots, x_n]$ and G a Gröbner basis of I with respect to a fixed monomial ordering $<$. Then, for any $p \in \mathbb{Q}[x_1, \ldots, x_n]$, the reduction of f modulo G is uniquely determined. In particular, $p \in I$ iff this reduction is zero, i.e.,* `Reduce`$(p, F, <) = 0$.

Classical algorithms for computing Gröbner bases of ideals start from a set of generators and construct iteratively new sets of generators until obtaining a Gröbner basis. The most popular algorithm for computing Gröbner bases is Buchberger's algorithm [10]. It is implemented in most of computer algebra software such as `Maple` and `Mathematica`. This algorithm has several variants and modern ones [11] make a large use of dedicated sparse linear algebra techniques and can be found in some general computer algebra systems such as `Magma` or `Maple` as well as in some dedicated systems like `FGb`.

8.4.1 Applications of Gröbner Bases

Gröbner bases are key objects for performing computations with polynomial ideals. As an illustration, we present in the following three important problems that can be solved through Gröbner bases computation.

The emptiness of the zero set. In several applications, a frequently asked question concerns the consistency of an algebraic system of equations, that is, the existence of common zeros in the algebraic closure $\overline{\mathbb{K}}$ of the coefficients field $\mathbb{K}$. Given an ideal $I = \langle f_1, \ldots, f_s \rangle \subset \mathbb{Q}[x_1, \ldots, x_n]$, this problem translates into testing if the variety associated to the ideal I, that is, $\mathcal{V}(I) = \{\alpha \in \mathbb{C}^n \mid \forall f \in I,\ f(\alpha) = 0\}$ is empty. According to the *Nullstellensatz theorem*, $\mathcal{V}(I)$ is empty if and only if $1 \in I$. Given a Gröbner basis G of I, this condition is equivalent to the existence of an element of G that belongs to $\mathbb{Q}$.

The ideal membership problem. Given $I = \langle f_1, \ldots, f_s \rangle \subset \mathbb{Q}[x_0, \ldots, x_n]$ and a polynomial $p \in \mathbb{Q}[x_1, \ldots, x_n]$, an important question consists in testing whether the polynomial p belongs to the ideal I. In particular this implies that the polynomial vanishes at the zero-set corresponding to the ideal I. If G denotes the Gröbner basis associated to I, then according to Theorem 3, this can be done by computing the normal form of p modulo G and checking that the latter is zero.

An important question that stems from the membership problem is the representation of p. Indeed, if $p \in I$, then by definition, there exist polynomials $q_1, \ldots, q_s$ in $\mathbb{Q}[x_1, \ldots, x_n]$ such that $P = q_1\, f_1 + \cdots + q_s\, f_s$. An interesting problem is then to determine effectively the polynomials $q_1, \ldots, q_s$. One natural approach is to compute the reduction of p modulo the polynomials of the Gröbner basis $g_1, \ldots, g_l$, and then express each g_i as a polynomial combination of $f_1, \ldots, f_s$ using the calculations performed during the construction of the Gröbner basis. In such a computation, we are interested in polynomials $q_1, \ldots, q_s$ with the minimum degree. It was proved (see for instance [12]) that, in general, the degree of such $q_1, \ldots, q_s$ is bounded by a value that is doubly exponential in the number of variables n, i.e. of the form d^{2^n} where d is the maximum degree of $p, f_1, \ldots, f_s$.

The elimination problem. If $I \subset \mathbb{Q}[x_1, \ldots, x_n]$ and i is an integer satisfying $1 \leq i \leq n$. The ideal $I_i = I \cap \mathbb{K}[x_{i+1}, \ldots, x_n]$, consisting of the elements of I that

do not depend on the variables $x_1, \ldots, x_i$, is called the *i-th elimination ideal* of I. These ideals play an important role in the computation with polynomial ideals and, in particular, for solving algebraic system of equations. Algorithmically, obtaining such ideals can be done by eliminating variables. A convenient way to do that is to compute Gröbner bases with respect to an appropriate ordering called *elimination ordering*.

Definition 11 A monomial ordering $<$ in $\mathbb{Q}[x_1, \ldots, x_r, x_{r+1}, \ldots, x_n]$ is an elimination ordering with respect to the block $[x_1, \ldots, x_r]$ if for any polynomial $p \in \mathbb{Q}[x_1, \ldots, x_r, x_{r+1}, \ldots, x_n]$, then we have:

$$\mathrm{LT}_<(p) \in \mathbb{Q}[x_{r+1}, \ldots, x_n] \Rightarrow p \in \mathbb{Q}[x_{r+1}, \ldots, x_n].$$

Then, a fundamental result gives a description of elimination ideals using the Gröbner bases computed with respect to a given elimination ordering.

Theorem 4 [6, Sect. 3.1] (Elimination theorem) *Let I be an ideal of $\mathbb{Q}[x_1, \ldots, x_n]$ and $i \in \{1, \ldots, n\}$. If G is a Gröbner basis for an elimination ordering with respect to the block $[x_1, \ldots, x_{i-1}]$, then $G_i = G \cap \mathbb{Q}[x_i, \ldots, x_n]$ is a Gröbner basis of the elimination ideal $I_i = I \cap \mathbb{Q}[x_i, \ldots, x_n]$.*

A well known elimination ordering is the lexicographic ordering described above. The above theorem shows in particular that a Gröbner basis computed with respect to the lexicographic ordering eliminates not only the first variable but also the first two variables, the first three variables and so on. In the context of solving algebraic system of equations, this provides a way to obtain a triangular description of the solutions. In the case of system with finitely many solutions, such a method yields a generalization of the classical Gaussian elimination for solving algebraic systems of equations. Computing the solutions then consists in solving inductively the obtained equations. Starting from the isolation of the roots of the polynomial in the last variable, then the resulting intervals are substituted in the next polynomial, and the isolation is performed again and so on.

Two important operations that stem from the elimination orderings are the *projection* and the *localization*, which are summarized in Propositions 7 and 8. To facilitate their illustrations, the following notation is needed. Given any subset $\mathcal{V}$ of $\mathbb{C}^n$ (d is an arbitrary positive integer), we denote by $\overline{\mathcal{V}}$ its *Zariski closure*, that is, the smallest algebraic variety of $\mathbb{C}^n$ containing $\mathcal{V}$. If $\mathcal{V}$ is a *constructible set* (i.e., defined by equations and inequations), then $\overline{\mathcal{V}}$ is also the closure for the usual topology.

Proposition 7 [6, Sect. 3.2] *Let $I \subset \mathbb{Q}[x_1, \ldots, x_n]$ be an ideal and denote by $V(I) \subset \mathbb{C}^n$ the corresponding algebraic variety. Consider the following projection map:*

$$\begin{aligned} \Pi_i : \mathbb{C}^n &\to \mathbb{C}^{n-i} \\ (\alpha_1, \ldots, \alpha_n) \in V(I) &\mapsto (\alpha_{i+1}, \ldots, \alpha_n) \in \Pi_i(V_{\mathbb{C}}). \end{aligned}$$

Then, we have

$$V(I_i) = \overline{\Pi_i(V)},$$

where I_i denotes the i-th elimination ideal of I.

Proposition 8 [6, Sect. 3.2] *Let $I \subset \mathbb{Q}[x_1, \ldots, x_n]$, $f \in \mathbb{Q}[x_1, \ldots, x_n]$, and t be a new indeterminate, then $\overline{V(I) \setminus V(f)} = V((I + \langle tf - 1\rangle) \cap \mathbb{Q}[x_1, \ldots, x_n])$. Moreover, if $G' \subset \mathbb{Q}[t, x_1, \ldots, x_n]$ is a Gröbner basis of $I + \langle tf - 1\rangle$ for an elimination ordering w.r.t. to $[t]$, then $G' \cap \mathbb{Q}[x_1, \ldots, x_n]$ is a Gröbner basis of:*

$$I : f^\infty := (I + \langle t\, f - 1\rangle) \cap \mathbb{Q}[x_1, \ldots, x_n].$$

The variety $\overline{V(I) \setminus V(f)}$ and the ideal $I : f^\infty$ are usually called the localization of $V(I)$ and I by f.

8.5 Certified Solutions of Zero-Dimensional Systems

In this section, we study the case of zero-dimensional systems, that is, systems with finitely many solutions in the algebraic closure of the coefficient field. For such systems, we will see that the quotient algebra of the corresponding ideal is a finite dimensional vector space. This fundamental property allows one to translate most of the questions about zero-dimensional systems into linear algebra questions in the corresponding quotient algebra. These questions can then be answered using classical linear algebra algorithms. Hence, starting from a system of polynomial equations, we can obtain many information about its solutions, e.g., counting their number, computing their symbolic representation or determining their multiplicities.

8.5.1 The Case of One Variable

To give a first idea of the link between zero-dimensional systems and the corresponding quotient algebras, let us start by considering the simple case of univariate polynomials and let us recall a classical result about the computation of the roots of such polynomials.

Given a polynomial in $\mathbb{Q}[x]$, $P(x) = \sum_{i=0}^{D} a_i\, x^i$ with $a_D \neq 0$, the quotient algebra $\frac{\mathbb{Q}[x]}{\langle P\rangle}$ is a $\mathbb{Q}$-vector space of dimension D, in which one can define the endomorphism of the multiplication by x

$$\begin{aligned} m_x : \tfrac{\mathbb{Q}[x]}{\langle P\rangle} &\to \tfrac{\mathbb{Q}[x]}{\langle P\rangle} \\ u &\mapsto \overline{x\, u}, \end{aligned}$$

which sends any u in $\frac{\mathbb{Q}[x]}{\langle P\rangle}$ to the remainder of the Euclidean division of $x\,u$ by P. We denote by $C(P)$ its matrix in the monomial basis $\{1, x, \ldots, x^{D-1}\}$, i.e.:

$$C(f) = \begin{pmatrix} 0 & 0 & 0 & \ldots & -\frac{a_0}{a_D} \\ 1 & 0 & 0 & \ldots & -\frac{a_1}{a_D} \\ 0 & 1 & 0 & \ldots & -\frac{a_2}{a_D} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & -\frac{a_{D-1}}{a_D} \end{pmatrix}.$$

This matrix is known as the *Frobenius companion matrix* of P and its characteristic polynomial is the polynomial P itself.

Theorem 5 *The eigenvalues of $C(P)$ are exactly the roots of $P(x)$ with the same multiplicities.*

Consequently, one can compute the roots of a univariate polynomial $P(x)$ by simply computing the eigenvalues of its Frobenius companion matrix. This example exhibits the role of multiplication endomorphisms for the characterization of the roots of a univariate polynomial. In fact, this approach can be generalized for characterizing the solutions of a zero-dimensional system defined by an ideal I in $\mathbb{Q}[x_1, \ldots, x_n]$. As for the case of one univariate polynomial, the quotient algebra corresponding to I, i.e., $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$ is a finite dimensional $\mathbb{Q}$-vector space, and a basis of it is given by the monomials that are irreducible modulo the ideal I [6]. The dimension of this vector space is the number of solutions of I counted with multiplicities, which we denote by D in the following.

The following result is a generalization of Theorem 5 for the case of ideals in $\mathbb{Q}[x_1, \ldots, x_n]$. The notation $\overline{P}$ denotes the normal form of P with respect to I.

Theorem 6 [7] *Let $h \in \mathbb{Q}[x_1, \ldots, x_n]$ and m_h be the multiplication endomorphism by h*

$$\begin{array}{rccc} m_h : & \frac{\mathbb{Q}[x_1,\ldots,x_n]}{I} & \to & \frac{\mathbb{Q}[x_1,\ldots,x_n]}{I} \\ & u & \mapsto & \overline{hu}. \end{array}$$

The eigenvalues of m_h are $h(\alpha)$, where $\alpha \in V(I)$, with multiplicity $\mu(\alpha)$.

According to Theorem 6, providing a basis $\mathcal{B}$ of $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$ and the matrices of the multiplication m_{x_i} by the variables $x_i, i = 1, \ldots, n$, one can compute all the coordinates of all the solutions $\alpha \in V(I)$. From the computation point of view, when $I \subset \mathbb{Q}[x_1, \ldots, x_n]$, one way to compute $\mathcal{B}$ as well as the matrices m_{x_i} is to use Gröbner bases.

Theorem 7 [6] *Let $I \subset \mathbb{Q}[x_1, \ldots, x_n]$ be a zero-dimensional ideal and G a Gröbner basis of I with respect to any monomial ordering $<$. Then, we have:*

- *For all $i = 1, \ldots, n$, there exists a polynomial $g_j \in G$ and a positive integer n_j such that $x_i^{n_j} = LM_<(g_j)$.*

- $\mathcal{B} := \{t = x_1^{e_1} \cdots x_n^{e_n} | (e_1, \ldots, e_n) \in \mathbb{N}^n \text{ and } e_i \leq n_i\} = \{w_1, \ldots, w_D\}$ *is a basis of* $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$ *as a* $\mathbb{Q}$*-vector space;*

Hence, given a Gröbner basis of a system, simply by looking at the leading terms of the basis, we are able to check if the system is zero-dimensional and, in the latter case, to deduce a basis of the corresponding quotient algebra. However, knowing all the coordinates of all the solutions of $V(I)$ is not sufficient since one needs to combine them suitably in order to get the actual solutions of $V(I)$, which is not an easy task. Alternatively, the usual approach, which we describe in the next section, is to compute a parametrization of the solutions.

Before going further, let mention the following important result which is a multivariate generalization of Hermite's theorem for counting the number of distinct roots of univariate polynomials [7].

Theorem 8 *Let* $h \in \mathbb{Q}[x_1, \ldots, x_n]$ *and* Her_h *be Hermite's quadratic form:*

$$\begin{array}{rccc} Her_h : & \frac{\mathbb{Q}[x_1,\ldots,x_n]}{I} & \to & \mathbb{Q} \\ & f & \mapsto & Trace\left(m_{f^2 h}\right). \end{array}$$

Then, we have:

- $rank(Her_h) = \sharp\{x \in V(I) \mid h(x) \neq 0\}$.
- $signature(Her_h) = \sharp\{x \in V(I) \cap \mathbb{R}^n \mid h(x) > 0\}$
  $-\sharp\{x \in V(I) \cap \mathbb{R}^n \mid h(x) < 0\}$,

where $\sharp$ *denotes the cardinality of a set.*

When $h = 1$, Theorem 8 yields an algorithm for counting the number of solutions in $V(I)$ as well as the number of solutions in $V(I) \cap \mathbb{R}^n$. This algorithm first constructs the matrix associated to Her_1 (the entries of this matrix are the $\mathrm{Trace}(m_{w_i w_j})$ where w_k is an element of $\mathcal{B}$) and then compute its rank (resp. signature) to get the number of solutions in $V(I)$ (resp. the number of solutions in $V(I) \cap \mathbb{R}^n$).

8.5.2 Univariate Representations of the Solutions

Suppose that $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$ is a $\mathbb{Q}$-vector space of dimension D and consider the vectors $1, \overline{x_1}, \ldots, \overline{x_1}^{D-1}$ in this vector space. If the latters are $\mathbb{Q}$-linearly independent, then they form a basis, and we can express $x_1^D, x_2, \ldots, x_n$ in $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$ as a $\mathbb{Q}$-linear combination of them which yields the following parametrization:

$$\begin{cases} f(x_1) = 0, \\ x_2 = g_2(x_1), \\ \quad \vdots \\ x_n = g_n(x_1). \end{cases} \tag{8.6}$$

The polynomial $\{f, x_2 - g_2, \ldots, x_n - g_n\}$ forms a Gröbner basis of I for the lexicographic monomial ordering $<_{\text{lex}}$ with $x_1 <_{\text{lex}} \ldots <_{\text{lex}} x_n$ [6].

Up to an eventual permutation of the variable's index (considering the vectors $1, \overline{x_i}, \ldots, \overline{x_i}^{D-1}$), the case (8.6) is known as the *Shape position* case.

On the other hand, one can consider a polynomial $h \in \mathbb{Q}[x_1, \ldots, x_n]$, a new independent variable t, and define the ideal $I_h := I + \langle t - h \rangle \subset \mathbb{Q}[t, x_1, \ldots, x_n]$ so that $V(I_h) = \{(\alpha, h(\alpha)) \mid \alpha \in V(I)\}$ (one can easily remark that $V(I_h)$ and $V(I)$ are in one-to-one correspondence). If $1, \overline{h}, \ldots, \overline{h}^{D-1}$ are $\mathbb{Q}$-linearly independent in $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I_h}$, then, we can also express $x_1, \ldots, x_n$ in $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I_h}$ as a linear combination of $1, \overline{h}, \ldots, \overline{h}^{D-1}$, which yields the following parameterization:

$$\begin{cases} f(t) = 0, \\ x_1 = g_1(t), \\ \quad \vdots \\ x_n = g_n(t). \end{cases} \tag{8.7}$$

However, in some cases (see the above example), one cannot get parametrizations of the forms (8.6) or (8.7).

Example 3 Consider the ideal $I := \langle x_1^2, x_1 x_2, x_2^2 \rangle$, which is already a Gröbner basis. According to Theorem 7, a basis of $\frac{\mathbb{Q}[x_1,x_2]}{I}$ is then $\mathcal{B}_{<_{\text{lex}}} := \{1, x_1, x_2\}$ and $D = \sharp\mathcal{B} = 3$ (the unique zero is $(0, 0)$ and has multiplicity 3).

As $x_1^2 \in I$ (resp. $x_2^2 \in I$), then $1, x_1, x_1^2$ (resp. $1, x_2, x_2^2$) are trivially $\mathbb{Q}$-linearly dependent in $\frac{\mathbb{Q}[x_1,x_2]}{I}$ and thus neither $1, x_1, \ldots, x_1^{D-1}$ nor $1, x_2, \ldots, x_2^{D-1}$ are linearly independant in $\frac{\mathbb{Q}[x_1,x_2]}{I}$. The ideal is not in *Shape position*. Let now take any $h \in \frac{\mathbb{Q}[x_1,x_2]}{I}$. The general expression of such an element is $h = ax_1 + bx_2 + c$, with $a, b, c \in \mathbb{Q}$, and it immediately turns out that $h^2 - 2ch - c^2 = 0$ in $\frac{\mathbb{Q}[x_1,x_2]}{I}$. Thus, for any $h \in \frac{\mathbb{Q}[x_1,x_2]}{I}$, $1, h, \ldots, h^{D-1}$ are $\mathbb{Q}$-linearly dependent in $\frac{\mathbb{Q}[x_1,x_2]}{I}$ which implies that the ideal I_h cannot be written under the form (8.7).

Mathematically, the two above situations ((8.6) and (8.7)) correspond to the case where the quotient algebra $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$ is cyclic, that is, when it is generated by the successive powers of an element of $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$. Such an element is called a *primitive element* of $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$. When $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$ is known to be cyclic, finding a primitive element is equivalent to finding what is called a *separating element* for the set of points defined by the variety $V(I)$.

Definition 12 Let h be a polynomial in $\mathbb{Q}[x_1, \ldots, x_n]$. Then, h is a separating element for $V(I)$ if and only if $x \in V(I) \mapsto h(x)$ is injective.

In addition a separating element can be found among a finite set of linear forms so as stated in the following theorem.

Theorem 9 *Suppose that* $\sharp V(I) = d$. *Then, the set*

$$Sep_d = \left\{ x_1 + i\, x_2 + \cdots + i^{n-1} x_n, i = 0, \ldots, n\frac{d(d-1)}{2} \right\},$$

contains at least one separating element for $V(I)$.

The computation of such a primitive element can be done by computing for each $h \in \text{Sep}_d$, the minimal integer d_h such that $1, \overline{h}, \ldots, \overline{h}^{d_h}$ are linearly dependent, and then selecting an h for which $d_h = D - 1$. The computation of the parametrization (8.7) then resumes to the computation of the coordinates of the vectors $x_1, x_2, \ldots, x_n$ in the basis $1, \overline{h}, \ldots, \overline{h^{D-1}}$. Note that another methods for obtaining such a parametrization is to compute a Gröbner basis of $I + \langle t - h \rangle$ with respect to the lexicographic monomial ordering $<_{\text{lex}}$ with $t < x_1 <_{\text{lex}} \ldots <_{\text{lex}} x_n$.

As mentioned before, the above strategy for computing a parametrization works only when a primitive element exists (i.e., $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$ is cyclic). This is the case for example when the considered ideal is radical (all the solutions have multiplicity one).

When $\frac{\mathbb{Q}[x_1,\ldots,x_n]}{I}$ is not cyclic, one can still compute a parametrization of the solutions using the so-called *Rational Univariate Representation* (RUR) [13].

Definition 13 Given any $h \in \mathbb{Q}[x_1, \ldots, x_n]$, we define:

- $f_h(t) = \prod_{\alpha \in V(I)} (t - h(\alpha))^{\mu(\alpha)}$,
- $g_{h,1}(t) = \sum_{\alpha \in V(I)} \mu(\alpha) \prod_{\beta \in V(I), \beta \neq \alpha} (t - h(\beta))$,
- $g_{h,v}(t) = \sum_{\alpha \in V(I)} \mu(\alpha)\, v(\alpha) \prod_{\beta \in V(I), \beta \neq \alpha} (t - h(\beta))$ for $v \in \{x_1, \ldots, x_n\}$.

If h separates $V(I)$, then the univariate polynomials $\{f_h(t), g_{h,1}(t), \ldots, g_{h,x_n}(t)\}$ define the so called Rational Univariate Representation of I associated to h.

The Rational Univariate Representation of I bears important properties which we summarize below.

- $f_h(t), g_{h,1}(t), \ldots, g_{h,x_n}(t)$ are polynomials in $\mathbb{Q}[t]$.
- The application

$$\begin{aligned} \phi_h : V(I) &\longrightarrow V(f_h) \\ x &\longmapsto h(x), \end{aligned}$$

  defines a bijection between $V(I)$ and $V(f_h)$, whose reciprocal is given by:

$$\begin{aligned} \phi_h^{-1} : V(f_h) &\longrightarrow V(I) \\ x &\longmapsto \left(\frac{g_{h,x_1}(x)}{g_{h,1}(x)}, \ldots, \frac{g_{h,x_n}(x)}{g_{h,1}(x)} \right). \end{aligned}$$

- ϕ_h preserves the multiplicities : $\mu(h(x)) = \mu(x)$.

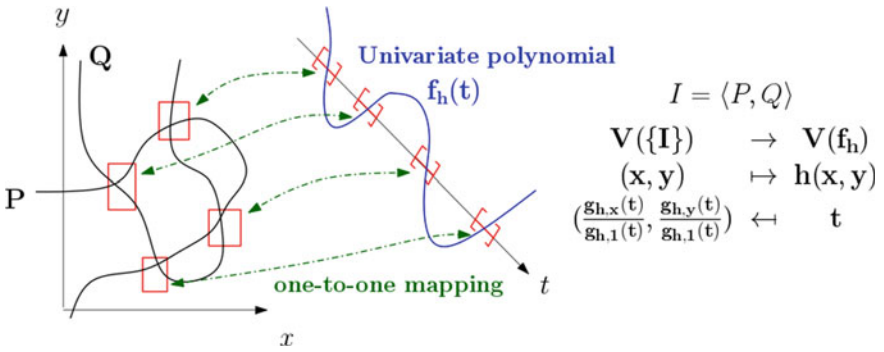


Fig. 8.2 A rational univariate representation of a zero-dimensional bivariate system $I = \langle P, Q \rangle$

The Rational Univariate Representation of an ideal I is a one-to-one mapping between the solutions of $V(I)$ and the roots of a univariate polynomial $f_h(t)$ (see Fig. 8.2). This representation is uniquely defined up to a separating element. Moreover, unlike classical parametrizations, such a representation preserves the multiplicities of the solutions, in the sense that the multiplicity of a solution in I is the multiplicity of the corresponding root in the polynomial $f_h(t)$. The latter property is critical in many problems where the information about the multiplicities is needed.

To compute a RUR, one has to solve the following two problems:

- Find a separating element h.
- Given any polynomial h, compute a RUR-Candidate $f_h, g_{h,1}, g_{h,x_1}, \ldots, g_{h,x_n}$ such that if h is a separating element, then the RUR-Candidate is a RUR.

According to [13], a RUR-Candidate can be explicitly computed when we know a suitable representation of $\mathbb{Q}[x_1, \ldots, x_n]/I$, which can be summarized as follows:

- $f_h = \sum_{i=0}^{D} a_i t^i$ is the characteristic polynomial of m_h. Let us denote by $\overline{f_h}$ its square-free part.
- For any $v \in \mathbb{Q}[X_1, \ldots, X_n]$, $g_{h,v} = g_{h,v}(t) = \sum_{i=0}^{d-1} \text{Trace}(m_{vh^i}) H_{d-i-1}(t)$, $d = \deg(\overline{f_h})$ and $H_j(T) = \sum_{i=0}^{j} a_i t^{i-j}$.

In [13], a strategy is proposed to compute a RUR for any system defined by a Gröbner basis for any ordering.

8.5.2.1 Applications of the Rational Univariate Representation

From Formal to Numerical Solutions. Computing a RUR reduces the resolution of a zero-dimensional system to solving a polynomial f_h with one variable and to evaluating n rational fractions $(\frac{g_{h,x_i}(t)}{g_{h,1}(t)}, i = 1 \ldots n)$ at the roots of f_h. The goal is thus

to compute all the real roots of f_h providing a numerical approximation with an arbitrary precision of the coordinates.

The isolation of the real roots of f_h can be done using the algorithm proposed in [9]. The output will be a list l_{f_h} of intervals with rational bounds such that for each real root α of f_h, there exists a unique interval in l_{f_h} which contains α. The second step consists in refining each interval in order to ensure that it does not contain any real root of $g_{h,1}$. Since f_h and $g_{h,1}$ are coprime, this computation is easy. Then, we can ensure that the rational functions can be evaluated by using interval arithmetics without any cancelation of the denominator. The last evaluation is performed by using multi-precision arithmetics (MPFI package—[14]). Moreover, the rational functions defined by the RUR are stable under numerical evaluation even if their coefficients are huge rational numbers. Thus, the isolation of the real roots does not involve huge compaction burden. To increase the precision of the result, it is only necessary to decrease the length of the intervals in l_{f_h} which can be easily done by bisection or using a certified Newton's algorithm. It is in particular quite simple to certify the sign of the coordinates.

Signs of Polynomials at the Roots of a System. Due to the presence of inequalities in semi-algebraic system, it is important to develop a method for computing the sign ($+, -, \neq$ or 0) of given multivariate polynomials at the real roots of a zero-dimensional system. Having a RUR $\{f_h, g_{h,1}, g_{h,x_1}, \ldots, g_{h,x_n}\}$ of I, one can translate the problem of computing the sign of a multivariate polynomial into a problem of computing the sign of a univariate polynomial. Indeed, let $P \in \mathbb{Q}[x_1, \ldots, x_n]$ be the polynomial to be evaluated at the real solution $\alpha = (\alpha_1, \ldots, \alpha_n) \in V(I)$ (α is the image of a root γ of $f_h(t)$ by the RUR mapping). One can define the polynomial $P_I(t)$ roughly as the numerator of the rational fraction $P\left(\frac{g_{h,x_1}}{g_{h,1}}, \ldots, \frac{g_{h,x_n}}{g_{h,1}}\right)$, that is the rational fraction obtained after substituting in P each variable x_i by $\frac{g_{h,x_i}}{g_{h,1}}$. Then the following result holds.

Theorem 10 *The sign of $P(x_1, \ldots, x_n)$ at the real solution $\alpha = (\alpha_1, \ldots, \alpha_n) \in V(I)$ is equal to the sign of $P_I(t)$ at the corresponding root γ of $f_h(t)$ via the RUR mapping.*

Accordingly, the problem of computing the sign of $P(x_1, \ldots, x_n)$ at a solution of $V(I)$ is reduced to the problem of computing the sign of $P_I(t)$ at a real root of $f_h(t)$. To solve the latter problem, a naive algorithm consists in isolating the real root of $f_h(t)$, so that the interval is also isolating for the product $P_I(t)\, f_h(t)$ and then evaluating the sign of $P_I(t)$ at the endpoints of this interval.

Consequently, in order to compute the sign of the polynomial $P(x_1, \ldots, x_n)$ at a solution of $V(I)$, it is sufficient to compute the sign of the polynomial $P_I(t)$ at a given root of $f_h(t)$.

Instead of straightforwardly *plugging* the formal coordinates provided by the RUR into P, we better extend the RUR by computing rational functions which coincide with the values of P at the roots of I. This can be done by using the general formula $g_{h,P} = \sum_{i=0}^{D-1} \mathrm{Trace}(m_{P\,h^i}) H_{D-i-1}(t)$ given in [13]. One can directly compute

the Trace(Pt^i) by reusing the computations already done if the RUR has already been computed. Hence, it is not more costly to compute the extended RUR than the classical one.

8.5.3 *Testing Structural Stability: The Zero-Dimensional Case*

In the following, we are going to show how Rational Univariate Representations can be used in order to solve the stability test problem mentioned in the introduction. For one or two dimensional systems, the test of the structural stability can be reduced to the study of algebraic zero-dimensional systems. Indeed, in the case of one dimensional system the stability condition translates into

$$D(z) \neq 0 \text{ for } |z| \leq 1,$$

or equivalently, the subset of $\mathbb{C}$ defined by $E := \{z \in \mathbb{C} \mid D(z) = 0, |z| \leq 1\}$ is empty. The set E can be viewed as a semi-algebraic set of $\mathbb{R}^2$. Indeed, if we note $z = x + i\,y$, where x (resp., y) is the real part (resp., the imaginary part) of z and i the imaginary unit, then the polynomial $D(z)$ can be rewritten as $D(x, y) = \mathcal{R}(x, y) + i\,\mathcal{I}(x, y)$, where $\mathcal{R}, \mathcal{I} \in \mathbb{Q}[x, y]$, and the inequality $|z| \leq 1$ as $x^2 + y^2 \leq 1$, which shows that:

$$E \approx \{(x, y) \in \mathbb{R}^2 | \mathcal{R}(x, y) = 0,\ \mathcal{I}(x, y) = 0, x^2 + y^2 \leq 1\}.$$

Then, the problem of testing the stability reduces to that of testing that the above semi-algebraic set does not have real solutions. Without loss of generality the system $S := \{\mathcal{R}(x, y) = 0,\ \mathcal{I}(x, y) = 0\}$ can be assumed to be zero-dimensional (i.e., has a finite number of complex solutions). In that case, the problem resumes to compute the sign of the real solutions of S at the polynomial $x^2 + y^2 - 1$.

Example 4 We consider the polynomial $D(z)=\frac{3}{2}z^5 - \frac{27}{2}z^4 + \frac{57}{2}z^3 + \frac{7}{2}z^2 - \frac{9}{2}z + \frac{1}{2}$. We first compute the zero-dimensional system S whose real solutions are in bijection with the complex roots of $D(z)$:

$$S := \begin{cases} \mathcal{R}(x, y) = \frac{3}{2}x^5 - 15x^3y^2 + \frac{15}{2}x\,y^4 - \frac{27}{2}\,x^4 + 81x^2y^2 - \frac{27}{2}y^4 + \frac{57}{2}x^3 \\ \qquad\qquad -\frac{171}{2}x\,y^2 + \frac{7}{2}x^2 - \frac{7}{2}y^2 - \frac{9}{2}x + \frac{1}{2} = 0, \\ \mathcal{I}(x, y) \ = \frac{171}{2}x^2y - \frac{9}{2}\,y - 15\,x^2y^3 + \frac{3}{2}\,y^5 + 54\,x\,y^3 - \frac{57}{2}y^3 + 7xy \\ \qquad\qquad -54x^3y + \frac{15}{2}\,x^4y = 0. \end{cases}$$

The system S is zero-dimensional and we can compute a Rational Univariate Representation of its solutions using the formulas given in Sect. 8.5.2 which yields:

$$
\begin{aligned}
f(t) &= 559872\,t^{25} - 25194240\,t^{24} + 544195584\,t^{23} - 7493513472\,t^{22} + 73628346816\,t^{21} \\
&\quad -547311691584t^{20} + 3183535332864t^{19} - 14780593319616t^{18} + 55362880574208t^{17} \\
&\quad -167896649845440\,t^{16} + 411029639424576\,t^{15} - 804050295433200\,t^{14} \\
&\quad +1232226241447500\,t^{13} - 1428873627636324\,t^{12} + 1177034305128192\,t^{11} \\
&\quad -6034409182022760\,t^{10} + 1261878032504435\,t^{9} + 22809165295113\,t^{8} \\
&\quad -11098557635568\,t^{7} + 17376699104892\,t^{6} - 9925212685221\,t^{5} + 2611676368585\,t^{4} \\
&\quad -821059361472\,t^{3} + 262536537420\,t^{2} - 42350188473\,t + 2455046453, \\
g_1(t) &= 13996800\,t^{24} - 604661760\,t^{23} + 12516498432\,t^{22} - 164857296384\,t^{21} \\
&\quad +1546195283136\,t^{20} - 10946233831680\,t^{19} + 60487171324416\,t^{18} \\
&\quad -266050679753088\,t^{17} + 941168969761536\,t^{16} - 2686346397527040\,t^{15} \\
&\quad +6165444591368640\,t^{14} - 11256704136064800\,t^{13} + 16018941138817500\,t^{12} \\
&\quad -17146483531635888\,t^{11} + 12947377356410112\,t^{10} - 6034409182022760\,t^{9} \\
&\quad +1135690229253987\,t^{8} + 182473322360904\,t^{7} - 77689903448976\,t^{6} \\
&\quad +104260194629352\,t^{5} - 49626063426105\,t^{4} + 10446705474340\,t^{3} \\
&\quad -2463178084416\,t^{2} + 525073074840\,t - 42350188473, \\
g_x(t) &= 25194240\,t^{24} - 1050879744\,t^{23} + 20976724224\,t^{22} - 265852699776\,t^{21} \\
&\quad +2391835843008\,t^{20} - 16175589523776\,t^{19} + 84921114868416\,t^{18} \\
&\quad -352340187356736\,t^{17} + 1164594239224128\,t^{16} - 3065803125993360\,t^{15} \\
&\quad +6371804589628464\,t^{14} - 10251200537235576\,t^{13} + 12302401061993148\,t^{12} \\
&\quad -10249204642846020\,t^{11} + 4995304129178172\,t^{10} - 576047210865300\,t^{9} \\
&\quad -590896493514297\,t^{8} + 232387793555778\,t^{7} - 215336160313290\,t^{6} \\
&\quad +124704312574422\,t^{5} - 32799357684699\,t^{4} + 9758271572934\,t^{3} \\
&\quad -3373050489686\,t^{2} + 598205563056\,t - 37550186449, \\
g_y(t) &= -37511424\,t^{23} + 1503816192\,t^{22} - 28660687488\,t^{21} + 344722614912\,t^{20} \\
&\quad -2925622473408\,t^{19} + 18543038368896\,t^{18} - 90562857236928\,t^{17} \\
&\quad +346475609384832\,t^{16} - 1044493268252400\,t^{15} + 2472748660136736\,t^{14} \\
&\quad -4535514360134424\,t^{13} + 6272956097279064\,t^{12} - 6229275628948668\,t^{11} \\
&\quad +4056309643855968\,t^{10} - 1442957641141788\,t^{9} + 203140683497376\,t^{8} \\
&\quad -32613756115554\,t^{7} - 114821122679658\,t^{6} + 73799941129998\,t^{5} \\
&\quad -22045846055586\,t^{4} + 8305034379450\,t^{3} - 2665289870974\,t^{2} + 418198960296\,t \\
&\quad -23825974876.
\end{aligned}
$$

Isolating numerically the real roots of $f(t)$ and substituting in $\frac{g_x(t)}{g_1(t)}$ and $\frac{g_y(t)}{g_1(t)}$ yields the following five real solutions:

$$
[x = -0.45367372,\, y = 0],\, [x = 0.14614706,\, y = 0],\, [x = 0.25639717,\, y = 0],
$$
$$
[x = 3.59132461,\, y = 0],\, [x = 5.45980486,\, y = 0].
$$

and we can easily remark (without further symbolic computations) that the three first solutions correspond to the roots of $D(z)$ that are inside the unit disk while the two last solutions correspond to the roots of $D(z)$ that are outside the unit disk, which implies that the system is not stable.

In the case of two dimensional systems, according to DeCarlo et al. [15], the structural stability condition, i.e.

$$
D(z_1, z_2) \neq 0 \text{ for } |z_1| \leq 1, |z_2| \leq 1,
$$

is equivalent to:

$$\begin{cases} D(z_1, 1) \neq 0 \text{ for } |z_1| \leq 1, \\ D(1, z_2) \neq 0 \text{ for } |z_2| \leq 1, \\ D(z_1, z_2) \neq 0 \text{ for } |z_1| = |z_2| = 1. \end{cases}$$

The two first conditions can easily be tested using classical stability tests (see for instance [16]), or the method presented above. For the last condition, if we note $z_j = x_j + i\, y_j$ testing the latter resumes to test that the following system

$$S := \begin{cases} \mathcal{R}(x_1, y_1, x_2, y_2) = 0, \\ \mathcal{I}(x_1, y_1, x_2, y_2) = 0, \\ x_1^2 + y_1^2 - 1 = 0, \\ x_2^2 + y_2^2 - 1 = 0, \end{cases}$$

where $D(x_1, y_1, x_2, y_2) = \mathcal{R}(x_1, y_1, x_2, y_2) + i\,\mathcal{I}(x_1, y_1, x_2, y_2)$, does not have real solutions. The system S consists of four polynomials in four variables and is generically zero-dimensional. One can thus compute the corresponding Rational Univariate Representation and use it to check the existence of real solutions.

Example 5 We consider the polynomial $D(z_1, z_2)=(12 + 10z_1 + 2z_1^2) + (6 + 5z_1 + z_1^2)z_2$ which is shown to be devoid from complex zero in $\mathbb{D}^2$ [17]. This polynomial yields the following zero-dimensional system

$$S := \begin{cases} R(x_1, y_1, x_2, y_2) = x_1^2 x_2 - 2x_1\, y_1 y_2 - y_1^2 x_2 + 2\, x_1^2 + 5x_1 x_2 - 2y_1^2 \\ \qquad\qquad -5\, y_1 y_2 + 10\, x_1 + 6x_2 + 12 = 0, \\ C(x_1, y_1, x_2, y_2) = x_1^2 y_2 + 2\, x_1 y_1 x_2 - y_1^2 y_2 + 4\, x_1 y_1 + 5\, x_1 y_2 \\ \qquad\qquad +5\, y_1 x_2 + 10\, y_1 + 6\, y_2 = 0, \\ x_1^2 + y_1^2 - 1 \qquad = 0, \\ x_2^2 + y_2^2 - 1 \qquad = 0, \end{cases}$$

whose solutions are encoded by the following Rational Univariate Representation:

$$\begin{aligned} &f(t) = 144\, t^4 + 337\, t^2 + 144, \; g_1(t) = 1152\, t^3 + 1348\, t, \\ &g_{x_1}(t) = -1680\; t^3 - 1820\, t, \quad g_{y_1}(t) = -1348\, t^2 - 1152, \\ &g_{x_2}(t) = -1440\, t^3 - 1685\, t, \quad g_{y_2}(t) = 900\, t^2 + 900. \end{aligned}$$

Performing numerical isolation on the polynomial $f(t)$, we obtain that it does not admit real roots, which implies that the system S does not have real solutions, and thus that the initial system is stable.

Note finally that one can avoid doubling the number of variables by opting for special transformations such as Möbius transformation (see [18] for details).

8.6 Real Roots of Positive Dimensional Systems

In this section, we review the principal approaches for studying systems of polynomials equations that admit an infinite number of complex zeros. As mentioned in the introduction, various questions can be asked about the zero set of such systems: deciding the emptiness, computing points in each connected component of the variety, etc.

We distinguish between two general strategies. The first one, which is described in the next section, is based on the classical Cylindrical Algebraic Decomposition (CAD) algorithm [19]. This algorithm, based on variable elimination, one after the other, provides a partition of the real space into cells in which the given polynomials keep their sign constant. It allows one to answer to more general questions such as deciding the truth of a first order formula, quantifier elimination, etc. However, its complexity, which is doubly exponential in the number of variable turns out to be its Achilles' heel, and prevents it from being used for system with more than two variables. The second strategy, described briefly in Sect. 8.6.2, is based on the determination of a function that reaches its extremum (at a finite number of points), on each connected component of the studied set. Putting in equation these extremum then allows one to reduce the problem to the study of zero-dimensional systems. These methods are referred as the critical point methods and lead to algorithms that have a single exponential complexity in the number of variables.

8.6.1 Cylindrical Algebraic Decomposition

Let start with some definitions that are used in the sequel.

A *semi-algebraic set* of $\mathbb{R}^n$ is a set of $\mathbb{R}^n$ that satisfies a logical combination of polynomial equations and inequalities with real coefficients. The set of semi-algebraic sets forms the smallest class $\mathcal{SA}_n$ of sets in $\mathbb{R}^n$ such that:

- If $P \in \mathbb{R}[x_1, \ldots, x_n]$, then $\{x \in \mathbb{R}^n \mid P(x) = 0\} \in \mathcal{SA}_n$.
- If $A \in \mathcal{SA}_n$ and $B \in \mathcal{SA}_n$, then $A \cup B$, $A \cap B$ and $\mathbb{R}^n \setminus A$ are in $\mathcal{SA}_n$.

Proposition 9 *Any semi-algebraic set of* $\mathbb{R}^n$ *is the union of a finite number of semi-algebraic sets of the form* $\{x \in \mathbb{R}^n \mid P(x) = 0,\ Q_1(x) > 0, \ldots,\ Q_l(x) > 0\}$, *where* $l \in \mathbb{N}$, *and* $P, Q_1, \ldots, Q_l \in \mathbb{R}[x_1, \ldots, x_n]$.

Definition 14 A function from $A \subset \mathbb{R}^m$ to $B \subset \mathbb{R}^n$ is *semi-algebraic* if the corresponding graph is semi-algebraic.

One knows that semi-algebraic sets of $\mathbb{R}$ decompose into an union of a finite number of points and open intervals. More generally, semi-algebraic sets of $\mathbb{R}^n$ decompose into a disjoint union of cells that are isomorphic to open hypercubes of different dimensions.

The demonstration of this property can be done by exhibiting an algorithm which, given a set of polynomials, decomposes $\mathbb{R}^n$ in cells where the sign of these polynomials is invariant.

The resulting decomposition allows one to answer several questions about the zero of the system among which for example: Does the system admit real solutions?

Definition 15 A *Cylindrical Algebraic Decomposition* of $\mathbb{R}^n$ is a sequence $C_1, \ldots, C_n$ such that each C_i is a partition of $\mathbb{R}^i$ in a finite number of semi-algebraic sets satisfying:

(a) Each cell C of C_1 is either a point or an open interval.
(b) For any $1 \leq k < n$ and any $C \in C_k$, there exists a finite number of continuous semi-algebraic functions $\Psi_{C,1} < \ldots < \Psi_{C,l_C} : C \longrightarrow \mathbb{R}$ such that the cylinder $C \times \mathbb{R}$ is the disjoint union of cells in C_{k+1} that are:

- either the graph of one the function Ψ_{C,i_C} :

$$A_{C,j} = \{(x', x_{k+1}) \in \mathbb{C} \times \mathbb{R} \mid x_{k+1} = \Psi_{C,j}(x')\},$$

- or the section of the cylinder bounded by the functions $\Psi_{C,j}$ et $\Psi_{C,j+1}$:

$$B_{C,j} = \{(x', x_{k+1}) \in \mathbb{C} \times \mathbb{R} \mid \Psi_{C,j}(x') < x_{k+1} < \Psi_{C,j+1}(x')\}.$$

Proposition 10 *Every cell of a Cylindrical Algebraic Decomposition of* $\mathbb{R}^n$ *is semi-algebraically homeomorphic to an open hypercube of the form* $(0, 1)^k$.

Given a set of polynomials F, a subset S of $\mathbb{R}^n$ is said to be *F-invariant* if the sign of each polynomial in F is constant inside S. In the following, we are going to show how to compute a Cylindrical Algebraic Decomposition adapted to a set of polynomial F, that is a decomposition of $\mathbb{R}^n$ into cells that are F-invariant. The resulting Cylindrical Algebraic Decomposition is then said to be F-invariant.

Example 6 Consider the polynomial $f = x - y^2 - 1$. We provide a Cylindrical Algebraic Decomposition adapted to f, that is, a partition of $\mathbb{R}^2$ into cells that are f-invariant (see Fig. 8.3).

The latter is given by the sequence C_1, C_2 where:

- C_1 is the partition of $\mathbb{R}$ that consists of $]-\infty, 1[, \{1\},]1, +\infty[$.
- C_2 is the partition of $\mathbb{R}^2$ that consists of the following semi-algebraic set:
  - $C_{2,1} = \{(x, y) \in \mathbb{R}^2 | x < 1\}$,
  - $C_{2,2} = \{(x, y) \in \mathbb{R}^2 | x = 1, y < 0\}$,
  - $C_{2,2} = \{(x, y) \in \mathbb{R}^2 | x = 1, y = 0\}$,
  - $C_{2,3} = \{(x, y) \in \mathbb{R}^2 | x = 1, y > 0\}$,
  - $C_{2,4} = \{(x, y) \in \mathbb{R}^2 | x > 1, x^2 - y^2 - 1 > 0, y < 0\}$,
  - $C_{2,5} = \{(x, y) \in \mathbb{R}^2 | x > 1, x^2 - y^2 - 1 = 0, y < 0\}$,
  - $C_{2,6} = \{(x, y) \in \mathbb{R}^2 | x > 1, x^2 - y^2 - 1 < 0\}$,

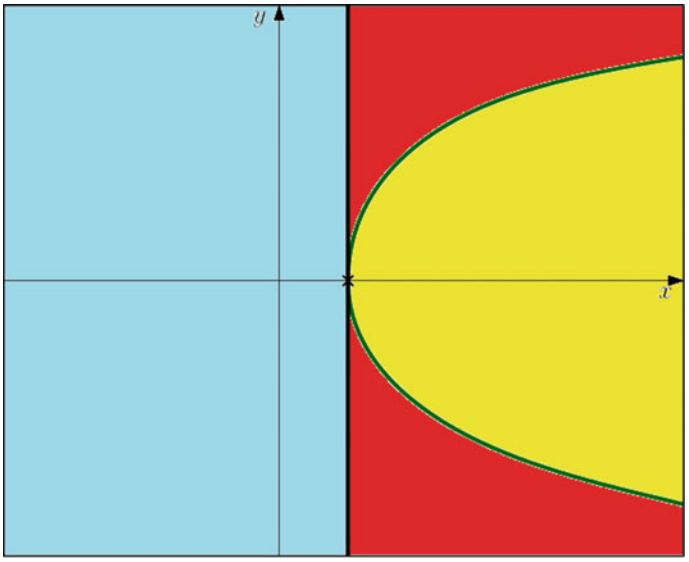


Fig. 8.3 Decomposition of $\mathbb{R}^2$ in $(x - y^2 - 1)$-invariant cells

- $C_{2,7} = \{(x, y) \in \mathbb{R}^2 | x > 1, x^2 - y^2 - 1 = 0, y > 0\}$,
- $C_{2,8} = \{(x, y) \in \mathbb{R}^2 | x > 1, x^2 - y^2 - 1 > 0, y > 0\}$,

and each cell $C_{2,i}$ for $i = 1, \ldots, 9$ is f-invariant.

If we have a look to the interval $]1, +\infty[$, the corresponding cylinder, that is, $\mathcal{C} :=]1, +\infty[\times \mathbb{R}$, is decomposed by means of the following semi-algebraic functions:

- $\Psi_{\mathcal{C},1} :]1, +\infty[\to \mathbb{R}$
  $y \mapsto -\sqrt{x - 1}$,
- $\Psi_{\mathcal{C},2} :]1, +\infty[\to \mathbb{R}$
  $y \mapsto \sqrt{x - 1}$.

More generally, we have the following result.

Proposition 11 *Let $P(x_1, \ldots, x_n) \in \mathbb{R}[x_1, \ldots, x_n]$, $C \subset \mathbb{R}^{n-1}$ be a connected semi-algebraic set and $k \leq d$ a positive integer such that for each point $\alpha = (\alpha_1, \ldots, \alpha_{n-1}) \in C$, the polynomial $P(\alpha, x_n)$ has degree d and admits exactly k complex roots. Then, there exist $l \leq k$ continuous semi-algebraic functions $\Psi_1 < \cdots < \Psi_l : C \longrightarrow \mathbb{R}$, such that for each $\alpha \in C$, the set of real roots of $P(\alpha, x_n)$ is exactly $\{\Psi_1(\alpha), \ldots, \Psi_l(\alpha)\}$. Moreover, for $i = 1 \ldots l$, the multiplicity of the roots $\Psi_i(\alpha)$ is constant for $\alpha \in C$.*

Let now consider a set of polynomials. We need to obtain results about the relative positions of their zeros. A basic result is the following.

Proposition 12 *Let P and Q be two polynomials in $\mathbb{R}[x_1, \ldots, x_n]$ and C a connected component of a semi-algebraic set of $\mathbb{R}^{n-1}$. Let suppose that the degree and*

the number of distinct complex roots of P and Q are constant above C and so that for their gcd (finite number of common solutions). Let $\xi, \zeta : C \to \mathbb{R}$ be two continuous semi-algebraic functions such that $P(\alpha, \xi(\alpha)) = 0$ and $Q(\alpha, \zeta(\alpha)) = 0$ for all $\alpha \in C$. If there exists $\beta \in C$ such that $\xi(\beta) = \zeta(\beta)$, then $\xi(\alpha) = \zeta(\alpha)$ for all $\alpha \in C$.

The two above propositions allow us to construct semi-algebraic functions that have the same properties as the functions used in a CAD of $\mathbb{R}^n$. These functions are actually the roots of P and Q with respect to the last variable. Hence, we almost reach the initial objective since outside these semi-algebraic functions, and under the hypotheses of the above propositions, the sign of P and Q is constant. It remains thenceforth to address the cases where the hypotheses of the propositions are not satisfied, that is:

- The components where the degree of P and Q varies, i.e., where the leading term vanishes; In that case, we need to perform the same operation on P (resp. Q) deprived from its leading term.
- The components where the degree of the gcd of P and Q varies, i.e., where the resultant of these polynomials vanishes.

Definition 16 Let $P_1, \ldots, P_r$ be polynomials in $\mathbb{R}[x_1, \ldots, x_n]$. We denote by $\text{PROJ}(P_1, \ldots, P_r)$ the minimal set of polynomials in $\mathbb{R}[x_1, \ldots, x_n]$ that satisfies the following conditions:

- If $\deg_{x_n}(P_i) = d \leq 2$, $\text{PROJ}(P_1, \ldots, P_r)$ contains all the non constant polynomials among the principal subresultants (Definition 6), $\sigma_j(P_i, \frac{\partial P_i}{\partial x_n})$, $j = 0, \ldots, d-1$ (variations of the number of roots of P_i).
- If $1 \leq d = \min(\deg_{x_n}(P_i), \deg_{x_n}(P_k))$, $\text{PROJ}(P_1, \ldots, P_r)$ contains all the non constant polynomials among the principal subresultants $\sigma_j(P_i, P_k)$, $j = 0, \ldots, d$ (variation of the number of common roots of two polynomials).
- If $\deg_{x_n}(P_i) \geqslant 1$ and $\text{lc}_{x_n}(P_i)$ is not constant, $\text{PROJ}(P_1, \ldots, P_r)$ contains $\text{lc}_{x_n}(P_i)$ and the set $\text{PROJ}(P_1, \ldots, P_r, \text{Trunc}(P_i))$[1] (case of non constant polynomials in x_n whose the leading term vanishes).
- If $\deg_{X_n}(P_i) = 0$ and P_i non constant, $\text{PROJ}(P_1, \ldots, P_r)$ contains P_i (constant polynomials in x_n).

A direct consequence of the propositions stated above is the following theorem.

Theorem 11 *Let $\{P_1, \ldots, P_r\}$ be a set of polynomials in $\mathbb{R}[x_1, \ldots, x_n]$ and C a connected semi-algebraic set $(P_1, \ldots, P_r)$-invariant. Then, there exist continuous semi-algebraic functions $\Psi_1 < \cdots < \Psi_l : C \to \mathbb{R}$ such that for any $\alpha \in C$, the set of $\{\Psi_1(\alpha), \ldots, \Psi_l(\alpha)\}$ is the set of roots of non-identically zero polynomials in $\{P_1, \ldots, P_r\}$. The graph of each Ψ_i and the sections of the cylinder $C \times \mathbb{R}$ bounded by the graphs of Ψ_i and Ψ_{i+1}, $i = 1, \ldots, l-1$ are connected semi-algebraic sets, homeomorphic to C or $C \times (0, 1)$ respectively, and $(P_1, \ldots, P_r)$-invariants.*

[1] $\text{Trunc}(P_i)$ refers to the polynomial obtained after reducing all the coefficients of P_i modulo $\text{lc}_{x_n}(P_i)$.

Having constructed a CAD of $\mathbb{R}^{n-1}$ adapted to $\{P_1, \ldots, P_r\}$, the above theorem allows us to extend the latter to a CAD of $\mathbb{R}^n$ adapted to $\{P_1, \ldots, P_r\}$. By iteratively constructing the set PROJ(.) from $P_1, \ldots, P_r$ (i.e. PROJ(PROJ(...))), one ends up, after $n-1$ steps, with a finite set of polynomials in x_1. The final step then consists in computing a CAD for these univariate polynomials. The real roots of these polynomials decompose the real axis into a finite number of points and open intervals. This algorithmic construction proves the following general result.

Theorem 12 *For any set of polynomials* $\{P_1, \ldots, P_r\}$ *in* $\mathbb{R}[x_1, \ldots, x_n]$*, there exists a CAD of* $\mathbb{R}^n$ *adapted to* $\{P_1, \ldots, P_r\}$.

Cylindrical Algebraic decomposition is implemented in most computer algebraic softwares such as `Maple` (in the package RegularChains[SemiAlgebraicSetTools]) or `Mathematica`.

8.6.1.1 CAD for Testing Structural Stability

Let us go back to the problem of testing the structural stability of multidimensional system and show how Cylindrical Algebraic Decomposition can be used in this context. Checking the structural stability of a two dimensional systems, i.e. $D(z_1, z_2) \neq 0$ for $|z_1| \leq 1$, $|z_2| \leq 1$, can be reduced, via the transformations $z_i = x_i + i\, y_i$, to testing that the following semi-algebraic set is empty.

$$S := \begin{cases} \mathcal{R}(x_1, y_1, x_2, y_2) = 0, \\ \mathcal{I}(x_1, y_1, x_2, y_2) = 0, \\ x_1^2 + y_1^2 \leq 1, \\ x_2^2 + y_2^2 \leq 1. \end{cases}$$

This can be done by computing a Cylindrical Algebraic Decomposition of $\mathbb{R}^4$ adapted to the polynomials $\mathcal{R}$, $\mathcal{I}$ and $x_i^2 + y_i^2 - 1$ for $i = 1, 2$, and then check if this decomposition contain cells satisfying the sign conditions of S.

Example 7 We consider the polynomial $D(z_1, z_2) = 6 + 5\, z_1 + z_2$. After transformation, the latter yields the following semi-algebraic set:

$$S := \begin{cases} \mathcal{R}(x_1, y_1, x_2, y_2) = 5\, x_1 + x_2 + 6 = 0, \\ \mathcal{I}(x_1, y_1, x_2, y_2) = 5\, y_1 + y_2 = 0, \\ x_1^2 + y_1^2 \leq 1, \\ x_2^2 + y_2^2 \leq 1. \end{cases}$$

Computing a CAD adapted to $\mathcal{I}$, $\mathcal{R}$, $x_1^2 + y_1^2 - 1$, $x_2^2 + y_2^2 - 1$ returns (after 2/3 min of computations) 1717 cells. Among these cells, 177 satisfy the above conditions which correspond to the real zeros of the system S. This implies that the input system is not stable.

If we consider the polynomial $D(z_1, z_2) = 2 - z_1 z_2$, the CAD associated with the polynomials of the corresponding system S returns (after 30 min of computations) 31655 cells and outputs 3687 real points satisfying the condition of S. Again the system is not stable.

In practice, we can observe that when the polynomial D is bivariate with total degree larger than 2 or has more than 2 variables (which yields semi-algebraic systems with at least six variables), the previous CAD-based approach fails to return an answer in a reasonable time. This is mainly due to the size of the output (the number of cells) which is doubly exponential in the number of variables. However, when we are only interested in deciding the emptiness of a real semi-algebraic set, this doubly exponential behavior can be overcome by opting for alternative methods, which we will describe in the next section.

8.6.2 Critical Point Methods

When we are only interested in deciding if a system of positive complex dimension has (or not) real roots, the Cylindrical Algebraic Decomposition might answer but this algorithm has a prohibitive complexity while it computes too much information. Alternatively, the so-called critical point methods allow one to compute at least one point in each semi-algebraically connected component of the studied semi-algebraic set and turn out to be, in general, much more efficient in practice.

Critical point methods are essentially based on the determination of a function that reaches its extrema (at a finite number of points), on each connected component of the studied set. Putting in equations these extremum then allow one to reduce the problem to the study of zero-dimensional systems which can be done using the algorithms described in Sect. 8.5 (which are known to be in a complexity that is single exponential in the number of variables). For a sake of simplicity, in the sequel, we will only consider the case of algebraic sets even if such methods can easily be extended to the case of semi-algebraic sets.

Let us start with some definitions needed in the sequel.

Definition 17 Let $V \subset \mathbb{C}^n$ be an algebraic variety and denote by $\mathcal{I}(V)$ the corresponding radical ideal (the set of polynomials that vanish on V).

- If f is a polynomial in $\mathbb{Q}[x_1, \ldots, x_n]$, the differential of f at a point $\alpha = (\alpha_1, \ldots, \alpha_n)$, denoted by $d_\alpha(f)$, is defined by:

$$d_\alpha(f) = \frac{\partial f}{\partial x_1}(x_1 - \alpha_1) + \cdots + \frac{\partial f}{\partial x_n}(x_n - \alpha_n).$$

- The tangent space of V at a point p, denoted by $T_\alpha(V)$, is the points of $\mathbb{C}^n$ on which the differential $d_\alpha(f)$ vanishes for all $f \in \mathcal{I}(V)$.

Definition 18 Let $V \subset \mathbb{C}^n$ be an algebraic variety, and $\varphi_1, \ldots, \varphi_s$ polynomials in $\mathbb{Q}[x_1, \ldots, x_n]$. Define the following polynomial application:

$$\begin{aligned} \varphi : V &\longrightarrow \mathbb{C}^m \\ \alpha &\longmapsto (\varphi_1(\alpha), \ldots, \varphi_m(\alpha)). \end{aligned}$$

- The set of critical points of φ restricted to V is the set of points of V such that the differential map $d_\alpha(\varphi) : T_\alpha(V) \longrightarrow \mathbb{C}^m$ is not surjective, or in other words, such that the rank of $d_\alpha(\varphi)$ is strictly smaller than m.

A fundamental result concerns the critical points of an application restricted to a *compact*[2] algebraic variety.

Theorem 13 [7] *Let $V \subset \mathbb{C}^n$ be a compact algebraic variety and $\varphi : V \longrightarrow \mathbb{C}^m$ a polynomial application. Then the set of the critical points of φ restricted to V intersect $V \cap \mathbb{R}^n$ in each of its connected components.*

In some simple cases, one can easily derive an algebraic characterization of the set of critical points of an application restricted to a variety. Indeed, given an algebraic variety $V \subset \mathbb{C}^n$ whose the corresponding radical ideal $\mathcal{I}(V)$ is generated by a finite number of polynomials $f_1, \ldots, f_s$. The tangent space at each point $p \in V$, $T_\alpha(V)$, is defined as the kernel of the linear application defined by the following matrix, which corresponds to the evaluation at α of the Jacobian matrix associated with the polynomials $f_1, \ldots, f_s$, namely:

$$\mathrm{Jac}(f_1, \ldots, f_s)_\alpha := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\alpha) & \cdots & \frac{\partial f_1}{\partial x_n}(\alpha) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_s}{\partial x_1}(\alpha) & \cdots & \frac{\partial f_s}{\partial x_n}(\alpha) \end{pmatrix}.$$

On the other hand, given a polynomial application $\varphi : V \longrightarrow \mathbb{C}^m$, the differential of φ at a point $\alpha \in V$ is the linear application which associates to each vector $v = (v_1, \ldots, v_n) \in T_\alpha(V)$, the vector $(d_\alpha(\varphi_1)(v), \ldots, d_\alpha(\varphi_m)(v))$, and whose matrix is defined by:

$$\mathrm{Jac}(\varphi_1, \ldots, \varphi_m)_\alpha := \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1}(\alpha) & \cdots & \frac{\partial \varphi_1}{\partial x_n}(\alpha) \\ \vdots & \vdots & \vdots \\ \frac{\partial \varphi_m}{\partial x_1}(\alpha) & \cdots & \frac{\partial \varphi_m}{\partial x_n}(\alpha) \end{pmatrix}.$$

A point α is said to be *critical* for φ if the rank of the above matrix is strictly smaller than m or in other words if its kernel has dimension larger or equal than one. Consequently, α is a critical point if there exists $(v_1, \ldots, v_n) \neq (0, \ldots, 0)$ such that

[2]Here, the term compact is used for subsets of the Euclidean space $\mathbb{R}^n$, which are closed and bounded regarding to the classical Euclidean topology.

$$\begin{cases} \frac{\partial \varphi_1}{\partial x_1}(\alpha)\, v_1 + \cdots + \frac{\partial \varphi_1}{\partial x_n}(\alpha) v_n = 0, \\ \vdots \\ \frac{\partial \varphi_m}{\partial x_1}(\alpha)\, v_1 + \cdots + \frac{\partial \varphi_m}{\partial x_n}(\alpha) v_n = 0, \end{cases}$$

under the following conditions that:

$$\begin{cases} \frac{\partial f_1}{\partial x_1}(\alpha)\, v_1 + \cdots + \frac{\partial f_1}{\partial x_n}(\alpha) v_n = 0, \\ \vdots \\ \frac{\partial f_s}{\partial x_1}(\alpha)\, v_1 + \cdots + \frac{\partial f_s}{\partial x_n}(\alpha) v_n = 0. \end{cases}$$

When the algebraic variety V is *smooth* and *equidimensional* of dimension d,[3] the rank of the Jacobian matrix of $f_1, \ldots, f_s$ has dimension $n - d$, and a point $\alpha \in V$ is a critical point of φ if we have

$$\text{Rank}(\text{Jac}(f_1, \ldots, f_s)_\alpha) + \text{Rank}(\text{Jac}(\varphi_1, \ldots, \varphi_m)_\alpha) < n - d + m,$$

that is, the rank of the following matrix

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\alpha) & \cdots\cdots & \frac{\partial f_1}{\partial x_n}(\alpha) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_s}{\partial x_1}(\alpha) & \cdots\cdots & \frac{\partial f_s}{\partial x_n}(\alpha) \\ \frac{\partial \varphi_1}{\partial x_1}(\alpha) & \cdots\cdots & \frac{\partial \varphi_1}{\partial x_n}(\alpha) \\ \vdots & \vdots & \vdots \\ \frac{\partial \varphi_m}{\partial x_1}(\alpha) & \cdots\cdots & \frac{\partial \varphi_m}{\partial x_n}(\alpha) \end{pmatrix},$$

is strictly smaller than $n - d + m$, or equivalently, if all its $(n - d + m, n - d + m)$ minors vanish on p. This yields the following theorem which gives a characterization of the critical points of a polynomial application restricted to a smooth and equidimensional variety.

Theorem 14 *Let $V \subset \mathbb{C}^n$ be a smooth and equidimensional variety of dimension d that is defined as the zero set of the radical ideal $\langle f_1, \ldots, f_s \rangle$ and φ: $\alpha \in \mathbb{C}^n \longrightarrow (\varphi_1(\alpha), \ldots, \varphi_m(\alpha)) \in \mathbb{C}^m$ a polynomial application. The set of critical points of φ restricted to V is the zero-set of the algebraic system that consists of:*

(a) The equations $f_1 = \cdots = f_m = 0$.
(b) The $(n - d - m, n - d - m)$ minors of the matrix $Jac(f_1, \ldots, f_s, \varphi_1, \ldots, \varphi_m)$.

Moreover, the above system is zero-dimensional, i.e., admits a finite number of zeros in $\mathbb{C}^n$.

[3] An algebraic variety is said to be equidimensional if all it irreducible components have the same dimension.

As an example, given an algebraic variety defined by a unique equation

$$V(f) = \{(\alpha_1, \ldots, \alpha_n) \in \mathbb{C}^n \mid f(\alpha_1, \ldots, \alpha_n) = 0\},$$

which we suppose *smooth* and *compact*, and considering the projection function $\Pi_{x_1} : (x_1, \ldots, x_n) \in \mathbb{C}^n \longrightarrow x_1 \in \mathbb{C}$, the set of critical points of Π_{x_1} restricted to $V(f)$ is finite and intersect each connected component of $V(f) \cap \mathbb{R}^n$. According to Theorem 14, this set of critical points can be defined as the zero-set of the system defined by $f = 0$ and the vanishing of $(2, 2)$ minors of the following matrix

$$\begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \\ 1 & 0 & \cdots & 0 \end{pmatrix},$$

that is, $\frac{\partial f}{\partial x_2} = 0, \ldots, \frac{\partial f}{\partial x_n} = 0$.

Example 8 Consider the sphere $\mathcal{S}$ defined by the equation $x^2 + y^2 + (z-1)^2 - 1 = 0$ and the projection $\Pi_x : (x, y, z) \in \mathbb{C}^3 \longrightarrow x$. According to Theorem 14, the critical points of Π_x restricted to $\mathcal{S}$ are the solutions of the following system

$$\begin{cases} x^2 + y^2 + (z-1)^2 - 1 = 0, \\ 2y = 0, \\ 2z - 2 = 0, \end{cases}$$

that is, the two real points $(1, 0, 1)$ and $(-1, 0, 1)$ of $\mathcal{S}$ (see Fig. 8.4).

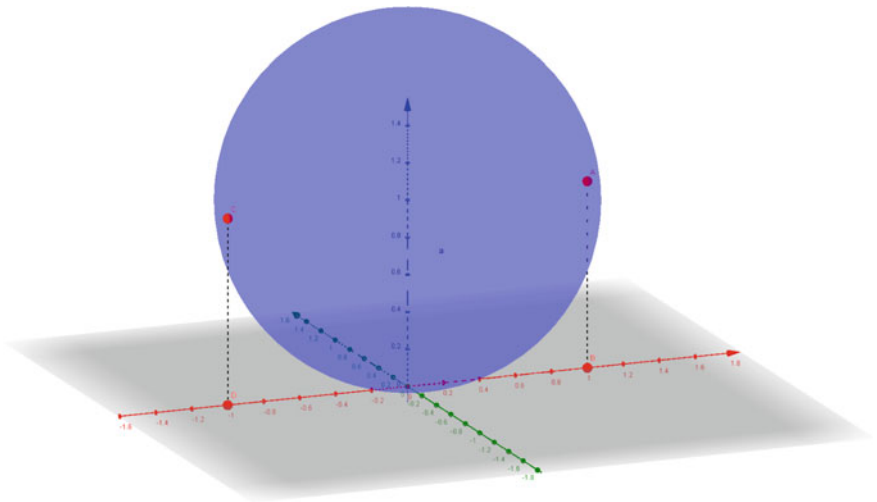


Fig. 8.4 Critical points of Π_x restricted to the variety $x^2 + y^2 + (z-1)^2 - 1 = 0$

Changing the function sometimes allows one to get rid of some assumptions. For example, to avoid the compactness assumption, one can consider the extrema of the distance function to some point A. When the point A is chosen *generic* enough and the set $V(f)$ is smooth, this allows one to reduce the problem to the resolution of a zero-dimensional. More precisely, the set of critical points of the distance function with respect to A is defined by

$$V(\mathcal{C}(A)) = \left\{ p \in \mathbb{C}^n \mid f(p) = 0 \wedge \mathrm{grad}_p(f) // \overrightarrow{\mathrm{Ap}} \right\},$$

where $\mathrm{grad}_p(f)$ is the gradient vector of f at the point $p = (p_1, \ldots, p_n)$. The points of $V(\mathcal{C}(A))$ are the zeros of the ideal generated by $\mathcal{C}(A) = \{f, f_{1,A}, \ldots, f_{r,A}\}$ where the $f_{i,A}$ are the 2×2 minors of the following matrix:

$$\begin{pmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \\ x_1 - a_1 & \cdots & x_n - a_n \end{pmatrix}.$$

The main algorithmic problem when using such a general strategy is that the assumptions made on the system cannot easily be checked (compactness, smoothness and equidimensionality) and/or bypassed, and the situation becomes more involved when dealing with systems of equations of the form $\{f_1 = 0, \ldots, f_s = 0\}$ rather than a unique one.

In [7] for instance, the authors first replace the system $\{f_1 = 0, \ldots., f_s = 0\}$ by the unique equation $f = \sum f_i^2$, then add an infinitesimal Ω and a new variable to switch to a bounded variety $f_\Omega = f^2 + (x_1^2 + \cdots + x_n^2 + x_{n+1}^2 - (\frac{1}{\Omega})^2)$ and then add a second infinitesimal ϵ to get a smooth and bounded variety defined by a unique equation $f_{\Omega,\epsilon} = (1-\epsilon) f_\Omega + \epsilon \left(x_1^{2(d_1+1)} + \cdots + x_n^{2(d_n+1)} + x_{n+1}^6 - 3(\frac{1}{\Omega} - 1)^{2(d_1+1)} \right)$. The algorithm then becomes rather simple since it "suffices" to study the system $\left\{ f_{\Omega,\epsilon} = 0, \frac{\partial f_{\Omega,\epsilon}}{\partial x_2} = 0, \ldots, \frac{\partial f_{\Omega,\epsilon}}{\partial x_{n+1}} = 0 \right\}$ and then take the limits (when $\Omega, \epsilon \to 0$) of the solutions. However, such a strategy turns out to be quite inefficient in practice, mainly because of the costly computations induced by the infinitesimal deformations as well as the degree increase produced by the sum of squares.

In [20], the authors avoid computing sum of squares as well as infinitesimal deformations by considering the distance function and recursively computing the critical points of the singular locus (which is another algebraic variety of smaller dimension).

The current state of the art algorithms and implementations (see [21, 22]) use extended notions of critical points/values (generalized critical values) to avoid the compactness assumption and prevent as much as possible either recursive call and/or costly decompositions.

References

1. Neumaier, A.: Introduction to Numerical Analysis. Cambridge University Press, Cambridge (2001)
2. Collins, G.E., Akritas, A.G.: Polynomial real root isolation using Descarte's rule of signs. In: Proceedings of the Third ACM Symposium on Symbolic and Algebraic Computation. ACM, pp. 272–275 (1976)
3. Rheinboldt, W.C.: Methods for Solving Systems of Nonlinear Equations, vol. 70. SIAM, Philadelphia (1998)
4. Mourrain, B., Pavone, J.: Subdivision methods for solving polynomial equations. J. Symb. Comput. **44**(3), 292–306 (2009)
5. Bouzidi, Y., Quadrat, A., Rouillier, F.: Computer algebra methods for testing the structural stability of multidimensional systems. In: 2015 IEEE 9th International Workshop on Multidimensional (nD) Systems (nDS). IEEE, pp. 1–6 (2015)
6. Cox, D., Little, J., O'Shea, D.: Ideals, Varieties, and Algorithms. Undergraduate Texts in Mathematics, 3rd edn. Springer, New York (2007)
7. Basu, S., Pollack, R., Roy, M.-F.: Algorithms in Real Algebraic Geometry. Algorithms and Computation in Mathematics, 2nd edn. Springer, Berlin (2006)
8. Emiris, I.Z., Mourrain, B., Tsigaridas, E.P.: The DMM bound: multivariate (aggregate) separation bounds. In: Watt, S. (ed.) ISSAC'10, Munich, Germany, pp. 243–250. ACM (2010)
9. Rouillier, F., Zimmermann, P.: Efficient isolation of polynomial real roots. J. Comput. Appl. Math. **162**(1), 33–50 (2003)
10. Buchberger, B.: In: Bose, N.K. (ed.) Gröbner Bases: An Algorithmic Method in Polynomial Ideal Theory. Recent Trends in Multidimensional Systems Theory. Reidel, Dordrecht (1985)
11. Faugère, J.-C.: A new efficient algorithm for computing Gröbner bases (F_4). J. Pure Appl. Algebra **139**(1–3), 61–88 (1999)
12. Brownawell, W.D., et al.: A pure power product version of the Hilbert Nullstellensatz. Mich. Math. J. **45**(3), 581–597 (1998)
13. Rouillier, F.: Solving zero-dimensional systems through the rational univariate representation. J. Appl. Algebra Eng. Commun. Comput. **9**(5), 433–461 (1999)
14. Revol, N., Rouillier, F.: Motivations for an arbitrary precision interval arithmetic and the MPFI library. Reliab. Comput. **11**, 1–16 (2005)
15. Decarlo, R.A., Murray, J., Saeks, R.: Multivariable Nyquist theory. Int. J. Control **25**(5), 657–675 (1977)
16. Bistritz, Y.: Zero location with respect to the unit circle of discrete-time linear system polynomials. Proc. IEEE **72**(9), 1131–1142 (1984)
17. Li, L., Xu, L., Lin, Z.: Stability and stabilisation of linear multidimensional discrete systems in the frequency domain. Int. J. Control **86**(11), 1969–1989 (2013)
18. Bouzidi, Y., Rouillier, F.: Certified algorithms for proving the structural stability of two dimensional systems possibly with parameters. In: MNTS 2016-22nd International Symposium on Mathematical Theory of Networks and Systems (2016)
19. Collins, G.: Quantifier elimination for real closed fields by cylindrical algebraic decomposition. Springer Lecture Notes in Computer Science, vol. 33, pp. 515–532 (1975)
20. Aubry, P., Rouillier, F., Din, M.S.E.: Real solving for positive dimensional systems. J. Symb. Comput. **34**(6), 543–560 (2002). http://www.sciencedirect.com/science/article/pii/S0747717102905638
21. RAGLIB: A library for real solving polynomial systems of equations and inequalities. http://www-salsa.lip6.fr/~safey/RAGLib/
22. Safey El Din, M., Schost, É.: Polar varieties and computation of one point in each connected component of a smooth real algebraic set. In: Proceedings of ISSAC 2003. ACM, pp. 224–231 (2003)

Chapter 9
A Review on Multiple Purely Imaginary Spectral Values of Time-Delay Systems

Islam Boussaada and Silviu-Iulian Niculescu

Abstract A standard framework in analyzing time-delay systems consists first, in identifying the associated crossing roots and secondly, then, in characterizing the local bifurcations of such roots with respect to small variations of the system parameters. Moreover, the dynamics of such spectral values are strongly related to their multiplicities (algebraic/geometric). This chapter review some new results by the authors from Boussaada and Niculescu (IEEE Trans Autom Control 61:1601–1606, [1]), Boussaada and Niculescu (Acta Applicandæ Mathematicæ 145(1):47–88, [2]), Boussaada and Niculescu (Proceeding of the 21st International Symposium on Mathematical Theory of Networks and Systems, pp. 1–8, [3]) allowing one to characterize the algebraic multiplicity of a quasipolynomial's crossing imaginary roots. First, we emphasize the link between the multiplicity characterization and functional Birkhoff matrices. Secondly, we elaborate a constructive bound for the multiplicity of a given crossing imaginary root. It is shown that Pólya-Szegő generic bound is never reached when the crossing frequency is different from zero.

Keywords Time-delay systems · Non-hyperbolic singular points · Local bifurcations · Multiple Hopf points · Bogdanov–Takens singularity

9.1 Introduction

The study of symmetries has became an important topic in the field of nonlinear dynamical systems since a wide range of applications displays equivariance conditions. This is essentially due to domains with geometrical symmetries aris-

I. Boussaada (✉)
IPSA & Laboratoire des Signaux et Systèmes, CNRS-CentraleSupélec-Université Paris Sud, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette cedex, France
e-mail: islam.boussaada@l2s.centralesupelec.fr

S.-I. Niculescu
Laboratoire des Signaux et Systèmes, CNRS-CentraleSupélec-Université Paris Sud, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette cedex, France
e-mail: Silviu.Niculescu@l2s.centralesupelec.fr

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_9

ing in fluid mechanics, structural mechanics, reaction-diffusion problems. Typical example for problems with symmetries are networks consisting in identical n-agents reaction problems with diffusion between neighboring agents. For instance, in [4] a network composed from four identical Brusselator chemical reactors is considered, through which it is observed that the existence of multi-dimensional irreducible representations of the symmetry group may force a spectral value to be multiple. In particular, it is emphasized that such underlying symmetries often induce *multiple Hopf bifurcation points*.[1] As another example where multiple Hopf points occur is reported in [5], where the loss of stability of the down hanging equilibrium position of tubes covering fluid is studied. Due to symmetries, it is shown that *Crossing Imaginary Roots* (CIR) (associated with a Multiple Hopf point) are necessarily double. In time-delay systems context, the paper [6] provides a characterization of the 1:1 resonant Hopf points (double Hopf points with the same frequency ω) in a 6-agents Bidirectional Associative Memory (BAM) neural network. Furthermore, multiple zero spectral value may occur in applications. The simplest case is known as the Bogdanov–Takens singularity which is characterized by an algebraic multiplicity two and a geometric multiplicity one. Cases with higher order multiplicities of the zero spectral value are known to us as generalized Bogdanov–Takens singularities. Those types of configurations are not only theoretical since they arise in concrete applications. Indeed, the Bogdanov–Takens singularity is identified in [7], where the case of two coupled scalar delay equations modeling a physiological control problem is studied. In [8] and [9], this type of singularity is also encountered in the study of coupled axial-torsional vibrations of some oilwell rotary drilling system. Moreover, the paper [10] is devoted to the analysis of such type of singularities where codimension two and three are studied, and the associated center manifolds are explicitly computed. It is commonly accepted that the time-delay induces desynchronizing and/or destabilizing effects on the dynamics. However, new theoretical developments in control of finite-dimensional dynamical systems suggest the use of delays in the control laws for stabilization purposes. For instance, [11] is concerned by the stabilization of the inverted pendulum by delayed control laws and provide concrete situations where the codimension of the zero spectral value exceeds the number of the coupled scalar equations modeling the inverted pendulum on cart.

In this chapter, we review results developed by the authors in [1, 3] about the characterization of multiple imaginary crossing roots for time-delay systems, see also [2] as an expended version of [3]. The remaining chapter is organized as follows: Sect. 9.2 is dedicated to formulate the problem and to recall some useful notions. Section 9.3 provides some results from [2] on LU-factorization of a class of functional Birkhoff matrices. Section 9.4 concerns the zero singularity. It includes some important results from [2, 3] allowing one to recover the generic Pólya-Szegő bound $\sharp_{PS}$. A resulting constructive framework is presented. Section 9.5 extends the last results to multiple

[1] An equilibrium point is called a Hopf point if the Jacobian at that point has a conjugate pair of purely imaginary spectral values $\pm i\omega$, $\omega > 0$. If there are two such pairs $\pm i\omega_1, \pm i\omega_2$ then it is called a double Hopf point. If additionally, $\omega_1 = \omega_2$ then it is called a 1:1 resonant double Hopf point.

CIRs with non zero frequencies. Under some sparsity conditions, a control oriented illustrative example is provided in Sect. 9.6. Some concluding remarks end the paper.

9.2 Problem Statement and Prerequisites

Consider the following infinite-dimensional system with N constant delays:

$$\dot{x} = \sum_{k=0}^{N} A_k x(t - \tau_k) \tag{9.1}$$

where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ denotes the state-vector, under appropriate initial conditions belonging to the Banach space of continous functions $\mathcal{C}([-\tau_N, 0], \mathbb{R}^n)$. Here τ_j, $j = 1, \ldots, N$ are strictly increasing positive constant delays with $\tau_0 = 0$ and $0 < \tau_1 < \tau_2 < \cdots < \tau_N$, the matrices $A_j \in \mathcal{M}_n(\mathbb{R})$ for $j = 0, \ldots, N$. It is well known that the asymptotic behavior of the solutions is determined by the roots of the characteristic equation [12, 13], that is a transcendental equation in the Laplace variable λ in which appear exponential terms induced by delays. More precisely, system (9.1) has a characteristic function $\Delta : \mathbb{C} \times \mathbb{R}_+^N \to \mathbb{C}$ of the form:

$$\Delta(\lambda, \tau) = \det\left(\lambda\, I - A_0 - \sum_{k=1}^{N} A_k\, e^{-\tau_k \lambda}\right) \tag{9.2}$$

or shorter, denoted $\Delta(\lambda)$, which gives

$$\begin{aligned} \Delta(\lambda) &= P_0(\lambda) + \sum_{M^k \in S_{N,n}} P_{M^k}(\lambda)\, e^{\sigma_{M^k}\, \lambda} \\ &= P_0(\lambda) + \sum_{k=1}^{\tilde{N}_{N,n}} P_{M^k}(\lambda)\, e^{\sigma_k\, \lambda} \end{aligned} \tag{9.3}$$

where $\sigma_{M^k} = -M^k\, \tau^T$, $\tau = (\tau_1, \ldots, \tau_N)$ is the delays vector and $S_{N,n}$ is the set of all the possible row vectors $M^k = (M_1^k, \ldots, M_N^k)$ belonging to $\mathbb{N}^N$ such that $1 \leqslant M_1^k + \cdots + M_N^k \leqslant n$ and $\tilde{N}_{N,n} = \#(S_{N,n})$. For instance,

$$\begin{aligned} S_{3,2} = \{&(1,0,0),\ (0,1,0),\ (0,0,1),\ (2,0,0),\ (1,1,0),\\ &(1,0,1),\ (0,2,0),\ (0,1,1),\ (0,0,2)\}, \end{aligned}$$

is first ordered by increasing sums $(\sum_{i=1}^{N} M_i^k)$ and then by lexicographical order, in this case, one has:

$$M^2 = (0,1,0) \quad \text{and} \quad \tilde{N}_{3,2} = 9.$$

Without any lost of generality, assume that P_0 is a monic polynomial of degree n in λ and the polynomials P_{M^k} are such that $\deg(P_{M^k}) = n - \sum_{s=1}^{N} M_s^k \leqslant (n-1)$ $\forall M^k \in S_{N,n}$. Let D_q denote the degree of the quasipolynomial $\sum_{k=1}^{\tilde{N}_{N,n}} P_{M_k}$. One can prove that the quasipolynomial function (9.3) admits an infinite number of zeros, see [13, 14]. The study of zeros of entire function [15] in the form (9.3) plays a crucial role in the analysis of the asymptotic stability of the zero solution of given system (9.1). Indeed, the zero solution is asymptotically stable if all the zeros of (9.3) are in the open left-half complex plane [16]. Accordingly to this observation, the parameter space which is spanned by the coefficients of the polynomials P_i, can be split into stability and instability domains (Nothing else than the so-called *D-decomposition*, see for instance [16] and references therein). These two domains are separated by a boundary, called the *critical boundary*, corresponding to the spectra consisting in roots with zero real parts. When the intersection of the spectrum with such a boundary is nonempty then the equilibrium point is said to be *nonhyperbolic*; which is the context of the present study. Furthermore, the local behavior at a nonhyperbolic singularity is described by the versal deformation of the singularity; that is, replacing the original vector field $f(.)$ by a perturbation-dependent vector field $g(., \epsilon)$ such that when the vector parameter vanish $\epsilon = (\epsilon_1, \ldots, \epsilon_k) = 0$, one has $f(.) = g(., \epsilon)|_{\epsilon=0}$. This deformation g is said to be *versal* if any other deformation occurs as a deformation induced from g and the number of its parameters k is minimal. The *codimension* of such a singularity is nothing else than the integer k. The notion of codimension is the tool allowing one to classify singularities for the Bifurcation theory. Recall that, the algebraic multiplicity of the zero spectral value is nothing else than the corresponding codimension, it represents a bound for the codimension when dealing with a CIR with non zero frequency.

Although the algebraic multiplicity of each spectral value of a time-delay system is finite (a direct consequence of Rouché Theorem, see [17]), to the best of the authors' knowledge, the estimation of *the upper bound of such a multiplicity for a given CIR* did not receive a complete characterization especially when the physical parameters of a given time-delay model are subject to algebraic constraints. It is worth mentioning that the root at the origin is invariant with respect to the delay parameters. However, its multiplicity is strongly dependent on the existing links between the delays and the other parameters of the system.

This chapter is devoted to review some new results of [1, 3] to give an answer to the question above. This work is motivated by the fact that the knowledge of such piece of information is crucial. First, in the linear analysis for time-delay systems, for instance, the analysis of sensitivity as well as the study local bifurcation. Secondly, when dealing with a nonlinear analysis and the computations of the center manifold are involved, see for instance [18–20].

The following result of [17] gives some valuable information allowing one to have a first estimation of such a bound for the multiplicity. proposition

Proposition 1 (Pólya-Szegö, [17], pp. 144) *Let* $\tau_1, \ldots, \tau_N$ *denote real numbers such that*

$$\tau_1 < \tau_2 < \ldots < \tau_N,$$

and $d_1, \ldots, d_N$ *positive integers satisfying*

$$d_1 \geqslant 1,\ d_2 \geqslant 1\ \ldots\ d_N \geqslant 1, \quad d_1 + d_2 + \cdots + d_N = D + N.$$

Let $f_{i,j}(s)$ *stands for the function* $f_{i,j}(s) = s^{j-1}\, e^{\tau_i s}$, *for* $1 \leqslant j \leqslant d_i$ *and* $1 \leqslant i \leqslant N$.
Let $\sharp$ *be the number of zeros of the function*

$$f(s) = \sum_{1 \leqslant i \leqslant N, 1 \leqslant j \leqslant d_i} c_{i,j}\ f_{i,j}(s),$$

that are contained in the horizontal strip $\alpha \leqslant \mathcal{I}(z) \leqslant \beta$.
Assuming that

$$\sum_{1 \leqslant k \leqslant d_1} |c_{1,k}| > 0,\ \ldots,\ \sum_{1 \leqslant k \leqslant d_N} |c_{N,k}| > 0,$$

then

$$\frac{(\tau_N - \tau_1)(\beta - \alpha)}{2\pi} - D + 1 \leqslant \sharp \leqslant \frac{(\tau_N - \tau_1)(\beta - \alpha)}{2\,\pi} + D + N - 1.$$

See also [21] for a modern formulation of the mentioned result. The proof of Pólya-Szegö result is mainly based on Rouché Theorem. It can be generically exploited to establish a bound for the multiplicity of the zero spectral value that we denote by $\sharp_{PS}$. Indeed, setting $\alpha = \beta = 0$ yields $\sharp_{PS} \leqslant D + N - 1$ where D stands for the sum of degrees of the involved polynomials corresponding to the quasipolynomial function f and N designates the associated number of polynomials. This gives a sharp bound in the case of *complete polynomials* i.e. polynomials having all their terms ordered from the greatest degree up to the constant term. Nevertheless, it is obvious that the Pólya-Szegö bound remains unchanged when certain coefficients $c_{i,j}$ vanish without affecting the degree of the quasipolynomial function. Such a remark allows us to claim that Pólya-Szegö bound does not take into account the algebraic constraints on the parameters. However, such constraints are commonly encountered in control problems due to models structures: explicit situations will be given in the next section concerned by motivating examples. Moreover, when one needs the explicit conditions on the system's parameters insuring a given multiplicity (bounded by $\sharp_{PS}$), then computations of the successive differentiations of the quasipolynomial have to be made.

In the present chapter, we emphasize a systematic approach allowing us to a sharper bound for CIR's multiplicity. Indeed, the proposed approach does not only take into account the algebraic constraints on the coefficients $c_{i,j}$ but it also furnishes appropriate conditions guaranteeing such a multiplicity. Furthermore, the symbolic approach we adopt in this study underlines the connexion between the codimension of the zero singularity problem and *incidence matrices* of the so-called *Confluent Vandermonde Matrix* as well as the *Birkhoff Matrix*, see for instance [2, 22–25]. To the best of the authors' knowledge, the first time the Vandermonde matrix appears in a control problem is reported in [26], where the controllability of a finite dimensional dynamical system is guaranteed by the invertibility of such a matrix, see [26, p.

121]. Next, in the context of time-delay systems, the use of Vandermonde matrix properties was proposed by [16, 27] when controlling one chain of integrators by delay blocks. Here we further exploit the algebraic properties of such matrices in a different context.

Initially, Birkhoff and Vandermonde matrices are derived from the problem of polynomial interpolation of an unknown function g, that can be presented in a general way by describing the interpolation conditions in terms of *Birkhoff incidence matrices*, see for instance [28]. For a given integers $n \geqslant 1$ and $r \geqslant 0$, the matrix

$$\mathcal{E} = \begin{pmatrix} e_{1,0} & \dots & e_{1,r} \\ \vdots & & \vdots \\ e_{n,0} & \dots & e_{n,r} \end{pmatrix},$$

is called an incidence matrix if $e_{i,j} \in \{0, 1\}$ for every i and j. Such a matrix contains the data providing the known information about the function g. Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ be such that $x_1 < \cdots < x_n$, the problem of determining a polynomial $\hat{P} \in \mathbb{R}[x]$ with degree less or equal to r that interpolates g at $(x, \mathcal{E})$, i.e. which satisfies the conditions:

$$\hat{P}^{(j)}(x_i) = g^{(j)}(x_i)$$

is known as the *Birkhoff interpolation problem*. An incidence matrix $\mathcal{E}$ is said to be *poised* if such a polynomial $\hat{P}$ is unique. This amounts to saying that the coefficients of the interpolating polynomial $\hat{P}$ are solutions of a linear square system with an associated square matrix $\Upsilon_{\mathcal{E}}$ that we call in the sequel by *Birkhoff matrix*. This matrix is parametrized in x and is shaped by $\mathcal{E}$. It turns out that the incidence matrix $\mathcal{E}$ is poised if and only if the Birkhoff matrix $\Upsilon_{\mathcal{E}}$ is non singular for all x such that $x_1 < \cdots < x_n$. The characterization of poised incidence matrices is solved for the interpolation problem for low degrees. For instance, the problem is still unsolved for any degree $n \geqslant 6$, see for instance [25, 29]. As an illustration of the above notions, let consider the reduced example from [29] with the incidence matrix

$$\mathcal{E} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \tag{9.4}$$

for which we associate the Birkhoff matrix

$$\Upsilon_{\mathcal{E}}^T = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 0 & 1 & 2x_1 & 3x_1^2 \\ 0 & 0 & 2 & 6x_2 \\ 0 & 1 & 2x_3 & 3x_3^2 \end{pmatrix}.$$

The interpolation problem is solvable if and only if

$$12\,x_3\,x_2 + 6\,x_1^2 - 12\,x_2\,x_1 - 6\,x_3^2$$

does not vanish for all values of x such that $x_1 < x_2 < x_3$. For the sake of the space limit, in the sequel, one can afford to replace the incidence matrix $\mathcal{E}$ by an appropriate vector $\mathcal{V}_\mathcal{E}$ reproducing exactly the same information, for instance, in the case of (9.4), one has $\mathcal{V}_\mathcal{E} = (x_1, x_1, \star, \star, x_2, \star, x_3)$. We point out that when no stars appear in $\mathcal{V}_\mathcal{E}$ and no any variable is repeated in the sequence defining $\mathcal{V}_\mathcal{E}$ then we are dealing with the classical Vandermonde matrix, otherwise (there are at least a repeated variable in $\mathcal{V}_\mathcal{E}$) the matrix $\Upsilon_\mathcal{E}$ is the so called Confluent Vandermonde matrix.

In the sequel, we associate to each given positive integer $s \geqslant 0$ and a given incidence matrix $\mathcal{E}$ (or equivalently $\mathcal{V}_\mathcal{E}$) a *functional Birkhoff matrix* which is the square matrix $\Upsilon_\mathcal{E}^s$ defined by:

$$\left\{\begin{array}{rl} \Phi =& [\Phi_1\ \Phi_2\ \ldots\ \Phi_M] \in \mathcal{M}_{l,\delta}(\mathbb{R}), \\ \Phi_i =& [f(\sigma_i)\ f^{(1)}(\sigma_i)\ \ldots\ f^{(d_i-1)}(\sigma_i)], \\ f(\sigma_i) =& g(\sigma_i).[1\ \ldots\ \sigma_i^{l-1}]^T, \quad \text{for} \quad 1 \leqslant i \leqslant M, \end{array}\right. \tag{9.5}$$

for a sufficiently regular function $g \in \mathcal{C}^k(\mathbb{R})$, see [30]. In the sequel, we are particularly interested in square matrices where it is assumed that $\sum_{i=1}^{M} d_i = \delta = l$. If $g(x) = 1$ then we are dealing with the so-called confluent Vandermonde matrix, see [30]. If, additionally, $d_i = 1$ for $i = 1, \ldots, N$, then, we recover the classical Vandermonde matrix and in this case $M = \delta$ since Φ is assumed to be a square matrix, see [31].

The explicit development of numeric/symbolic algorithms for LU-factorization and inversion of the Vandermonde and confluent Vandermonde matrices [32, 33] is still an attracting topic due to their specific structure and their implications in various applications, see for instance [26, 34] and references therein. The authors propose in [2] an explicit recursive formula for the LU-factorization for three configurations of the functional Birkhoff matrix defined by (9.5). To the best of the authors' knowledge, such an explicit formulas seems to be unavailable in the mathematical literature and linear algebra textbooks, see for instance [25]. The Birkhoff matrix configurations we consider are: the first one, the regular case (no stars in $\mathcal{V}_\mathcal{E}$), that is the functional confluent Vandermonde matrix. The second configuration is when the polynomials associated with the delays in the quasipolynomial are sparse, that is, $\mathcal{V}_\mathcal{E}$ contains stars. For instance, it is the case for the variable blocks x_2 and x_3 in the example (9.4).

Furthermore, as a byproduct of the approach, we will first present a different proof for the Pólya-Szegő bound $\sharp_{PS}$ of the origin multiplicity deduced from Proposition 1, then we will establish a sharper bound for such a multiplicity under the hypothesis of the nondegeneracy of an appropriate Birkhoff matrix.

The following notations are adopted. Let ξ stand for the vector composed from x_i counting with their repetition d_i through columns of Υ, that is

$$\xi = (\underbrace{x_1, \ldots, x_1}_{d_1}, \ldots, \underbrace{x_M, \ldots, x_M}_{d_M}).$$

For instance, one has $\xi_1 = x_1$ and $\xi_{d_1+d_2+1} = \xi_{d_1+d_2+d_3} = x_3$. According to the above notations and under the setting $d_0 = 0$, without any loss of generality: $\xi_k = \xi_{d_0+\cdots+d_r+\alpha} = \xi_{\sum_{l=0}^{\varrho(k)-1} d_l + \varkappa(k)}$, for $0 \leqslant r \leqslant M-1$ and $\alpha \leqslant d_{r+1}$, where $\varrho(k)$ denotes the index of component of x associated with ξ_k, that is $x_{\varrho(k)} = \xi_k$, and by $\varkappa(k)$ the order of ξ_k in the sequence of ξ composed only by $x_{\varrho(k)}$. Obviously, we have $\varrho(k) = r+1$ and $\varkappa(k) = \alpha$.

9.3 Functional Birkhoff Matrices

We start this section by defining some results on functional confluent Vandermonde matrices that will be useful for the remaining paper. For the sake of simplicity, since we are concerned by the regular case, $\Upsilon_{\mathcal{E}}$ will be denoted Υ.

It is well known that Vandermonde and confluent Vandermonde matrices V can be factorized into a lower triangular matrix L and an upper triangular matrix U where $V = LU$, see for instance [35, 36]. In what follows, we show that the same applies for the functional confluent Vandermonde matrix (9.5) by establishing explicit formulas for L and U where $\Upsilon = LU$. The factorization is *unique* if no row or column interchanges are made and if it is specified that the diagonal elements of L are unity. The following theorem concerning (9.5) with $s = n+1$ will be used in the sequel but, by the same way, it can be easily adapted for any positive integer s. The following result is proved in [2] using a 2D recurrence.

Theorem 1 ([2]) *Given the functional confluent Vandermonde matrix* (9.5) *with incidence vector* $\mathcal{V}_{\mathcal{E}}$ *wanting stars, the unique LU-factorization with unitary diagonal elements* $L_{i,i} = 1$ *is given by the formulae:*

$$\begin{cases} L_{i,1} = x_1^{i-1} \quad \textit{for} \quad 1 \leqslant i \leqslant \delta, \\ U_{1,j} = \Upsilon_{1,j} \quad \textit{for} \quad 1 \leqslant j \leqslant \delta, \\ L_{i,j} = L_{i-1,j-1} + L_{i-1,j}\,\xi_j \quad \textit{for} \quad 2 \leqslant j \leqslant i, \\ U_{i,j} = (\varkappa(j)-1)\,U_{i-1,j-1} + U_{i-1,j}\left(x_{\varrho(j)} - \xi_{i-1}\right) \\ \qquad \textit{for} \quad 2 \leqslant i \leqslant j. \end{cases} \tag{9.6}$$

The explicit computation determinant of the functional confluent Vandermonde matrix Υ follows directly from (9.6) as explained in the next corollary:

Corollary 1 ([2]) *The determinant of the functional confluent Vandermonde matrix* Υ *is given by:*

$$det(\Upsilon) = \prod_{j=1}^{\delta} \left(U_{j,j}\right),$$

where for $1 \leqslant j \leqslant \delta$, $U_{j,j}$ *are defined by:*

$$\begin{cases} U_{1,1} = x_1^{n+1}, \\ U_{j,j} = U_{j-1,j} \left(x_{\varrho(j)} - \xi_{j-1} \right) \text{ when } j > 1 \text{ and } \varkappa(j) = 1, \\ U_{j,j} = (\varkappa(j) - 1)\, U_{j-1,j-1} \quad \text{otherwise}. \end{cases}$$

Moreover, the diagonal elements of the matrix U *associated with the functional confluent Vandermonde matrix* Υ *are obtained as follows:*

$$\begin{cases} U_{1,1} = x_1^{n+1}, \\ U_{j,j} = x_{k+1}^{n+1} \prod_{l=1}^{k} (x_{k+1} - x_l)^{d_l} \quad \text{when} \quad j = 1 + d_k \quad \text{for} \quad 1 \leqslant k \leqslant M-1, \\ U_{j,j} = (j - 1 - d_k)\, U_{j-1,j-1} \quad \text{when} \quad d_k + 1 < j \leqslant d_{k+1} \quad \text{for} \quad 1 \leqslant k \leqslant M-1. \end{cases}$$

Moreover, the functional confluent Vandermonde matrix Υ *is invertible if and only if we have* $x_i \neq 0$ *and* $x_i \neq x_j \; \forall \; 1 \leqslant i \neq j \leqslant \delta$.

9.4 The Multiple Zero Singularity

9.4.1 Recovering Polya-Segö Generic Bound

In this section we focus on *the regular case*, that is when all the polynomials of the delayed part of the studied quasipolynomial are complete. However, the complementary configuration, when the polynomials of the delayed part are *sparse*, that is, when the incidence vector $\mathcal{V}_{\mathcal{E}}$ contains a star or a sequence of successive stars will be considered in the next section.

In view of the obtained results on functional confluent Vandermonde matrix we are now able to prove the following proposition. Let us define $a_{i,j}$ the coefficient of the monomial λ^j for the polynomial P_{M^i} for $1 \leqslant i \leqslant \tilde{N}_{N,n}$ and note $P_{M^0} = P_0$. Thus, we have $a_{0,n} = 1$ and $a_{i,k} = 0 \, \forall k \geqslant d_i = n - \sum_{s=1}^{N} M_s^i$, where $d_i - 1$ is nothing else than the degree of P_{M^i}.

Proposition 2 ([2]) *The multiplicity of the zero root for the generic quasipolynomial function* (9.3) *cannot be larger than* $\sharp_{PS} = D + \tilde{N}_{N,n}$, *where* D *is the sum of degrees of the involved polynomials and* $\tilde{N}_{N,n} + 1$ *the number of the corresponding polynomials. Moreover, such a bound is reached if and only if the parameters of* (9.3) *satisfy simultaneously for* $0 \leqslant k \leqslant \sharp_{PS} - 1$*:*

$$a_{0,k} = - \sum_{i \in S_{N,n}} \left(a_{i,k} + \sum_{l=0}^{k-1} \frac{a_{i,l} {\sigma_i}^{k-l}}{(k-l)!} \right). \tag{9.7}$$

Remark 1 In the generic case, the Pólya-Szegő bound $\sharp_{PS}$ is completely recovered. The proof of Proposition 2 gives an alternative method for identifying such a bound.

Remark 2 When the coefficients of a given time-delay system (9.1) are fixed, we can similarly consider the generic case accompanied with an appropriate algebraic constraint additionally to an inequality constraint due to dealing with positive delays. When written in terms of the coefficients of the associated quasipolynomial (9.3), the algebraic constraint becomes $\mathfrak{C}(a) = 0$ additionally to the inequality constraint $\tau_k > 0$.

Remark 3 The above claim can be interpreted as follows. Under the hypothesis:

$$\Delta(i\omega) = 0 \Rightarrow \omega = 0 \tag{H}$$

that is all the imaginary roots are located at the origin, then the dimension of the projected state on the center manifold associated with zero singularity for Eq. (9.3) is less or equal to its number of nonzero coefficients minus one. Indeed, under (H), the codimension of the zero spectral value is equal to the dimension of the state on the center manifold since in general the state's dimension on the center manifold is nothing else than the sum of the dimensions of the generalized eigenspaces associated with the spectral values having a zero real part.

We first need to introduce some notations. Let denote by $\Delta^{(k)}(\lambda)$ the k-th derivative of $\Delta(\lambda)$ with respect to the variable λ. We say that zero is an eigenvalue of algebraic multiplicity $m \geqslant 1$ for (9.1) if $\Delta(0) = \Delta^{(k)}(0) = 0$ for all $k = 1, \ldots, m-1$ and $\Delta^{(m)}(0) \neq 0$. In what follows, we assume that $\sigma_k \neq \sigma_{k'}$ for any $k \neq k'$ where $k, k' \in S_{N,n}$. Indeed, when for some value of the delay vector τ, there exists some $k \neq k'$ such that $\sigma_k = \sigma_{k'}$ then, the number of auxiliary delays and the number of polynomials is reduced by considering a new family of polynomials $\tilde{P}$ such that $\tilde{P}_{M^k} = P_{M^k} + P_{M^{k'}}$.

Since we are dealing only with the values of $\Delta_k(0)$, we suggest to translate the problem into the parameter space (the space of the coefficients of the P_i), this will be more appropriate and we will consider parametrization by σ. In the following lemma we introduce an m-set of multivariate algebraic functions (polynomials) vanishing at zero when the multiplicity of the zero root of the transcendental equation $\Delta(\lambda) = 0$ is equal to m. The following lemma allows one to establish an m-set of multivariate algebraic equations (polynomials) vanishing at zero when the multiplicity of the zero root of the transcendental equation $\Delta(\lambda) = 0$ is equal to m.

Lemma 1 ([2]) *Zero is a root of $\Delta^{(k)}(\lambda)$ for $k \geqslant 0$ if and only if the coefficients of P_{M^j} for $0 \leqslant j \leqslant \tilde{N}_{N,n}$ satisfy the following assertion*

$$a_{0,k} = -\sum_{i \in S_{N,n}} \left[a_{i,k} + \sum_{l=0}^{k-1} \frac{a_{i,l} {\sigma_i}^{k-l}}{(k-l)!} \right].$$

Proof (*Proof of Proposition 2*:) The condition (9.7) follows directly from Lemma 1. In what follows, we recover the bound $\sharp_{PS}$ by using explicitly the Vandermonde

matrices. Then, when assuming that some coefficients of the quasipolynomial vanish without affecting its degree, we show that a sharper bound can be related to the number of nonzero parameters rather than the degree.

Let define

$$\nabla_k(\lambda) = \sum_{i=0}^{N} \frac{d^k}{d\lambda^k} P_i(\lambda) + \sum_{l=0}^{k-1} \left((-1)^{l+k} \binom{k}{l} \sum_{i=1}^{N} \tau_i^{k-l} \frac{d^l}{d\lambda^l} P_i(\lambda) \right).$$

Elementary computations show that $\Delta^k(0) = 0 \equiv \nabla_k$ for $k = 0, \ldots, \sharp_{PS}$. We shall consider the variety associated with the vanishing of the polynomials ∇_k, that is $\nabla_0(0) = \cdots = \nabla_{m-1}(0) = 0$ and $\nabla_m(0) \neq 0$ and we aim to find the maximal m (codimension of the zero singularity).

Let us exhibit the first elements of the family ∇_k

$$\begin{cases} \nabla_0(0) = \sum_{s=0}^{\tilde{N}_{N,n}} a_{s,0} = 0, \\ \nabla_1(0) = \sum_{s=0}^{\tilde{N}_{N,n}} a_{s,1} + \sum_{s=1}^{\tilde{N}_{N,n}} a_{s,0}\, \sigma_s = 0, \\ \nabla_2(0) = 2 \sum_{s=0}^{\tilde{N}_{N,n}} a_{s,2} + 2 \sum_{s=1}^{\tilde{N}_{N,n}} a_{s,1} \sigma_s + \sum_{s=1}^{N} a_{s,0} \sigma_s^2 = 0. \end{cases}$$

If we consider $a_{i,j}$ and the σ_k's as variables, the obtained algebraic system is nonlinear and solving it in all generality (without attributing values for n and N) becomes a very difficult task. Indeed, even by using Gröbner basis methods [37], this task is unsolvable since the set of variables depends on N and n. However, considering $a_{i,j}$ as variables and the σ_k's as parameters gives the problem a linear aspect as it can be seen from (9.7). Let adopt the following notation: $a_0 = (a_{0,0},\ a_{0,1},\ \ldots,\ a_{0,n-1})^T$ and $a_i = (a_{i,0},\ a_{i,1},\ \ldots,\ a_{i,d_i-1})^T$ for $1 \leqslant i \leqslant \tilde{N}_{N,n}$. Next, denote by $\sigma = (\sigma_1,\ \sigma_2,\ \ldots,\ \sigma_{\tilde{N}_{N,n}})$ and $a = (a_1,\ a_2,\ \ldots,\ a_{\tilde{N}_{N,n}})^T$. Consider now the ideal I_1 generated by the n polynomials

$$I_1 = < \nabla_0(0),\ \nabla_1(0),\ \ldots,\ \nabla_{n-1}(0) > .$$

As it can be seen from (9.7) and Lemma 1, the variety V_1 associated with the ideal I_1 has the following linear representation $a_0 = \Upsilon_1\, a$ such that $\Upsilon_1 \in \mathcal{M}_{n, D_q + \tilde{N}_{N,n}}(\mathbb{R}[\sigma])$ where D_q is the degree of $\sum_{k=1}^{\tilde{N}_{N,n}} P_{M_k}$ and $D_q = D - n$ (D the degree of the quasipolynomial (9.3)). In some sense, in this variety there are no any restriction

on the components of a if a_0 is left free. Since $a_{0,k} = 0$ for all $k > n$, the remaining equations consist in an algebraic system only in a and parametrized by σ. Now consider the ideal I_2 generated by the $D_q + \tilde{N}_{N,n}$ polynomials defined by

$$I_2 =< \nabla_{n+1}(0),\ \nabla_{n+2}(0),\ \ldots,\ \nabla_{D+\tilde{N}_{N,n}}(0) > .$$

It can be observed that the variety V_2 associated with I_2 can be written as $\Upsilon_2\, a = 0$, which is nothing else that an homogeneous linear system with $\Upsilon_2 \in \mathcal{M}_{D_q+\tilde{N}_{N,n}}(\mathbb{R}[\sigma])$. More precisely, Υ_2 is nothing else than the functional confluent Vandermonde matrix (9.5), with $x = \sigma$, $s = n$, $M = \tilde{N}_{N,n}$ and $\delta = D_q + \tilde{N}_{N,n}$, which is associated with some incidence vector:

$$\mathcal{V}_{\mathcal{E}} = (\underbrace{\sigma_{M^1}, \ldots, \sigma_{M^1}}_{n-\sum_{s=1}^{N} M_s^1}, \ldots, \ldots, \sigma_{M^{\tilde{N}_{N,n}}}, \ldots, \sigma_{M^{\tilde{N}_{N,n}}}). \tag{9.8}$$

Now, using Corollary 1 and the assumption that σ_i's are distinct non zero auxiliary delays we can conclude that the determinant of Υ_2 cannot vanish. Thus the only solution for this subsystem is the zero solution, that is $a = 0$.

Finally, consider the polynomial defined by $\nabla_n(0)$, by Lemma 1

$$\nabla_n(0) = 0 \Leftrightarrow 1 = -\sum_{i=1}^{\tilde{N}_{N,n}} \sum_{s=0}^{n-1} \frac{a_{i,s}\sigma_i^{n-s}}{(n-s)!}.$$

Substituting the unique solution of V_2 into the last equality leads to an incompatibility result. In conclusion, the maximal codimension of the zero singularity is less or equal to $D_q + \tilde{N}_{N,n} + n$, which is exactly Pólya-Szegő bound $\sharp_{PS} = \underbrace{D_q + (n+1)}_{D+\tilde{N}_{N,n}}$.

Remark 4 It is noteworthy that the codimension of the zero singularity may decrease if the vector parameter a_0 is not left free. Indeed, if some parameter component $a_{0,k}$ is fixed for $0 \leqslant k \leqslant n-1$, then the variety associated with the first ideal I_1 may impose additional restrictions on the vector parameter a.

9.4.2 On Beyond of Pólya-Szegő Bound

Polynomials in nature (e.g. from applications) are not necessarily generic. They often have some additional structures which we would like to take into account in the multiplicity bound.

Proposition 3 ([2]) *Consider a quasipolynomial function* (9.3) *containing one or several incomplete polynomials, for which an incidence vector* $\mathcal{V}_{\tilde{\mathcal{E}}}$ *is associated—which is nothing than* (9.8)*—such that the vanishing coefficients are replaced by stars.*

When the associated functional Birkhoff matrix $\Upsilon_{\tilde{\mathcal{E}}}$ is nonsingular then the multiplicity of the zero root for the quasipolynomial function (9.3) cannot be larger than n plus the number of nonzero coefficients of the polynomial family $(P_{M^k})_{M^k \in S_{N,n}}$.

Proof (*Proof of Proposition 3*) Similarly as in the proof of Proposition 2: when z coefficients from the polynomial family $(P_{M^k})_{M^k \in S_{N,n}}$ vanish without affecting the degree of the quasipolynomial, then $a^T \in \mathbb{R}^{D_q + \tilde{N}_{N,n} - z}$ and thus the matrix Υ_2 of the proof of Proposition 2 becomes $\Upsilon_{\tilde{\mathcal{E}}} \in \mathcal{M}_{D_q + \tilde{N}_{N,n} - z}(\mathbb{R}[\sigma])$. This proves that the maximal codimension of the zero singularity is less or equal to $D_q + \tilde{N}_{N,n} - z + n < \sharp_{PS}$.

Remark 5 Obviously, the number of non-zero coefficients of a given quasipolynomial function is bounded by the corresponding degree. Thus, the bound elaborated in Proposition 3 is sharper than $\sharp_{PS}$, even in the generic case, that is when all the parameters of the quasipolynomial are left free, these two bounds are equal. Indeed, in the generic case, that is when the number of the left free parameters is optimal, the Pólya-Szegő bound which is equal to $\sharp_{PS} = D + \tilde{N}_{N,n} = n + D_q + \tilde{N}_{N,n}$, which is nothing else than n plus the number of parameters of the polynomial family $(P_{M^k})_{M^k \in S_{N,n}}$.

9.5 Multiple Crossing Imaginary Roots with Non Zero Frequency

In this section we consider generic quasipolynomials and we aim at characterizing their crossing imaginary roots $\lambda_c = j\omega$, $\omega \neq 0$. Let us set $c_k = \cos(\sigma_k \omega)$ and $s_k = \sin(\sigma_k \omega)$ and, for a given real positive number x, we denote by $\lfloor x \rfloor$ the integer part of x or equivalently the floor function at x.

Lemma 2 ([1]) *An imaginary complex number $z = j\,\omega$ is a root of $\partial_z^k \Delta(z, \tau) = 0$ for $k \geqslant 0$ if and only if the coefficients of P_i for $0 \leqslant i \leqslant \tilde{N}$ satisfy*

$$\alpha_k + \tilde{\alpha}_k = 0 \text{ and } \beta_k + \tilde{\beta}_k = 0, \tag{9.9}$$

where

$$\begin{cases} \alpha_0 = \displaystyle\sum_{\substack{i=0,\\ i \text{ even}}}^{n} (-1)^{\lfloor \frac{i}{2} \rfloor} \omega^i a_{0,i} \text{ and } \beta_0 = \sum_{\substack{i=1,\\ i \text{ odd}}}^{n} (-1)^{\lfloor \frac{i}{2} \rfloor} \omega^i a_{0,i}, \\ \alpha_k = -\partial_\omega \beta_{k-1} \text{ and } \beta_k = \partial_\omega \alpha_{k-1} \ \forall k \geqslant 1, \\ \tilde{\alpha}_k = \displaystyle\sum_{i=1}^{\tilde{N}} \sum_{l=0}^{d_i} a_{i,l} \frac{\partial^l \left(c_i {\sigma_i}^k \right)}{\partial \sigma_i^l} \text{ and } \tilde{\beta}_k = \sum_{i=1}^{\tilde{N}} \sum_{l=0}^{d_i} a_{i,l} \frac{\partial^l \left(s_i {\sigma_i}^k \right)}{\partial \sigma_i^l}. \end{cases}$$

The proof of Lemma 2 can be found in [1].

Let us set $\gamma_k = (\alpha_k + \tilde{\alpha}_k)^2 + (\beta_k + \tilde{\beta}_k)^2$ and, for a given positive integer m, let α^m, β^m stand for the vectors:

$$\begin{cases} V_m(a_0, \omega) = -(\alpha_0, \ldots, \alpha_{m-1})^\top, \\ W_m(a_0, \omega) = -(\beta_0, \ldots, \beta_{m-1})^\top. \end{cases} \tag{9.10}$$

Let $\varrho = \sum_{i=1}^{\tilde{N}} d_i$ and define the functional confluent matrices A_m and B_m belonging to $\mathcal{M}_{m,\varrho}(\mathbb{R})$ and characterized by the incidence vector

$$\xi = (\underbrace{\sigma_1, \ldots, \sigma_1}_{d_1}, \ldots, \underbrace{\sigma_{\tilde{N}}, \ldots, \sigma_{\tilde{N}}}_{d_{\tilde{N}}})$$

as well as the respective functions $g_A(x) = \cos(\omega x)$ and $g_B(x) = \sin(\omega x)$. As a direct consequence of the above Lemma 2, one has:

Proposition 4 ([1]) *An imaginary crossing root $z = j\,\omega$ for (9.3) is of multiplicity m if and only if one of the following equivalent assertions holds:*

(a) The variety $\mathcal{V}_m$ is non empty and $\gamma_{m+1} \neq 0$, where

$$\mathcal{V}_m = \left\{ \begin{array}{l} (\omega, a_{0,0}, \ldots, a_{\tilde{N},d_{\tilde{N}}}, \sigma_1, \ldots, \sigma_{\tilde{N}}) \in \mathbb{R}^{n+\delta+1} \times \mathbb{R}_+^{\tilde{N}}, \\ \gamma_k = 0 \text{ for } k = 0, \ldots, m-1 \end{array} \right\}.$$

(b) The frequency ω satisfies the following linear system:

$$\begin{cases} A_m(\omega, \sigma).p = V_m(a_0, \omega), \\ B_m(\omega, \sigma).p = W_m(a_0, \omega). \end{cases} \tag{9.11}$$

Proof Both assertions of the above proposition are direct consequences of Lemma 2. They follow by considering two ideals consisting in the real part $\alpha_k + \tilde{\alpha}_k$, (respectively the imaginary part $\beta_k + \tilde{\beta}_k$) of the successive derivatives of the quasipolynomial function. A careful inspection of the coefficients $\tilde{\alpha}_k$ and $\tilde{\beta}_k$ from Lemma 2 allows to construct the linear system (9.11), where A_m and B_m are functional confluent Vandermonde matrices.

Corollary 2 ([1]) *If the square matrices A_ϱ and B_ϱ are non degenerate then the multiplicity of any imaginary crossing root $z = j\,\omega_0$ for (9.3) is bounded by ϱ.*

Proof For $m = \varrho + 1$, the inconsistency of (9.11) follows from the Kronecker-Rouché-Capelli Theorem, see [38]. Indeed, the rows of the functional confluent Vandermonde matrix $A_{\varrho+1}$ (respectively $B_{\varrho+1}$) are functionally dependent but lin-

early independent provided that A_ϱ, (respectively B_ϱ) is non degenerate. Thus, the rank of the augmented matrix $A_{\varrho+1} : V_{\varrho+1}$, (resp. $B_{\varrho+1} : W_{\varrho+1}$), which is equal to $\varrho + 1$, is greater than the rank of the matrix $A_{\varrho+1}$, (resp. $B_{\varrho+1}$).

Corollary 3 ([1]) *The functional confluent Vandermonde matrix A_ϱ, respectively B_ϱ, is non degenerate if and only if $\forall\ 1 \leqslant i \neq j \leqslant \tilde{N}$, $\omega\sigma_i \neq \pi/2 + k\pi$ and $\sigma_i \neq \sigma_j$, respectively $\omega\sigma_i \neq k\pi$ and $\sigma_i \neq \sigma_j$.*

9.6 Illustration on Inverted Pendulum: An Effective Approach versus Pólya-Szegő Bound

A natural consequence of Propositions 2–3 is to explore the situation when the codimension of zero singularity reaches its upper bound. Starting the section by a generic example, we show the convenience of the proposed approach even in the case of coupling delays. Then the obtained symbolic results are applied to identify an effective sharp bound in the case of concrete physical system (with constraints on the coefficients), namely, the stabilization of an inverted pendulum on cart via a multi-delayed feedback.

We associate to the general planar time-delay system with two positive delays $\tau_1 \neq \tau_2$ the quasipolynomial function:

$$\begin{aligned}\Delta(\lambda) = \lambda^2 + a_{0,0,1}\lambda + a_{0,0,0} + \left(a_{1,0,0} + a_{1,0,1}\lambda\right)\mathrm{e}^{\lambda\sigma_{1,0}} \\ + \left(a_{0,1,0} + a_{0,1,1}\lambda\right)\mathrm{e}^{\lambda\sigma_{0,1}} \\ + a_{2,0,0}\mathrm{e}^{\lambda\sigma_{2,0}} + a_{1,1,0}\mathrm{e}^{\lambda\sigma_{1,1}} + a_{0,2,0}\mathrm{e}^{\lambda\sigma_{0,2}}.\end{aligned} \tag{9.12}$$

Generically, the multiplicity of the zero singularity is bounded by $\sharp_{PS} = 9$. However, in what follows, we present two configurations where such a bound cannot be reached. The first corresponds to the case when $\sigma_i = \sigma_j$ for $i \neq j$ and the second, when some components vector vanish:

$$a = (a_{1,0,0}, a_{1,0,1}, a_{0,1,0}, a_{0,1,1}, a_{2,0,0}, a_{1,1,0}, a_{0,2,0})^T.$$

Formula (9.7) allows us to explicitly compute the confluent Vandermonde matrices Υ_1 and Υ_2 and the expression of $\nabla_2(0)$ from the proof of Proposition 2 such that $\Upsilon_1 a = a_0$, $\nabla_2(0) = 0$ and $\Upsilon_2 a = 0$ where $a_0 = (a_{0,0,0}, a_{0,0,1})^T$:

$$\Upsilon_1 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ \sigma_{1,0} & 1 & \sigma_{0,1} & 1 & \sigma_{2,0} & \sigma_{1,1} & \sigma_{0,2} \end{bmatrix},$$

$$\nabla_2(0) - 2 =$$

$$\begin{bmatrix} {\sigma_{1,0}}^2 & 2\sigma_{1,0} & {\sigma_{0,1}}^2 & 2\sigma_{0,1} & {\sigma_{2,0}}^2 & {\sigma_{1,1}}^2 & {\sigma_{0,2}}^2 \end{bmatrix} a$$

$$\Upsilon_2 =$$

$$\begin{bmatrix} {\sigma_{1,0}}^3 & 3\sigma_{1,0}{}^2 & {\sigma_{0,1}}^3 & 3\sigma_{0,1}{}^2 & {\sigma_{2,0}}^3 & {\sigma_{1,1}}^3 & {\sigma_{0,2}}^3 \\ {\sigma_{1,0}}^4 & 4\sigma_{1,0}{}^3 & {\sigma_{0,1}}^4 & 4\sigma_{0,1}{}^3 & {\sigma_{2,0}}^4 & {\sigma_{1,1}}^4 & {\sigma_{0,2}}^4 \\ {\sigma_{1,0}}^5 & 5\sigma_{1,0}{}^4 & {\sigma_{0,1}}^5 & 5\sigma_{0,1}{}^4 & {\sigma_{2,0}}^5 & {\sigma_{1,1}}^5 & {\sigma_{0,2}}^5 \\ {\sigma_{1,0}}^6 & 6\sigma_{1,0}{}^5 & {\sigma_{0,1}}^6 & 6\sigma_{0,1}{}^5 & {\sigma_{2,0}}^6 & {\sigma_{1,1}}^6 & {\sigma_{0,2}}^6 \\ {\sigma_{1,0}}^7 & 7\sigma_{1,0}{}^6 & {\sigma_{0,1}}^7 & 7\sigma_{0,1}{}^6 & {\sigma_{2,0}}^7 & {\sigma_{1,1}}^7 & {\sigma_{0,2}}^7 \\ {\sigma_{1,0}}^8 & 8\sigma_{1,0}{}^7 & {\sigma_{0,1}}^8 & 8\sigma_{0,1}{}^7 & {\sigma_{2,0}}^8 & {\sigma_{1,1}}^8 & {\sigma_{0,2}}^8 \\ {\sigma_{1,0}}^9 & 9\sigma_{1,0}{}^8 & {\sigma_{0,1}}^9 & 9\sigma_{0,1}{}^8 & {\sigma_{2,0}}^9 & {\sigma_{1,1}}^9 & {\sigma_{0,2}}^9 \end{bmatrix}.$$

As shown in the proof of Proposition 2, Υ_2 is a singular matrix when $\sigma_i = \sigma_j$ for $i \neq j$. For instance, when $\sigma_{2,0} = \sigma_{0,1}$ that is $2\tau_1 = \tau_2$, then the bound of multiplicity of the zero singularity decreases since the polynomials $P_{2,0}$ and $P_{0,1}$ will be collected $\tilde{P}_{0,1} = P_{0,1} + P_{2,0}$.

Consider now a system of two coupled equations with two delays modeling a friction free inverted pendulum on cart. The adopted model is studied in [11, 39–41] and in the sequel we keep the same notations. In the dimensionless form, the dynamics of the inverted pendulum on a cart in Fig. 9.1 is governed by the following second-order differential equation:

$$\left(1 - \frac{3\epsilon}{4}\cos^2(\theta)\right)\ddot{\theta} + \frac{3\epsilon}{8}\dot{\theta}^2\sin(2\theta) - \sin(\theta) + U\cos(\theta) = 0, \tag{9.13}$$

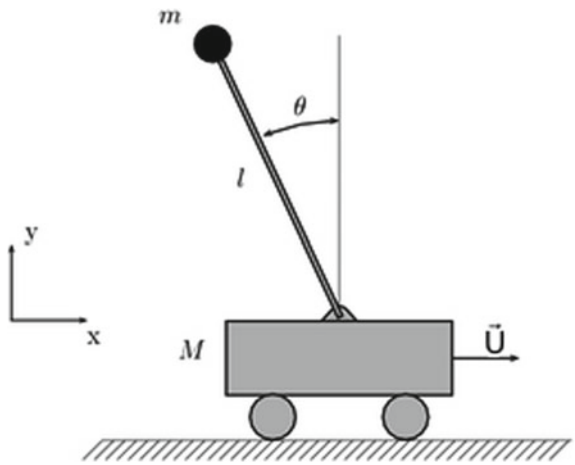


Fig. 9.1 Inverted pendulum on a cart

where $\epsilon = m/(m+M)$, M is the mass of the cart and m is the mass of the pendulum and D represents the control law that is the horizontal driving force. A generalized Bogdanov–Takens singularity with codimension three is identified in [40] by using $U = a\,\theta(t-\tau) + b\,\dot{\theta}(t-\tau)$. Motivated by the technological constraints, it is suggested in [11] to avoid the use of the derivative gain that requires the estimation of the angular velocity that can induce harmful errors for real-time simulations and a multi-delayed-proportional controller $U = a_{1,0}\,\theta(t-\tau_1) + a_{2,0}\,\theta(t-\tau_2)$ is proposed. This choice is argued by the accessibility of the delayed state by some simpler sensor. By this last controller choice and by setting $\epsilon = \frac{3}{4}$, the associated quasipolynomial function Δ becomes:

$$\Delta(\lambda) = \lambda^2 - \frac{16}{7} + \frac{16\,a_1}{7}\,\mathrm{e}^{-\lambda\,\tau_1} + \frac{16\,a_2}{7}\,\mathrm{e}^{-\lambda\,\tau_2}.$$

A zero singularity with codimension three is identified in [11]. Moreover, it is shown that the upper bound of the codimension for the zero singularity for (9.13) is three (can be easily checked by (9.7)) and this configuration is obtained when the gains and delays satisfy simultaneously the following conditions:

$$a_{1,0} = -\frac{7}{-7+8\,\tau_1},\ a_{2,0} = \frac{8{\tau_1}^2}{-7+8\,{\tau_1}^2},\ \tau_2 = \frac{7}{8\,\tau_1}.$$

However, using Pólya-Szegő result, one has $\sharp_{PS} = D - 1 = (3+2+2) - 1 = 6$ which exceeds the effective bound which is three. This is a further justification for the algebraic constraints on the parameters imposed by the physical model, for instance the vanishing of $a_{0,1}$.

Let now consider the sparse case associated with the control law

$$U = a_{1,0}\,\theta(t-\tau_1) + a_{2,1}\,\dot{\theta}(t-\tau_2).$$

The quasipolynomial function Δ becomes:

$$\Delta(\lambda) = \lambda^2 - \frac{16}{7} + a_{1,0}\mathrm{e}^{-\lambda\,\tau_1} + \lambda\,a_{2,1}\,\mathrm{e}^{-\lambda\,\tau_2}.$$

Using Pólya-Szegő result, one has $\sharp_{PS} = D - 1 = (3+2+3) - 1 = 7$. However, using the Proposition 3, one knows that the zero multiplicity cannot be larger than 4. Indeed, the multiplicity 4 is reached only when $a = \frac{16}{7}$, $b = \frac{4\sqrt{42+28\,\sqrt{3}}}{7}$, $\tau_1 = \frac{\sqrt{42+28\,\sqrt{3}}}{4}$, $\tau_2 = \frac{1}{336}\left(42+28\,\sqrt{3}\right)^{3/2} - \frac{\sqrt{42+28\,\sqrt{3}}}{8}$.

Remark 6 The developed framework can be useful in the analysis of a wide range of applications modeled by time-delay systems. For instance, the analysis of a double-inverted pendulum is given in [42] and a biological model describing a vector disease is given in [2].

9.7 Concluding Remarks

This chapter addressed the problem of identifying the maximal dimension of the eigenspace associated with a zero/multiple CIR singularity for time-delay systems as well as the explicit conditions guaranteeing such a configuration. Under the assumption that all the imaginary roots are located at the origin or respectively (at any other crossing frequency) our result gives the relation between d the maximal dimension of the projected state on the center manifold associated with the generalized Bogdanov–Takens singularities (multiple Hopf singularity) from one side and N the number of the delays and n the degree of the polynomial P_0 from the other side. It is shown that the bound deduced from Polya-Segö result [17] is never reached when the crossing frequency is different from zero. Finally, the effective method elaborated in this paper emphasizes the connexions between the multiple roots of quasipolynomials and incidence matrices of a some functional Birkhoff matrices.

Acknowledgements The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

1. Boussaada, I., Niculescu, S.I.: Tracking the algebraic multiplicity of crossing imaginary roots for generic quasipolynomials: a Vandermonde-based approach. IEEE Trans. Autom. Control. **61**, 1601–1606 (2016)
2. Boussaada, I., Niculescu, S.-I.: Characterizing the codimension of zero singularities for time-delay systems. Acta Applicandæ Mathematicæ **145**(1), 47–88 (2016)
3. Boussaada, I., Niculescu, S.-I.: Computing the codimension of the singularity at the origin for delay systems: the missing link with Birkhoff incidence matrices. In: Proceeding of the 21st International Symposium on Mathematical Theory of Networks and Systems, pp. 1–8 (2014)
4. Dellnitz, M., Werner, B.: Computational methods for bifurcation problems with symmetries-with special attention to steady state and Hopf bifurcation points. J. Comput. Appl. Math. **26**(1–2), 97–123 (1989)
5. Steindl, A., Troger, H.: Bifurcations of the equilibrium of a spherical double pendulum at a multiple eigenvalue. In: Küpper, T., Seydel, R., Troger, H. (eds.) Bifurcation: Analysis, Algorithms, Applications, vol. 79 of ISNM 79, pp. 277–287. Birkhäuser, Basel (1987)
6. Zhang, C., Zheng, B., Wang, L.: Multiple hopf bifurcations of symmetric BAM neural network model with delay. Appl. Math. Lett. **22**(4), 616–622 (2009)
7. Hale, J.K., Huang, W.: Period doubling in singularly perturbed delay equations. J. Diff. Eq. **114**, 1–23 (1994)
8. Boussaada, I., Mounier, H., Niculescu, S.-I., Cela, A.: Analysis of drilling vibrations: a time-delay system approach. In: Proceedings of the 20th Mediterranean Conference on Control and Automation, pp. 1–5 (2012)
9. Marquez, M.S., Boussaada, I., Mounier, H., Niculescu, S.-I.: Analysis and Control of Oilwell Drilling Vibrations. Advances in Industrial Control. Springer, Berlin (2015)
10. Campbell, S., Yuan, Y.: Zero singularities of codimension two and three in delay differential equations. Nonlinearity **22**(11), 2671 (2008)
11. Boussaada, I., Morarescu, I.-C., Niculescu, S.-I.: Inverted pendulum stabilization: characterization of codimension-three triple zero bifurcation via multiple delayed proportional gains. Syst. Control Lett. **82**, 1–9 (2015)

12. Diekmann, O., Gils, S.V., Lunel, S.V., Walther, H.: Delay Equations. Applied Mathematical Sciences, Functional, Complex, and Nonlinear Analysis, vol. 110. Springer, New York (1995)
13. Bellman, R., Cooke, K.: Differential-difference Equations. Academic Press, New York (1963)
14. Ahlfors, L.V.: Complex Analysis. McGraw-Hill, Inc., New York (1979)
15. Levin, B.: Distribution of Zeros of Entire Functions. Translations of Mathematical Monographs. AMS, Providence, Rhode Island (1964)
16. Michiels, W., Niculescu, S.-I.: Stability and Stabilization of Time-Delay Systems, Advances in Design and Control, Society for Industrial and Applied Mathematics (SIAM), vol. 12 (2007)
17. Pólya, G., Szegő, G.: Problems and Theorems in Analysis, vol. I: Series, Integral Calculus, Theory of Functions. Springer, New York (1972)
18. Carr, J.: Application of Center Manifold Theory. Springer, Berlin (1981)
19. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields. Springer, Berlin (2002)
20. Kuznetsov, Y.: Elements of Applied Bifurcation Theory. Applied Mathematics Sciences, Vol. 112, 2nd edn. Springer, New York (1998)
21. Wielonsky, F.: A Rolle's theorem for real exponential polynomials in the complex domain. J. Math. Pures Appl. **4**, 389–408 (2001)
22. Björck, A., Elfving, T.: Algorithms for confluent Vandermonde systems. Numer. Math. **21**, 130–137 (1973)
23. Gautshi, W.: On inverses of Vandermonde and confluent Vandermonde matrices. Numer. Math. **4**, 117–123 (1963)
24. Gautshi, W.: On inverses of Vandermonde and confluent Vandermonde matrices II. Numer. Math. **5**, 425–430 (1963)
25. Gonzalez-Vega, L.: Applying quantifier elimination to the Birkhoff interpolation problem. J. Symb. Comp. **22**(1), 83–104 (1996)
26. Kailath, T.: Linear Systems. Prentice-Hall information and system sciences series. Prentice Hall International, Upper Saddle River (1998)
27. Niculescu, S.-I., Michiels, W.: Stabilizing a chain of integrators using multiple delays. IEEE Trans. Aut. Cont. **49**(5), 802–807 (2004)
28. Lorentz, G., Zeller, K.: Birkhoff interpolation. SIAM J. Numer. Anal. **8**(1), 43–48 (1971)
29. Rouillier, F., Din, M., Schost, E.: Solving the Birkhoff interpolation problem via the critical point method: an experimental study. In: Richter-Gebert, J., Wang, D. (eds.) Automated Deduction in Geometry, vol. 2061 of LNCS, pp. 26–40. Springer, Berlin (2001)
30. Ha, T., Gibson, J.: A note on the determinant of a functional confluent Vandermonde matrix and controllability. Linear Algebr. Appl. **30**, 69–75 (1980)
31. Olver, P.: On multivariate interpolation. Stud. Appl. Math. **116**, 201–240 (2006)
32. Melkemi, L., Rajeh, F.: Block lu-factorization of confluent Vandermonde matrices. Appl. Math. Lett. **23**(7), 747–750 (2010)
33. Respondek, J.S.: On the confluent Vandermonde matrix calculation algorithm. App. Math. Lett. **24**(2), 103–106 (2011)
34. Hou, S.-H., Pang, W.-K.: Inversion of confluent Vandermonde matrices. Comput. Math. Appl. **43**(12), 1539–1547 (2002)
35. Oruc, H.: Factorization of the Vandermonde matrix and its applications. Appl. Math. Lett. **20**(9), 982–987 (2007)
36. Melkemi, L.: Confluent Vandermonde matrices using Sylvester's structures. Research Report of the Ecole Normale Supérieure de Lyon **98–16**, 1–14 (1998)
37. Cox, D., Little, J., O'Shea, D.: Ideals, Varieties, and Algorithms. An Introduction to Computational Algebraic Geometry and Commutative Algebra, Springer, New York (2007)
38. Shafarevich, I., Remizov, A.: Matrices and determinants. In Linear Algebra and Geometry, pp. 25–77. Springer, Berlin (2013)
39. Atay, F.M.: Balancing the inverted pendulum using position feedback. Appl. Math. Lett. **12**(5), 51–56 (1999)
40. Sieber, J., Krauskopf, B.: Bifurcation analysis of an inverted pendulum with delayed feedback control near a triple-zero eigenvalue singularity. Nonlinearity **17**, 85–103 (2004)

41. Sieber, J., Krauskopf, B.: Extending the permissible control loop latency for the controlled inverted pendulum. Dyn. Syst. **20**(2), 189–199 (2005)
42. Boussaada, I., Irofti, D., Niculescu, S.-I.: Computing the codimension of the singularity at the origin for time-delay systems in the regular case: A Vandermonde-based approach. In: Proceedings of the 13th European Control Conference, pp. 1–6 (2014)

Chapter 10
Controlled and Conditioned Invariance for Polynomial and Rational Feedback Systems

Christian Schilli, Eva Zerz and Viktor Levandovskyy

Abstract We consider polynomially nonlinear state-space systems and given algebraic varieties. A variety V is said to be controlled invariant w.r.t. a given system if we can find a polynomial state feedback law that causes the closed loop system to have V as an invariant set. If this task can be achieved by a polynomial output feedback law, V is called controlled and conditioned invariant. This concept leads to the problem of determining the intersection of a certain (affine) submodule of a free module over a polynomial ring with a free module over a subalgebra of this ring. Moreover, we expand the set of feedbacks, which can be chosen to make V invariant, to rational ones. We show how to decide whether a variety is controlled invariant for a given system with rational feedback and, in the single output case, if a feedback law can be chosen which just depends on the output. The key point of this consideration will be the intersection of a "fractional" (affine) module with a vector space over the subfield generated by the output and we give a method to do so.

Keywords Polynomial control systems · State-space systems · Invariant sets · Controlled invariant varieties · State feedback · Polynomial feedback · Controlled and conditioned invariance · Output feedback · Symbolic computation · Gröbner bases · Affine intersection · Rational feedback · Fractional modules

This work was supported by DFG Graduiertenkolleg 1632 "Experimental and constructive algebra".

C. Schilli (✉) · E. Zerz · V. Levandovskyy
Lehrstuhl D für Mathematik, RWTH Aachen University,
Pontdriesch 14-16, 52062 Aachen, Germany
e-mail: christian.schilli@rwth-aachen.de
URL: http://www.math.rwth-aachen.de/

E. Zerz
e-mail: eva.zerz@math.rwth-aachen.de

V. Levandovskyy
e-mail: Viktor.Levandovskyy@math.rwth-aachen.de

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_10

10.1 Introduction

Let us first define some objects which are frequently used throughout this paper. Let $K \in \{\mathbb{R}, \mathbb{C}\}$ be a field and $n \in \mathbb{N}$. Then we call $R = K[x_1, \ldots, x_n] = K[\underline{x}]$ the polynomial ring in n variables. Further let $I \subseteq R$ be an ideal of R. We intend to study the variety $V = \mathcal{V}(I) \subseteq K^n$, the common zero set of all polynomials in I, and control systems of the form

$$\begin{aligned} \dot{x}(t) &= f(x(t)) + g(x(t))u(t), \quad f \in R^n, \ g \in R^{n\times m} \\ y(t) &= h(x(t)), \qquad\qquad\quad h \in R^p \end{aligned} \tag{10.1}$$

with $m, p \in \mathbb{N}$. Here $x(t)$ is called the state of the control system at time t, whereas $y(t)$ is the output and $u(t)$ the input at time t. We wish to determine an input function $u(\cdot)$ such that V becomes invariant for (10.1). Let $S = K[h_1, \ldots, h_p] \subseteq R$ be the subalgebra of R generated by the components h_i of h, $T = K[y_1, \ldots, y_p] = K[\underline{y}]$ be another polynomial ring in p variables and $R_1 = K[x_1, \ldots, x_n, y_1, \ldots, y_p] = K[\underline{x}, \underline{y}]$. If k, l are natural numbers, $m_1, \ldots, m_k \in R_1^l$ and $R_2 \subseteq R_1$ is a subring, we write

$$\langle m_1, \ldots, m_k \rangle_{R_2} := \left\{ \sum_{i=1}^{k} a_i m_i \mid a_i \in R_2 \right\}$$

for the R_2-module generated by $m_1, \ldots, m_k$. Further, for any ring D appearing in this chapter, we denote the ith standard basis vector of D^k by e_i and we write $(v)_i$ for the ith component of a vector $v \in D^k$, where $i \in \{1, \ldots, k\}$.

This chapter is organized as follows: In Sect. 10.2, we will recall some definitions and known results concerning *invariant varieties*, which have been investigated in [11]. Further, we give a new result which allows us to compute the variety V, if all vector fields, for which V is invariant, are given. In the following Sect. 10.3, we consider *controlled invariant varieties*, on the one hand with polynomial, on the other hand with rational feedback laws. Section 10.4 deals with *controlled and conditioned varieties*, based on a definition given in [5]. We will see that this notion leads to the problem of intersecting *affine submodules* of R^m with S^m in the polynomial feedback case and of intersecting *affine fractional R-modules* of $\mathrm{Quot}(R)^m$ with $\mathrm{Quot}(S)^m$ in the rational feedback case. Here, $\mathrm{Quot}(D)$ denotes the quotient field of the commutative domain D. Some solutions will be presented as well as some examples to illustrate these methods.

10.2 Invariant Varieties of Autonomous Systems

For the moment, we will consider autonomous systems of the form

$$\dot{x}(t) = F(x(t)), \quad x(0) = x_0, \tag{10.2}$$

where $F \in R^n$ and $x_0 \in K^n$.

Definition 1 Let $\varphi(t, x_0)$ be the solution of (10.2) at time t, where $t \in J(x_0)$, the maximal existence interval of $\varphi(\cdot, x_0)$. We say that $V \subseteq K^n$ is an *invariant set with respect to* F if $x_0 \in V$ implies $\varphi(t, x_0) \in V$ for all $t \in J(x_0)$.

If V is any variety in K^n, then we define the *vanishing ideal of* V

$$\mathcal{J}(V) = \{p \in R \mid p(v) = 0 \text{ for all } v \in V\}.$$

We recall the following result from [11]:

Theorem 1 *Let a variety $V = \mathcal{V}(I)$ be given, where the ideal I is generated by elements $p_1, \ldots, p_k \in R$.*

(a) If we have

$$(\partial_1 p_i)F_1 + \cdots + (\partial_n p_i)F_n \in I \tag{10.3}$$

for all $i \in \{1, \ldots, k\}$, then V is invariant w.r.t. F.

(b) If V is invariant w.r.t. F, then

$$(\partial_1 p_i)F_1 + \cdots + (\partial_n p_i)F_n \in \mathcal{J}(V)$$

for all $i \in \{1, \ldots, k\}$.

Thus, if $\mathcal{J}(V) = I$ holds, then condition (10.3) for all $1 \leqslant i \leqslant k$ is necessary and sufficient for V being invariant w.r.t. F.

Now let $p_1, \ldots, p_k \in R$ and I, V be as in the theorem above. For all $i \in \{1, \ldots, k\}$ we define

$$\mathcal{N}_i = \ker(\partial_1 p_i, \ldots, \partial_n p_i, p_1, \ldots, p_k) \subseteq R^{n+k}$$

and set $\mathcal{M}_i := \pi(\mathcal{N}_i)$, where π denotes the projection on the first n components. Finally, let

$$\mathcal{M} := \bigcap_{i=1}^{k} \mathcal{M}_i \subseteq R^n. \tag{10.4}$$

Again, the next result can be found in [11].

Theorem 2 *We have*

$$\mathcal{M} = \{F \in R^n \mid F \text{ satisfies (10.3) for all } 1 \leqslant i \leqslant k\}.$$

Now assume that $\mathcal{J}(V) = \mathcal{J}(\mathcal{V}(I)) = I$. Then Theorem 2 says that V is invariant w.r.t. $F \in R^n$ if and only if $F \in \mathcal{M}$. For this reason, we call $\mathcal{M}$ the *module of admissible vector fields of* V.

Because of

$$\sum_{l=1}^{n}(\partial_l p_i)\cdot\left(\sum_{j=1}^{n} q_j e_j\right)_l = \sum_{l=1}^{n}(\partial_l p_i)\cdot q_l \in I$$

for all $i = 1,\ldots,k$ and $q_j \in I$, we always have

$$\mathcal{M} \supseteq I \cdot R^n. \tag{10.5}$$

Suppose now that the module $\mathcal{M}$ of admissible vector fields of a variety V is given. A natural question arising is whether we may compute the ideal I (or the variety V, resp.) just by the knowledge of $\mathcal{M}$. The next theorem gives an answer to this question under some kind of "smoothness" assumption:

Theorem 3 *Let* $J = \langle \partial_i p_j \mid i = 1,\ldots,n,\ j = 1,\ldots,k\rangle$. *If* I *and* J *satisfy*

$$I = \mathcal{J}(\mathcal{V}(I)) \text{ and } \mathcal{V}(I)\cap\mathcal{V}(J) = \emptyset,$$

then we can compute I *just by the knowledge of* $\mathcal{M}$ *in the following way:*
For $i = 1,\ldots,n$ *let*

$$I_i := \{q \in R \mid qe_i \in \mathcal{M}\}.$$

Then

$$I = \bigcap_{i=1}^{n} I_i.$$

Proof First note that the I_i defined in the assertion are ideals in R, since $\mathcal{M}$ is an R-module. Because of (10.5) we have $I \subseteq \bigcap_{i=1}^{n} I_i$.

For the other inclusion let $q \in \bigcap_{i=1}^{n} I_i$, which means that $qe_i \in \mathcal{M}$ for all $i = 1,\ldots,n$. This yields

$$\sum_{l=1}^{n}(\partial_l p_j)\cdot(qe_i)_l = q\cdot\partial_i p_j \in I \text{ for all } i = 1,\ldots,n,\ j = 1\ldots,k,$$

and thus

$$q(x)\cdot\partial_i p_j(x) = 0 \text{ for all } i = 1,\ldots,n,\ j = 1\ldots,k \text{ and } x \in \mathcal{V}(I).$$

If $q(x) \neq 0$ for one element $x \in \mathcal{V}(I)$, then $\partial_i p_j(x) = 0$ for all $i = 1,\ldots,n$, $j = 1\ldots,k$, so $x \in \mathcal{V}(J)\cap\mathcal{V}(I)$, which contradicts the assumption. We conclude $q(x) = 0$ for all $x \in \mathcal{V}(I)$, hence $q \in \mathcal{J}(\mathcal{V}(I)) = I$, which completes the proof. ■

Remark 1 For an arbitrary R-module $\mathcal{N} \subseteq R^n$ we define the ideal

$$\operatorname{ann}(R^n/\mathcal{N}) := \{q \in R \mid q[x] = [0] \text{ for all } x \in R^n\},$$

called the *annihilator of* $R^n/\mathcal{N}$. With the notations of Theorem 3 we get

$$\begin{aligned}\operatorname{ann}(R^n/\mathcal{M}) &= \{q \in R \mid q[x] = [0] \text{ for all } x \in R^n\} = \{q \in R \mid qx \in \mathcal{M} \text{ for all } x \in R^n\} \\ &= \{q \in R \mid qe_i \in \mathcal{M} \text{ for all } i = 1, \ldots, n\} = \bigcap_{i=1}^{n} I_i\end{aligned}$$

(note that the second to last equality stays true, since R is commutative). Thus, under the assumptions of Theorem 3, we may conclude that

$$I = \operatorname{ann}(R^n/\mathcal{M}).$$

Example 1 Consider $R = \mathbb{R}[x, y, z]$, $p_1 = xy - z$ and $p_2 = xz - y$, which generate the ideal $I = \langle p_1, p_2\rangle \subseteq R$. We claim that the vanishing set $V = \mathcal{V}(I) = \mathcal{V}(p_1) \cap \mathcal{V}(p_2)$ is, as the intersection of two hypersurfaces defined by p_1 and p_2, a variety with dimension 1, which is nonsmooth.

The Jacobian matrix of I with respect to p_1, p_2 is given by

$$\text{Jac} = \begin{pmatrix} y & x & -1 \\ z & -1 & x \end{pmatrix}.$$

The generic rank of Jac is 2 and thus

$$\dim(V) = 3 - \operatorname{rank}_R(\text{Jac}) = 3 - 2 = 1.$$

Further the ideal generated by the 2×2 minors of Jac is given by

$$L = \langle -y + xz, xy + z, x^2 - 1\rangle$$

and $\mathcal{V}(L) = \{(1, 0, 0), (-1, 0, 0)\}$ is exactly the set of singular points of V (i.e. the points p in which $\operatorname{rank}_K(\text{Jac}(p)) < \operatorname{rank}_R(\text{Jac})$). Thus $\mathcal{V}(I)$ is not smooth, but the ideal J generated by all partial derivatives of p_1 and p_2 is $J = \langle 1\rangle = \mathbb{R}[x, y, z]$, thus $\mathcal{V}(I) \cap \mathcal{V}(J) = \emptyset$ and the assumption of Theorem 3 is satisfied.

We use SINGULAR to do some computations:

```
ring R=0,(x,y,z),dp;
LIB "matrix.lib";
poly p1=xy-z;
poly p2=xz-y;
matrix k1[1][5]=diff(p1,x),diff(p1,y),diff(p1,z),p1,p2;
module n1=syz(k1);
```

```
module m1=submat(n1,1..3,1..ncols(n1));
matrix k2[1][5]=diff(p2,x),diff(p2,y),diff(p2,z),p1,p2;
module n2=syz(k2);
module m2=submat(n2,1..3,1..ncols(n2));
module M=intersect(m1,m2);
```

SINGULAR returns the following module of admissible vector fields of V:

$$\mathcal{M} = \left\langle \begin{pmatrix} 0 \\ y \\ z \end{pmatrix}, \begin{pmatrix} 0 \\ z \\ y \end{pmatrix}, \begin{pmatrix} xz - y \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ xz \\ z \end{pmatrix}, \begin{pmatrix} 0 \\ z \\ xz \end{pmatrix}, \begin{pmatrix} y^2 - z^2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} xy - z \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} x^2 - 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle.$$

Now we can compute the ideals I_i defined in Theorem 3:

```
module e1=[1,0,0];
module in1=intersect(M,e1);
ideal I1=submat(in1,1,1..ncols(in1));
module e2=[0,1,0];
module in2=intersect(M,e2);
ideal I2=submat(in2,2,1..ncols(in2));
module e3=[0,0,1];
module in3=intersect(M,e3);
ideal I3=submat(in3,3,1..ncols(in3));
```

This yields

$$I_1 = \langle xz - y, y^2 - z^2, xy - z \rangle,\ I_2 = \langle xz - y, y^2 - z^2, xy - z \rangle,$$
$$I_3 = \langle xz - y, y^2 - z^2, xy - z, x^2 - 1 \rangle.$$

Finally, the intersection of these three ideals may be derived by

```
ideal I=intersect(I1,I2,I3);
```

which gives the expected result

$$\bigcap_{i=1}^{3} I_i = \langle xz - y, \underbrace{y^2 - z^2}_{zp_1 - yp_2}, xy - z \rangle = I.$$

10.3 Controlled Invariant Varieties

Let us now look at the original system (10.1). For the input $u(\cdot)$, we use a state feedback law, that is, $u(t) = \alpha(x(t))$ for an $\alpha \in R^m$. Plugging this into (10.1) yields the closed loop system

$$\dot{x}(t) = f(x(t)) + g(x(t))\alpha(x(t)) = (f + g\alpha)(x(t)).$$

Definition 2 If $V \subseteq K^n$ is a variety, we call it *controlled invariant w.r.t. system* (10.1) if there is an $\alpha \in R^m$ such that V is invariant w.r.t. $F := f + g\alpha$.

This definition as well as Theorems 1 and 2 immediately yield the following:

Corollary 1 *Provided that $\mathcal{J}(V) = I$ holds, V is controlled invariant w.r.t. system (10.1) if and only if there is an $\alpha \in R^m$ with $f + g\alpha \in \mathcal{M}$, which is equivalent to $f \in \mathcal{M} + \mathrm{im}(g)$.*

If we assume that the given ideal I fulfills the assumption of Corollary 1, a state feedback law making V invariant can be obtained as follows: Because R is a noetherian ring, there is a finite system of generators for the R-module $\mathcal{M}$. Collecting these generators in a matrix M, we can test whether V is controlled invariant by answering the following question: Can we write f as an R-linear combination of the columns of M and g? If the answer is yes, we can find $\alpha \in R^m$ and β with entries in R and of appropriate size such that

$$f = M\beta - g\alpha,$$

and thus $f + g\alpha \in \mathcal{M}$, that is, α is a feedback law that makes V invariant w.r.t. (10.1).

Definition 3 If there exists $\alpha \in R^m$ satisfying $f + g\alpha \in \mathcal{M}$, then we call α *admissible feedback law for* V w.r.t. (10.1).

10.3.1 Nonuniqueness of Admissible Feedback Laws

Let us determine the nonuniqueness of admissible feedback laws α for V w.r.t. (10.1). For this, let $\alpha_1, \alpha_2 \in R^m$ fulfill $f + g\alpha_1 \in \mathcal{M}$ and $f + g\alpha_2 \in \mathcal{M}$. Then there are β_1, β_2 with entries in R having appropriate sizes such that $f + g\alpha_1 = M\beta_1$ and $f + g\alpha_2 = M\beta_2$. Subtraction yields

$$0 = M(\beta_1 - \beta_2) - g(\alpha_1 - \alpha_2) = (M, -g)\begin{pmatrix}\beta_1 - \beta_2 \\ \alpha_1 - \alpha_2\end{pmatrix}.$$

We conclude that the set of all state feedback laws making V an invariant variety is given by

$$\alpha + \pi(\ker(M, -g)), \tag{10.6}$$

where $\alpha \in R^m$ is a particular solution, i.e. $f + g\alpha \in \mathcal{M}$, and π denotes the projection on the last m components.

Example 2 Let $R = \mathbb{R}[x, y, z]$ and the variety $V = \mathcal{V}(I)$ be defined by the ideal $I = \langle 2x^2 + 2y^2 - 1, 2z^2 - 1\rangle$. Further consider control system (10.1) with

$$f = \begin{pmatrix} z \\ y \\ x \end{pmatrix} \in R^3 \text{ and } g = \begin{pmatrix} x & 0 \\ 0 & y \\ z & 0 \end{pmatrix} \in R^{3\times 2}.$$

In the same way as in Example 1 we compute $\mathcal{M} = \text{im}(M)$ to find

$$M = \begin{pmatrix} -y & 2xy & 2z^2-1 & 0 & 0 & 2x^2+2y^2-1 & 0 \\ x & 2y^2-1 & 0 & 2z^2-1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2z^2-1 & 0 & 2x^2+2y^2-1 \end{pmatrix}.$$

If we want to decide whether V is controlled invariant for the system, we check whether $f \in \mathcal{M} + \text{im}(g)$:

```
module L=M,-g;
NF(f,std(L));
```

SINGULAR returns 0, which means that $f \in \mathcal{L} = \mathcal{M} + \text{im}(g)$. Now we compute an admissible feedback law:

```
matrix coeff=lift(L,f);
matrix alpha=submat(coeff,8..9,1);
```

The `lift` command computes a representation of f as a linear combination of the generators of $\mathcal{L}$. The last 2 components of this result are the coefficients of g and equal $\alpha = \begin{pmatrix} -2xz \\ -2xz - 1 \end{pmatrix}$. Finally we want to find the nonuniqueness of admissible feedbacks. For this we compute

```
matrix s=syz(L);
matrix P=submat(s,8..9,1..ncols(s));
```

Let $\mathcal{P} = \text{im}(P)$. Then the set of admissible state feedbacks is given by $\alpha + \mathcal{P}$

$$= \begin{pmatrix} -2xz \\ -2xz - 1 \end{pmatrix} + \left\langle \begin{pmatrix} 2z^2-1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2x^2+2y^2-1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2z^2-1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2x^2+2y^2-1 \end{pmatrix} \right\rangle,$$

which also can be written as $\alpha + I \cdot R^2$.

In the last example, the nonuniqueness of admissible feedback laws is just determined by the module $I \cdot R^m$. We want to give some characterisation whether this is the case in terms of g and I.

For this we consider a slightly generalised framework: Let $I \subseteq R$ be an ideal and $\mathcal{M} \subseteq R^n$ be an R-module such that $\mathcal{M} \supseteq I \cdot R^n$. Further let $\mathcal{N} \subseteq R^n$ be an R-module, and $M \in R^{n\times l}$, $N \in R^{n\times m}$ be matrices over R with $\mathrm{im}(M) = \mathcal{M}$ and $\mathrm{im}(N) = \mathcal{N}$. Now we define another R-module $\mathcal{P} := \pi(\ker(M, -N))$, where π is the projection onto the last m components.

Lemma 1 *We always have*

$$\mathcal{M} \cap \mathcal{N} \supseteq I \cdot \mathcal{N} \text{ and } \mathcal{P} \supseteq I \cdot R^m.$$

Proof Since $\mathcal{N} \supseteq I \cdot \mathcal{N}$ and $\mathcal{M} \supseteq I \cdot R^n \supseteq I \cdot \mathcal{N}$ the first inclusion is clear. For the second let $p_i \in I$ be arbitrary. Then

$$N\left(\sum_{i=1}^{m} p_i e_i\right) = \sum_{i=1}^{m} p_i (N e_i) \in I \cdot \mathcal{N} \subseteq I \cdot R^n \subseteq \mathcal{M},$$

so there is $y \in R^l$ with $N(\sum_{i=1}^m p_i e_i) = My$. Thus we have

$$0 = (M, -N) \begin{pmatrix} y \\ \sum_{i=1}^m p_i e_i \end{pmatrix},$$

which means $\sum_{i=1}^m p_i e_i \in \pi(\ker(M, -N)) = \mathcal{P}$. ■

Theorem 4 *The following statements are equivalent:*

(a) $\mathcal{P} = I \cdot R^m$.
(b) $\ker(N) \subseteq I \cdot R^m$ *and* $\mathcal{M} \cap \mathcal{N} = I \cdot \mathcal{N}$.

Proof 1. $\Rightarrow$ 2.: If $y \in \ker(N)$, then $(M, -N)\begin{pmatrix} 0 \\ y \end{pmatrix} = 0$, so $y \in \mathcal{P} = I \cdot R^m$. Now $f \in \mathcal{M} \cap \mathcal{N}$ implies $f = Mx = Ny$ for some $x \in R^l$, $y \in R^m$. This means $y \in \mathcal{P} = I \cdot R^m$, so we can write $y = \sum_{i=1}^m p_i e_i$, where $p_i \in I$. But then

$$f = Ny = \sum_{i=1}^{m} p_i (N e_i) \in I \cdot \mathcal{N}.$$

2. $\Rightarrow$ 1.: If $y \in \mathcal{P}$, then there exists $x \in R^l$ with $0 = (M, -N)\begin{pmatrix} x \\ y \end{pmatrix}$, so we have

$$Mx = Ny \in \mathcal{M} \cap \mathcal{N} = I \cdot \mathcal{N}$$

Let $p_i \in I$ and $y_i \in R^m$ such that $Ny = \sum_{i=1}^r p_i (N y_i) = N(\sum_{i=1}^r p_i y_i)$. This means $N(y - \sum_{i=1}^r p_i y_i) = 0$, so $y - \sum_{i=1}^r p_i y_i \in \ker(N) \subseteq I \cdot R^m$ and finally $y \in I \cdot R^m$. ■

To come back to our original setting we just put $\mathcal{M}$ as defined in (10.4), $\mathcal{N} = \mathrm{im}(g)$ and $\mathcal{P} = \pi(\ker(M, -g))$ to obtain the following result:

Corollary 2 *We have* $\mathcal{P} = I \cdot R^m$ *iff* $\mathcal{M} \cap \mathrm{im}(g) = I \cdot \mathrm{im}(g)$ *and* $\ker(g) \subseteq I \cdot R^m$.

Example 3 For the data in Example 2, we derive $\mathcal{M} \cap \mathrm{im}(g) =$

$$\langle \begin{pmatrix} x(2z^2-1) \\ 0 \\ z(2z^2-1) \end{pmatrix}, \begin{pmatrix} 0 \\ y(2z^2-1) \\ 0 \end{pmatrix}, \begin{pmatrix} x(2x^2+2y^2-1) \\ 0 \\ z(2x^2+2y^2-1) \end{pmatrix}, \begin{pmatrix} 0 \\ y(2x^2+2y^2-1) \\ 0 \end{pmatrix} \rangle = I \cdot \mathrm{im}(g).$$

Further we have $\ker(g) = 0$. Using Corollary 2, this shows $\mathcal{P} = I \cdot R^2$ as already seen in Example 2, but derived with different methods.

Now we keep g from the system of Example 2, but consider the ideal $I = \langle p_1, p_2 \rangle$, where $p_1 = xy - z$, $p_2 = xz - y$ (c.f. Example 1). Then we have

$$\begin{aligned} \mathcal{M} \cap \mathrm{im}(g) &= \langle \begin{pmatrix} xp_1 \\ 0 \\ zp_1 \end{pmatrix}, \begin{pmatrix} 0 \\ yp_1 \\ 0 \end{pmatrix}, \begin{pmatrix} xp_2 \\ 0 \\ zp_2 \end{pmatrix}, \begin{pmatrix} 0 \\ yp_2 \\ 0 \end{pmatrix}, \begin{pmatrix} x(x^2-1) \\ 0 \\ z(x^2-1) \end{pmatrix}, \begin{pmatrix} 0 \\ y(x^2-1) \\ 0 \end{pmatrix} \rangle \\ &= I \cdot \mathrm{im}(g) + (x^2-1) \cdot \mathrm{im}(g) \supsetneq I \cdot \mathrm{im}(g) \end{aligned}$$

and thus $\mathcal{P} \neq I \cdot R^2$. In fact, one may derive

$$\mathcal{P} = I \cdot R^2 + (x^2 - 1)R^2.$$

10.3.2 Rational Feedback

In this section we want to answer the question whether there are some new admissible feedbacks, when we not just allow them to be polynomial, but also rational functions in the state $x(t)$. In order to do this, we refine some of our notations: Let

$$Q := \mathrm{Quot}(R) := \left\{ \frac{p}{q} \mid p, q \in R, q \neq 0 \right\}$$

be the quotient field of R. For a matrix $A \in R^{n \times m}$ and $S \in \{R, Q\}$ we write

$$\mathrm{im}_S(\cdot A) = \{xA \mid x \in S^{1 \times n}\} \text{ and } \ker_S(\cdot A) = \{x \in S^{1 \times n} \mid xA = 0\}$$

for the left image (or left kernel, resp.) of A over S and

$$\mathrm{im}_S(A \cdot) = \{Ax \mid x \in S^m\} \text{ and } \ker_S(A \cdot) = \{x \in S^m \mid Ax = 0\}$$

for the right image (or right kernel, resp.) of A over S.

Lemma 2 *Let $A \in R^{n\times m}$ be given. Consider matrices $B \in R^{q\times n}$ and $\tilde{A} \in R^{n\times r}$ satisfying*

$$\mathrm{im}_R(\cdot B) = \ker_R(\cdot A) \text{ and } \ker_R(B\cdot) = \mathrm{im}_R(\tilde{A}\cdot).$$

Then the following hold:

(a) There exists $X \in R^{r\times m}$ with $A = \tilde{A}X$.
(b) There exists $Y \in Q^{m\times r}$ with $\tilde{A} = AY$.

Proof 1. If we have

$$\mathrm{im}_R(A\cdot) \subseteq \mathrm{im}_R(\tilde{A}\cdot), \tag{10.7}$$

then each column a_i of A can be written as $a_i = \tilde{A}x_i$ for some $x_i \in R^r$, and putting $X = (x_1, \ldots, x_m) \in R^{r\times m}$ yields 1. But (10.7) is true, since for all $y = Ax \in \mathrm{im}_R(A\cdot)$ the assumption implies $By = BAx = 0$ and thus $y \in \ker_R(B\cdot) = \mathrm{im}_R(\tilde{A}\cdot)$.

2. First we prove that we already have

$$\mathrm{im}_Q(\cdot B) = \ker_Q(\cdot A). \tag{10.8}$$

The relation $BA = 0$ gives us $\mathrm{im}_Q(\cdot B) \subseteq \ker_Q(\cdot A)$. Now let $y = \frac{\tilde{y}}{d} \in \ker_Q(\cdot A)$ for some $\tilde{y} \in R^{1\times n}, 0 \neq d \in R$. Then $0 = yA = \frac{\tilde{y}}{d} \cdot A$ implies $\tilde{y}A = 0$, and thus $\tilde{y} \in \ker_R(\cdot A) = \mathrm{im}_R(\cdot B)$. So there is $\tilde{x} \in R^{1\times q}$ with $\tilde{y} = \tilde{x}B$ and we set $x := \frac{\tilde{x}}{d}$ to obtain the desired result $y = \frac{\tilde{y}}{d} = \frac{\tilde{x}}{d} \cdot B = xB \in \mathrm{im}_Q(\cdot B)$. Similarly one can show

$$\ker_Q(B\cdot) = \mathrm{im}_Q(\tilde{A}\cdot), \tag{10.9}$$

and completely analogous to (10.7) we get $\mathrm{im}_Q(A\cdot) \subseteq \mathrm{im}_Q(\tilde{A}\cdot)$. Since Q is a field we can use dimension formulas for linear maps, and the fact that row and column rank of matrices coincide. This gives us

$$\begin{aligned}\dim(\mathrm{im}_Q(A\cdot)) = \mathrm{rank}(A) &\overset{(10.8)}{=} n - \mathrm{rank}(B)\\ &\overset{(10.9)}{=} \mathrm{rank}(\tilde{A}) = \dim(\mathrm{im}_Q(\tilde{A}\cdot)).\end{aligned}$$

We may conclude

$$\mathrm{im}_Q(A\cdot) = \mathrm{im}_Q(\tilde{A}\cdot)$$

and now the same arguments as in the proof of 1. yield assertion 2. ■

Remark 2 The results of Lemma 2 can also be obtained by using exact sequences, extension modules, and the fact that Q is a flat R-module. The definition of extension modules and methods to compute them can be found e.g. in [3] or [10].

With this preparation we can give a method to find admissible feedback laws in Q^m, just by doing computations over R.

Lemma 3 *Let $k \in R^{q\times n}$, $\tilde{g} \in R^{n\times r}$ be matrices satisfying*

$$\mathrm{im}_R(\cdot k) = \ker_R(\cdot g) \text{ and } \ker_R(k\cdot) = \mathrm{im}_R(\tilde{g}\cdot).$$

Then the following are equivalent:

(a) There exists $\alpha \in R^r$ such that $f + \tilde{g}\alpha \in \mathcal{M}$.
(b) There exists $\tilde{\alpha} \in Q^m$ such that $f + g\tilde{\alpha} \in \mathcal{M}$.

Proof Using Lemma 2 we can find matrices X over R and Y over Q satisfying

$$g = \tilde{g}X \text{ and } \tilde{g} = gY.$$

Now, if $f + \tilde{g}\alpha \in \mathcal{M}$, where $\alpha \in R^r$, then $\tilde{\alpha} := Y\alpha \in Q^m$ satisfies

$$\mathcal{M} \ni f + \tilde{g}\alpha = f + gY\alpha = f + g\tilde{\alpha}.$$

This shows 1. $\Rightarrow$ 2. For the converse let $\tilde{\alpha} \in Q^m$ with $f + g\tilde{\alpha} \in \mathcal{M}$ and $\alpha_0 \in R^m$, $0 \neq d \in R$ such that $\tilde{\alpha} = \frac{\alpha_0}{d}$. There exists $m \in \mathcal{M}$ with

$$f - m = -g\tilde{\alpha} = -g \cdot \frac{\alpha_0}{d},$$

which implies

$$d(f - m) = -g\alpha_0 \in \mathrm{im}_R(g\cdot).$$

Then

$$kd(f - m) = -kg\alpha_0 = 0 \text{ and thus } k(f - m) = 0.$$

This shows $f - m \in \ker_R(k\cdot) = \mathrm{im}_R(\tilde{g}\cdot)$, so there is $\alpha \in R^r$ such that $f - m = -\tilde{g}\alpha$ or reformulated

$$f + \tilde{g}\alpha \in \mathcal{M},$$

which completes the proof. ■

Similarly as in the polynomial case, we want to find out how nonunique a rational admissible feedback law α can be. If $\alpha_1, \alpha_2 \in Q^m$ are such that $f + g\alpha_1 \in \mathcal{M}$ and $f + g\alpha_2 \in \mathcal{M}$, then $g(\alpha_1 - \alpha_2) \in \mathcal{M}$. Thus the set of all rational admissible feedback laws has the form

$$\alpha + \{\bar{\alpha} \in Q^m \mid g\bar{\alpha} \in \mathcal{M}\},$$

where $\alpha \in Q^m$ is one particular rational feedback law making V invariant. Since $\mathcal{M}$ is just an R-module and we are looking for elements $\bar{\alpha} \in Q^m$, we get some mixup of structures:

Lemma 4 *Let $\mathcal{M} = \mathrm{im}_R(M\cdot) \subseteq R^n$ be an R-module and $g \in R^{n\times m}$. Consider $k \in R^{q\times n}$ and $\tilde{g} \in R^{n\times r}$ constructed in Lemma 3 as well as $Y \in Q^{m\times r}$ with $\tilde{g} = gY$. Then we have*

$$\{\bar{\alpha} \in Q^m \mid g\bar{\alpha} \in \mathcal{M}\} = \ker_Q(g\cdot) + Y\pi(\ker_R((M, -\tilde{g})\cdot)),$$

where π denotes the projection onto the last r components.

Proof If $\bar{\alpha} = \beta + \gamma$, where $\beta \in \ker_Q(g\cdot)$ and $\gamma = Y\gamma_1$, $\gamma_1 \in \pi(\ker_R((M, -\tilde{g})\cdot))$, then there is γ_2 such that $0 = (M, -\tilde{g})\begin{pmatrix}\gamma_2\\ \gamma_1\end{pmatrix} = M\gamma_2 - \tilde{g}\gamma_1$. This shows

$$g\bar{\alpha} = \underbrace{g\beta}_{=0} + g\gamma = gY\gamma_1 = \tilde{g}\gamma_1 = M\gamma_2 \in \mathcal{M}.$$

This gives us "$\supseteq$".

For the other inclusion let $\bar{\alpha} \in Q^m$ with $g\bar{\alpha} = Mx$ for some x. Then $kg\bar{\alpha} = 0$, which implies $g\bar{\alpha} \in \mathrm{im}_R(\tilde{g}\cdot)$, so there is $\hat{x}$ satisfying $\tilde{g}\hat{x} = gY\hat{x} = g\bar{\alpha}$. Thus $\bar{\alpha} - Y\hat{x} \in \ker_Q(g\cdot)$ and because of $\tilde{g}\hat{x} - Mx = \tilde{g}\hat{x} - g\bar{\alpha} = 0$ we may conclude $\hat{x} \in \pi(\ker_R((M, -\tilde{g})\cdot))$. Finally

$$\bar{\alpha} \in \ker_Q(g\cdot) + Y\pi(\ker_R((M, -\tilde{g})\cdot))$$

which we wanted to show. ■

Definition 4 If $Y \in Q^{m\times r}$ and $\mathcal{N} \subseteq R^r$ is an R-module, we can write

$$Y\mathcal{N} = \langle\frac{\tilde{c_1}}{d_1}, \ldots, \frac{\tilde{c_k}}{d_k}\rangle_R = \frac{1}{d}\cdot\langle c_1, \ldots, c_k\rangle_R \text{ for some } c_i, \tilde{c_i} \in R^m, d_i, d \in R.$$

Then we call $Y\mathcal{N}$ a *fractional R-module*.

Remark 3 Definition 4 is based on the notion of "fractional ideals", see e.g. [2, Chap. VII] or [7, Chap. 11] for an introduction. This concept has turned out to be useful in system theoretic applications before, cf. [9] or [8] and the references therein.

Example 4 Let R, I, V, $\mathcal{M}$ be as in Example 1 and the control system be defined by

$$\dot{x} = f(x) + g(x)u = \begin{pmatrix}-y\\ x\\ z\end{pmatrix} + \begin{pmatrix}0 & z\\ -z & 0\\ y & -x\end{pmatrix}u.$$

First we check whether $f \in \mathcal{M} + \mathrm{im}_R(g\cdot)$:

```
matrix f[3][1]=-y,x,z;
matrix g[3][2]=0,z,-z,0,y,-x;
module pol=M,-g;
NF(f,std(pol));
```

SINGULAR returns a nonzero result, which means that $f \notin \mathcal{M} + \mathrm{im}_R(g\cdot)$. So we try to find rational feedbacks, which make V invariant w.r.t. f and compute the matrices $k, \tilde{g}$ from Lemma 3:

```
matrix k=transpose(syz(transpose(g)));
matrix gt=syz(k);
```

This yields

$$k = \begin{pmatrix} x & y & z \end{pmatrix} \quad \text{and} \quad \tilde{g} = \begin{pmatrix} 0 & -y & -z \\ -z & x & 0 \\ y & 0 & x \end{pmatrix}.$$

We want to find out if $f \in \mathcal{M} + \mathrm{im}_R(\tilde{g}\cdot)$:

```
module rat=M,-gt;
NF(f,std(rat));
```

Now the result is 0 and we may conclude that there indeed exists a rational admissible feedback law $\tilde{\alpha}$, which we may obtain in the following way:

```
matrix coeff=lift(rat,f);
matrix alpha=submat(coeff,ncols(M)+1..ncols(M)+ncols(gt),
1);
```

The outcome

$$\alpha = \begin{pmatrix} 0 \\ \frac{1}{2}z - 1 \\ -\frac{1}{2}y \end{pmatrix} \text{ satisfies } f + \tilde{g}\alpha = \begin{pmatrix} 0 \\ \frac{1}{2}xz \\ -\frac{1}{2}xy + z \end{pmatrix} \in \mathcal{M}.$$

To get $\tilde{\alpha}$ we have to find a transformation matrix $Y \in Q^{2\times 3}$ with $\tilde{g} = gY$:

```
ring Q=(0,x,y,z),w,dp;
matrix g=imap(r,g);
matrix gt=imap(r,gt);
matrix Y=lift(g,gt);
```

This yields

$$Y = \frac{1}{z} \cdot \begin{pmatrix} z & -x & 0 \\ 0 & -y & -z \end{pmatrix} \text{ and thus } \tilde{\alpha} = Y\alpha = \frac{1}{z} \cdot \begin{pmatrix} -\frac{1}{2}xz + x \\ y \end{pmatrix}.$$

Now this $\tilde{\alpha}$ fulfills $f + g\tilde{\alpha} \in \mathcal{M}$.

In order to derive the nonuniqueness of rational admissible feedbacks we compute the R-module $\mathcal{N} = \pi(\ker_R((M, -\tilde{g})\cdot))$:

```
setring R;
matrix s=syz(rat);
module N=submat(s,ncols(M)+1..ncols(M)+ncols(gt),1..
ncols(s));
```

Since $\ker_Q(g\cdot) = 0$ the set $\{\bar{\alpha} \in Q^3 \mid g\bar{\alpha} \in \mathcal{M}\}$, which determines the nonuniqueness of rational admissible feedback laws, is equal to

$$Y\mathcal{N} = \frac{1}{z} \cdot \langle \begin{pmatrix} xy - z \\ y^2 - z^2 \end{pmatrix}, \begin{pmatrix} xz^2 - yz \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ xz^2 - yz \end{pmatrix}, \begin{pmatrix} y^2z - z^3 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ y^2z - z^3 \end{pmatrix}, \begin{pmatrix} 0 \\ xyz - z^2 \end{pmatrix}, \begin{pmatrix} x^2z - z \\ 0 \end{pmatrix} \rangle.$$

10.4 Controlled and Conditioned Invariant Varieties

Let us now come back to the polynomial setting and the notation used in Sects. 10.1 and 10.2. So far, we have seen how to find a state feedback function α such that V is invariant for (10.1) with $u(t) = \alpha(x(t))$, if this is possible at all. However, in general it is restrictive to assume that the full state is available for the feedback. Instead, we would like to use only the output $y(t) = h(x(t))$ of (10.1) for the feedback. In view of this, we ask the following question: Is it possible to choose an output feedback law $u(t) = \beta(y(t)) = \beta(h(x(t)))$ making V invariant?

Definition 5 Given a variety V, we call it *conditioned and controlled invariant w.r.t. system* (10.1) if there is a polynomial state feedback law $\alpha \in R^m$ as in Definition 2 which additionally takes the form $\alpha = \beta(h_1, \ldots, h_p)$ for some $\beta \in T^m$, i.e.

$$\alpha \in S^m = K[h_1, \ldots, h_p]^m. \tag{10.10}$$

If we wish to decide whether a variety V satisfies Definition 5, we have to determine the set (10.6) and check whether one of its elements lies in S^m. More generally, we intend to compute the set

$$(v + \mathcal{P}) \cap K[h_1, \ldots, h_p]^m \tag{10.11}$$

for a given R-module $\mathcal{P} \subseteq R^m$, $v \in R^m$, and $h_1, \ldots, h_p \in R$. The main goal of this section is to analyse the structure of this set and to give algorithms for its computation. The algebraic methods we present are not limited to $K \in \{\mathbb{R}, \mathbb{C}\}$, so let K be an arbitrary field.

10.4.1 Intersection of an Ideal and a Subalgebra

Our first step will be to give a method to compute (10.11) in the special case $v = 0$ (or $v \in \mathcal{P}$). For this, we define two K-algebra homomorphisms

$$\begin{aligned} \phi : T \to R,\ & y_j \mapsto h_j,\ j = 1, \ldots, p, \text{ and} \\ \psi : R_1 \to R,\ & x_i \mapsto x_i,\ i = 1, \ldots, n, \\ & y_j \mapsto h_j,\ j = 1, \ldots, p. \end{aligned}$$

The next theorem gives some properties of these maps:

Theorem 5 *Let ϕ and ψ be defined as above.*

(a) We have $\mathrm{im}(\phi) = \mathrm{im}(\psi|_T) = S$. Hence we may write $\phi : T \twoheadrightarrow S$ and thus $\phi(\phi^{-1}(I)) = I \cap S$ for all subsets $I \subseteq R$.

(b) $\psi|_R = \mathrm{id}_R$, $\psi|_T = \phi$, so we have the following commutative diagram:

$$\begin{array}{ccc} R_1 & \xrightarrow{\psi} & R \\ \uparrow & & \uparrow \\ T & \xrightarrow{\phi} & S \end{array}$$

(c) If $I \subseteq R$ is an ideal, then

$$\phi^{-1}(I) = (J_\phi + I) \cap T \text{ and } \psi^{-1}(I) = J_\phi + I, \text{ where} \tag{10.12}$$

$$J_\phi := \langle y_j - \phi(y_j) \mid j = 1, \ldots, p \rangle_{R_1}. \tag{10.13}$$

Proof 1. and 2. can be verified easily and 3. follows from [6], Theorem 3.2.1. ■

Since an algorithm for the elimination of variables needed in (10.12) is well known (see e.g. [4, Chap. 1]), we can use Theorem 5 to find an algorithmic method to compute the S-module $I \cap S$ for any ideal $I \subseteq R$.

For arbitrary $m \in \mathbb{N}$, we define the extension maps to free modules T^m (resp. R_1^m) induced by ϕ (resp. ψ), also denoted by ϕ (resp. ψ):

$$\phi : T^m \to R^m, \quad \sum_{i=1}^{m} a_i e_i \mapsto \sum_{i=1}^{m} \phi(a_i) e_i, \tag{10.14}$$
$$\psi : R_1^m \to R^m, \quad \sum_{i=1}^{m} a_i e_i \mapsto \sum_{i=1}^{m} \psi(a_i) e_i.$$

Here $a_i \in T$ (resp. $a_i \in R_1$) are arbitrary polynomials.

After a short computation, we get

$$\phi\left(\sum_{i=1}^{s} a_i t_i\right) = \sum_{i=1}^{s} \phi(a_i)\phi(t_i), \ s \in \mathbb{N}, \ a_i \in T, \ t_i \in T^m, \tag{10.15}$$
$$\psi\left(\sum_{i=1}^{s} b_i r_i\right) = \sum_{i=1}^{s} \psi(b_i)\psi(r_i), \ s \in \mathbb{N}, \ b_i \in R_1, \ r_i \in R_1^m.$$

If $m_1, \ldots, m_k \in R_1^m$ and R_2 is a subring of R_1, then (10.15) yields

$$\psi(\langle m_1, \ldots, m_k \rangle_{R_2}) = \langle \psi(m_1), \ldots, \psi(m_k) \rangle_{\psi(R_2)}. \tag{10.16}$$

Analogously to Theorem 5, we obtain the following corollary by considering the individual components of the elements of T^m (resp. R_1^m):

Corollary 3 *The extension maps ϕ and ψ have the following properties:*

(a) $\mathrm{im}(\phi) = \mathrm{im}(\psi|_{T^m}) = S^m$, *hence* $\phi(\phi^{-1}(\mathcal{P})) = \mathcal{P} \cap S^m$ *for all subsets* $\mathcal{P} \subseteq R^m$.
(b) $\psi|_{R^m} = \mathrm{id}_{R^m}$, $\psi|_{T^m} = \phi$.
(c) If $\mathcal{P} \subseteq R^m$ *is an* R*-module, then*

$$\phi^{-1}(\mathcal{P}) = (J_\phi^m + \mathcal{P}) \cap T^m \text{ and } \psi^{-1}(\mathcal{P}) = J_\phi^m + \mathcal{P}, \text{ where}$$
$$J_\phi^m := \langle (y_j - \phi(y_j)) \cdot e_i \mid j = 1, \ldots, p, i = 1, \ldots, m \rangle_{R_1}$$

With this result, we are able to compute the intersection of an R-module $\mathcal{P} \subseteq R^m$ with S^m for a finitely generated subalgebra S of R. It will be the basic and frequently used tool for the determination of (10.11).

10.4.2 Intersection of an Affine Ideal and a Subalgebra

Assume there exists $q \in (v + \mathcal{P}) \cap S^m$. Then $q = v + p \in S^m$ for some $p \in \mathcal{P}$ and for every other $q' \in (v + \mathcal{P}) \cap S^m$, there is $p' \in \mathcal{P}$ with $q' = v + p'$. We conclude

$$q' - q = v + p' - v - p = p' - p \in \mathcal{P} \cap S^m,$$

i.e. $q' \in q + \mathcal{P} \cap S^m$. On the other hand, if $q' \in q + \mathcal{P} \cap S^m$, we see

$$q' = q + p' = v + p + p' \in v + \mathcal{P}.$$

Thus

$$(v + \mathcal{P}) \cap S^m = q + \mathcal{P} \cap S^m. \tag{10.17}$$

So our objective is to determine one particular element q of $(v + \mathcal{P}) \cap S^m$ and then we'll get the whole set using (10.17) and Corollary 3. In the following we will use some methods from the theory of Gröbner bases (for an introduction see e.g. [1, Chaps. 1 and 3]): Let $<_{\text{el}}$ be an elimination order (see [1, 2.3.1]) on the monomials of $R_1 = K[\underline{x}, \underline{y}]$ with

$$y^\mu <_{el} x^\nu \text{ for all } \nu \in \mathbb{N}^n \setminus \{0\}, \mu \in \mathbb{N}^p \tag{10.18}$$

(for example the lexicographical order ([1, 1.4.2]) with $x_1 > \cdots > x_n > y_1 > \cdots > y_p$ will do the task). We define the order $<_{\text{TOP}}$ for monomials $qe_i, re_j \in R_1^m$, $q, r \in R_1$ by

$$qe_i <_{\text{TOP}} re_j \Leftrightarrow \begin{cases} q <_{el} r \\ \text{ or if } q = r \\ i <_{\mathbb{N}} j. \end{cases}$$

It is easy to see that this is indeed a term order on R_1^m, defined in [1, 3.5.1]. If G is a Gröbner basis of $\mathcal{P}$ [1, 3.5.13] and $w \in R_1^m$, then we abbreviate the normal form of w w.r.t. G [1, p. 155] by $\text{NF}(w, G)$. Further the leading monomial of w [1, p. 143] is denoted by $\text{lm}(w)$.

The next result treats the computation of (10.17) in the special case $S = T = K[\underline{y}]$.

Lemma 5 *Let $v \in R^m$ and $\mathcal{P}$ be a submodule of $K[\underline{x}, \underline{y}]^m$. Moreover let G be a Gröbner basis of $\mathcal{P}$ w.r.t. $<_{TOP}$. Then the following statements are equivalent:*

(a) $(v + \mathcal{P}) \cap T^m \neq \emptyset$.
(b) $\text{NF}(v, G) \in T^m$.

Proof 2. $\Rightarrow$ 1. : If $w = \text{NF}(v, G) \in T^m$, then there is $p \in \mathcal{P}$ with $w = v + p$. This shows

$$w \in (v + \mathcal{P}) \cap T^m.$$

1. $\Rightarrow$ 2. : Let $w = v + p \in T^m$, where $p \in \mathcal{P}$. If $w = 0$, then $\text{NF}(v, G) = 0$ and we are finished. Otherwise there is $\mu \in \mathbb{N}^p$ and $k \in \{1, \ldots, m\}$ with $y^\mu e_k = \text{lm}(w)$ and we have

$$\hat{w} := \mathrm{NF}(v, G) = \mathrm{NF}(v + p, G) = \mathrm{NF}(w, G).$$

Assume $\hat{w} \notin T^m$. We can find $j \in \{1, \ldots, n\}$ and $i \in \{1, \ldots, m\}$ satisfying

$$x_j e_i \leqslant_{\mathrm{TOP}} \mathrm{lm}(\hat{w}) \leqslant_{\mathrm{TOP}} \mathrm{lm}(w) = y^{\mu} e_k.$$

This implies $y^{\mu} \geqslant_{el} x_j$, which is a contradiction to (10.18). ■

For the general case (i.e. $T = K[\underline{y}]$, $S = K[h_1, \ldots, h_p]$) we define the $K[\underline{x}, \underline{y}]$-module

$$\mathcal{P}_1 := \psi^{-1}(\mathcal{P}) = \mathcal{P} + J_{\phi}^m$$

(c.f. Corollary 3).

Theorem 6 *Let G be a Gröbner basis of $\mathcal{P}_1$ w.r.t. $<_{\mathrm{TOP}}$. We have*

$$(v + \mathcal{P}) \cap S^m = \psi((v + \mathcal{P}_1) \cap T^m).$$

This and Lemma 5 imply the equivalence of the following statements:

(a) $\mathrm{NF}(v, G) \in T^m$.
(b) $(v + \mathcal{P}_1) \cap T^m \neq \emptyset$.
(c) $(v + \mathcal{P}) \cap S^m \neq \emptyset$.

Proof First, let $t \in (v + \mathcal{P}_1) \cap T^m$. Corollary 3, part (a) yields $\psi(t) \in S^m$ and there is an element $w \in \mathcal{P}_1 = \psi^{-1}(\mathcal{P})$ with $t = v + w$. This gives us

$$\psi(t) = \psi(\underbrace{v}_{\in R^m}) + \underbrace{\psi(w)}_{\in \mathcal{P}} = v + \psi(w) \in v + \mathcal{P}.$$

For the inclusion $\subseteq$ let $q \in (v + \mathcal{P}) \cap S^m$, i.e.

$$q = v + w = \psi(t) \text{ for some } w \in \mathcal{P},\ t \in T^m.$$

Since $v, w \in R^m$ we have $\psi(v) = v$ and $\psi(w) = w$, hence

$$\psi(t - v - w) = q - v - w = 0.$$

Using 3. of Corollary 3 we may conclude

$$t - v - w =: j \in \ker(\psi) = \psi^{-1}(0) = J_{\phi}^m.$$

This yields

$$t = v + \underbrace{w + j}_{\in \mathcal{P} + J_\phi^m = \mathcal{P}_1} \in (v + \mathcal{P}_1) \cap T^m,$$

and thus

$$q = \psi(t) \in \psi((v + \mathcal{P}_1) \cap T^m),$$

which proves the assertion. ■

We give a short summary of the last results and a procedure to decide if a variety V is controlled and conditioned invariant: Given f, g, h and I described in the introduction, we compute α and $\mathcal{P} \subseteq R^m$ as mentioned in Sect. 10.3. If there is no α such that $f + g\alpha$ is admissible for V, then we are done and V is not controlled and conditioned invariant. Otherwise we use the results from Sect. 10.4 to compute

$$(\alpha + \mathcal{P}) \cap S^m,$$

where $S := K[h_1, \ldots, h_p]$. For this we define $J_\phi^m \subseteq K[\underline{x}, \underline{y}]^m$ as described in Corollary 3 and compute a Gröbner basis G of $J_\phi^m + \mathcal{P}$ w.r.t. $<_{\text{TOP}}$ defined above. Then the elements of $\psi(G \cap K[\underline{y}]^m)$ generate $\mathcal{P} \cap S^m$. Now compute $t = \text{NF}(\alpha, G)$. If t does not just depend on $\underline{y}$, we may conclude $(\alpha + \mathcal{P}) \cap S^m = \emptyset$ and V is not controlled and conditioned invariant. Otherwise set $\alpha^* := \psi(t)$ to get

$$(\alpha + \mathcal{P}) \cap S^m = \alpha^* + \mathcal{P} \cap S^m,$$

the set of admissible output feedbacks for V.

Example 5 Let $I = \langle x_1x_2 - x_3, x_1x_3 - x_2 \rangle \subseteq \mathbb{R}[x_1, x_2, x_3] =: R$ and the following I/O-system on the variety $V = \mathcal{V}(I)$ be given:

$$\dot{x} = f(x) + g(x)u = \begin{pmatrix} x_1x_2x_3^2 \\ 2x_1x_2x_3 \\ -x_1^2x_2x_3 - 2x_1x_2^2 \end{pmatrix} + \begin{pmatrix} 0 & -x_3 \\ x_3 & 0 \\ -x_2 & x_1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

$$y(x) = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} x_1x_2 \\ x_3^2 \end{pmatrix}.$$

Similarly as in Example 2 one may check that $f + g\alpha \in \mathcal{M}$, where $\mathcal{M} = \text{im}_R(M\cdot)$ is the module of admissible vector fields of I and

$$\alpha = \begin{pmatrix} -\frac{1}{2}x_1^2x_3 - \frac{3}{2}x_1x_2 \\ x_2^2 \end{pmatrix}.$$

The nonuniqueness of admissible feedback laws can be derived as

$$\mathcal{P} = \pi(\ker_R(M, -g)\cdot) = I \cdot R^2 + \langle (x^2 - 1)e_1 \rangle.$$

Now our goal is to decide if V is controlled and conditioned invariant. To achieve this we compute the intersection of the set of admissible feedbacks and the subalgebra generated by the output of the system:

$$(\alpha + \mathcal{P}) \cap K[h_1, h_2] = \left(\begin{pmatrix} -\frac{1}{2}x_1^2 x_3 - \frac{3}{2}x_1 x_2 \\ x_2^2 \end{pmatrix} + I \cdot R^2 + \langle (x^2 - 1)e_1 \rangle \right) \cap K[x_1 x_2, x_3^2].$$

At first we derive $\mathcal{P} \cap K[h_1, h_2]$. For this we define the ring $R_1 = \mathbb{R}[x_1, x_2, x_3, y_1, y_2]$ equipped with the $<_{\text{TOP}}$-order from above and the subalgebra generated by h:

```
int m=2;
ring R1=0,(x(1..3),y(1..2)),lp;
ideal h=x(1)*x(2),x(3)^2;
```

Now we build the module J_ϕ^m from Corollary 3 and compute a Gröbner basis of $J_\phi^m + \mathcal{P}$:

```
module Jphim=(y(1)-h[1])*freemodule(m),(y(2)-h[2])
*freemodule(m);  module pre=Jphim,P;
std(pre);
```

SINGULAR computes 13 generators for this Gröbner basis, from which the first two elements just depend on y_1, y_2 and we map these elements with ψ:

```
subst(std(pre)[1],y(1),h[1],y(2),h[2]);
subst(std(pre)[2],y(1),h[1],y(2),h[2]);
```

The resulting module is given by

$$\mathcal{P} \cap K[h_1, h_2] = (x_1^2 x_2^2 - x_3^2) K[h_1, h_2]^2.$$

Finally we compute

```
matrix t=NF(alpha,std(pre));
```

which yields

$$t = \begin{pmatrix} -2y_1 \\ y_2 \end{pmatrix}$$

and

```
subst(t,y(1),h[1],y(2),h);
```

gives us

$$\alpha^* = \psi(t) = \begin{pmatrix} -2x_1x_2 \\ x_3^2 \end{pmatrix} \in (\alpha + \mathcal{P}) \cap K[h_1, h_2]^2.$$

Thus V is controlled and conditioned invariant and the whole set of admissible output feedbacks is given by

$$(\alpha + \mathcal{P}) \cap K[h_1, h_2]^2 = \alpha^* + \mathcal{P} \cap K[h_1, h_2]^2 = \begin{pmatrix} -2x_1x_2 \\ x_3^2 \end{pmatrix} + (x_1^2x_2^2 - x_3^2)K[h_1, h_2]^2.$$

10.4.3 Rational Output Feedback

After the considerations in Sect. 10.3.2 it seems natural to investigate controlled and conditioned invariance in the rational case. In the last part of this chapter, we will again use the notations from Sect. 10.3.2 and restrict to the MISO case (multi input, single output), i.e. $p = 1$, thus $S = K[h]$ for a polynomial $h \in R \setminus K$ and $T = K[y]$ for a variable y. Let $K(h)$ be the subfield of $Q = \mathrm{Quot}(R)$ generated by h, which means $K(h) = \mathrm{Quot}(S)$. Further we assume without loss of generality that in control system (10.1), the given matrix $g \in R^{n\times m}$ has full column rank over Q, i.e. $\ker_Q(g\cdot) = 0$. Otherwise, let $r := \mathrm{rank}_Q(g)$. We may rearrange the columns of g such that $g = (g_1, g_2)$, where $g_1 \in R^{n\times r}$ and $\mathrm{rank}_Q(g_1) = r$. Then $g_2 = g_1V$ for a matrix V with entries in Q and thus

$$gu = (g_1, g_2)\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = g_1u_1 + g_2u_2 = g_1(\underbrace{u_1 + Vu_2}_{=:\tilde{u}}).$$

Instead of (10.1) we may consider the system

$$\dot{x}(t) = f(x(t)) + g_1(x(t))\tilde{u}(t), \quad y(t) = h(x(t)),$$

where g_1 has full column rank.

Suppose now that the set $\mathcal{A}$ of rational admissible feedback laws is not empty. Using this assumption and Lemma 4, we may write $\mathcal{A}$ as

$$\mathcal{A} = \alpha + Y\mathcal{N},$$

where $\alpha \in \mathcal{A}$, $\mathcal{N} \subseteq R^q$ is an R-module and $Y \in Q^{m\times q}$. Similarly as in the polynomial case, we wish to determine the set

$$\mathcal{O} = \mathcal{A} \cap K(h)^m = (\alpha + Y\mathcal{N}) \cap K(h)^m = \alpha^* + (Y\mathcal{N} \cap K(h)^m),$$

where α^* is one particular element of $\mathcal{O}$.

For this, let us first state some properties of the subalgebra $K[h]$:

Lemma 6 *(a) $K[h]$ is a PID.*
(b) If $p, q \in K[y] \setminus \{0\}$ with $\deg(p) < \deg(q)$ then $\deg(p \circ h) < \deg(q \circ h)$.
(c) If $pq \in K[h]$, where $p \in R, 0 \neq q \in K[h]$, then $p \in K[h]$.

Proof 1. The map

$$K[y] \to K[h], \quad p(y) \mapsto p(h)$$

is a ring isomorphism, so $K[y] \cong K[h]$ and since $K[y]$ is a PID, $K[h]$ is as well.
2. Let $m = \deg(p), n = \deg(q)$ and $a_1, \ldots, a_m, b_1, \ldots, b_n \in K, a_m \neq 0, b_n \neq 0$ with $p = \sum_{i=0}^{m} a_i y^i, q = \sum_{i=0}^{n} b_i y^i$. Then

$$\deg(p \circ h) = \deg\left(\sum_{i=0}^{m} a_i h^i\right) = m \cdot \deg(h) < n \cdot \deg(h) = \deg\left(\sum_{i=0}^{n} b_i h^i\right) = \deg(q \circ h).$$

3. Let $r \in K[h]$ satisfy $r = pq$. Then there are $\tilde{q}, \tilde{r} \in K[y]$ with $q = \tilde{q}(h)$ and $r = \tilde{r}(h)$. We wish to show that $\tilde{q}l = \tilde{r}$ for some $l \in K[y]$, which yields

$$pq = r = \tilde{r}(h) = \tilde{q}(h)l(h) = ql(h)$$

and thus $p = l(h) \in K[h]$. Since $K[y]$ is a Euclidean domain, we may write $\tilde{r} = \tilde{q}l + s$ for some $l, s \in K[y]$ with $s = 0$ (which shows the assertion) or $\deg(s) < \deg(\tilde{q})$. We have

$$r = \tilde{r}(h) = \tilde{q}(h)l(h) + s(h) = ql(h) + s(h).$$

Now $q|r$ implies $q|s(h)$ and thus

$$\deg(\tilde{q} \circ h) = \deg(\tilde{q}(h)) = \deg(q) \leqslant \deg(s(h)) = \deg(s \circ h),$$

which contradicts part 2. ■

Let us first take care of the case, where $\mathcal{N}$ is an ideal of R, i.e. $q = 1$. Then $\mathcal{P} := Y\mathcal{N}$ is a fractional ideal of R and we have

$$\mathcal{P} = \langle \frac{b_1}{d_1}, \ldots, \frac{b_k}{d_k} \rangle_R = \langle \frac{c_1}{d}, \ldots, \frac{c_k}{d} \rangle_R = \frac{1}{d} \cdot \langle c_1, \ldots, c_k \rangle_R \subseteq Q,$$

for some $b_i, d_i \in R \setminus \{0\}, i = 1, \ldots, k$ and $d = \operatorname{lcm}(d_1, \ldots, d_k) \neq 0, c_i = \frac{d}{d_i} \cdot b_i$.

In a first step we wish to compute generators for $\mathcal{P} \cap K(h)$. In order to do this, consider the following construction: Let $e \in R$ be a divisor of d. Since $K[h]$ is a

principal ideal domain, the ideal $\langle e\rangle_R \cap K[h]$ is generated by one element $s_e \in K[h]$. Now $s_e \in \langle e\rangle_R$, so there is $r_e \in R$ with $s_e = r_e e$. Further we define $\mathcal{C}_e$ as the ideal quotient of $\langle c_1, \ldots, c_k\rangle_R$ by $\langle \frac{d}{e}\rangle_R$. Summarized:

$$\langle e\rangle_R \cap K[h] = \langle s_e\rangle_{K[h]} = e \cdot \langle r_e\rangle_{K[h]} \text{ and } \mathcal{C}_e := \langle c_1, \ldots, c_k\rangle_R : \langle \frac{d}{e}\rangle_R.$$

Again, the ideal $(r_e\mathcal{C}_e) \cap K[h]$ is generated by one element $k_e \in K[h]$ and $k_e = r_e l_e$ for some $l_e \in R$, i.e.

$$(r_e\mathcal{C}_e) \cap K[h] = \langle k_e\rangle_{K[h]} = r_e\langle l_e\rangle_{K[h]}.$$

Finally let

$$\mathcal{L}_e := \begin{cases} 0 & \text{if } r_e = 0 \\ \langle l_e\rangle_{K[h]} & \text{if } r_e \neq 0. \end{cases}$$

This construction yields the following result:

Lemma 7 *We have*

$$\mathcal{P} \cap K(h) = \sum_{e|d} \left\{ \frac{p}{e} \mid p \in \mathcal{L}_e \right\} = \langle \frac{l_e}{e} \mid e|d \rangle_{K[h]}.$$

Proof The second equality is clear, we just prove the first one.

"$\supseteq$": Since $\mathcal{P} \cap K(h)$ is closed under addition, it suffices to show $\{\frac{p}{e} \mid p \in \mathcal{L}_e\} \subseteq \mathcal{P} \cap K(h)$ for every divisor e of d. Let $e \in R$ with $e|d$ and $0 \neq p \in \mathcal{L}_e$ (if $\mathcal{L}_e = 0$ the assertion is clear). Then $r_e \neq 0$ and $r_e\mathcal{L}_e \subseteq r_e\mathcal{C}_e$ is equivalent to $\mathcal{L}_e \subseteq \mathcal{C}_e$, since R is a domain. Thus $p \in \mathcal{C}_e$, which implies

$$\frac{p}{e} = \frac{\frac{d}{e} \cdot p}{\frac{d}{e} \cdot e} = \frac{\frac{d}{e} \cdot p}{d} \in \langle \frac{c_1}{d}, \ldots, \frac{c_k}{d}\rangle_R = \mathcal{P}.$$

Now $r_e p \in r_e\mathcal{L}_e \subseteq K[h]$ and $r_e e = s_e \in K[h]$ yield

$$\frac{p}{e} = \frac{r_e p}{r_e e} \in K(h).$$

"$\subseteq$": Let $0 \neq \frac{1}{d}\sum_{i=1}^{k} b_i c_i \in \mathcal{P} \cap K(h)$, $b_i \in R$. Then there are $n, r \in R \setminus \{0\}$ satisfying

$$\frac{r}{n}\sum_{i=1}^{k} b_i c_i \in K[h] \text{ and } \frac{r}{n} \cdot d \in K[h].$$

Let $q = \gcd(n, r), \tilde{n}, \tilde{r} \in R \setminus \{0\}$ with $r = q\tilde{r}, n = q\tilde{n}$. This yields

$$\frac{\tilde{r}}{\tilde{n}} \sum_{i=1}^{k} b_i c_i = \frac{r}{n} \sum_{i=1}^{k} b_i c_i \in K[h] \text{ and } \frac{\tilde{r}}{\tilde{n}} \cdot d = \frac{r}{n} \cdot d \in K[h].$$

Thus $\tilde{n} | \sum_{i=1}^{k} b_i c_i$ and $\tilde{n} | d$ and we may define

$$e := \frac{d}{\tilde{n}} \in R \setminus \{0\} \text{ and } p := \frac{1}{\tilde{n}} \sum_{i=1}^{k} b_i c_i \in R \setminus \{0\}.$$

Obviously $e | d$ and

$$\frac{1}{d} \sum_{i=1}^{k} b_i c_i = \frac{1}{\frac{\tilde{r}}{\tilde{n}} \cdot d} \cdot \frac{\tilde{r}}{\tilde{n}} \sum_{i=1}^{k} b_i c_i = \frac{\tilde{r} p}{\tilde{r} e} = \frac{p}{e}.$$

It remains to show $p \in \mathcal{L}_e$. We have $\frac{d}{e} \cdot p = \tilde{n} p \in \langle c_1, \ldots, c_k \rangle_R$, so $p \in \mathcal{C}_e$ and

$$0 \neq \tilde{r} e = \frac{\tilde{r}}{\tilde{n}} \cdot \tilde{n} e = \frac{\tilde{r}}{\tilde{n}} \cdot d \in \langle e \rangle_R \cap K[h] = e \langle r_e \rangle_{K[h]}$$

implies $r_e \neq 0$ and $\tilde{r} = a r_e$ for some $a \in K[h] \setminus \{0\}$. Further $a r_e p = \tilde{r} p \in K[h]$. Using Lemma 6, part (c), we may conclude $r_e p \in K[h]$ and thus $r_e p \in (r_e \mathcal{C}_e) \cap K[h] = r_e \mathcal{L}_e$, which shows $p \in \mathcal{L}_e$. ■

Theorem 7 *In the situation above, assume $\mathcal{P} \cap K(h) \neq 0$ and let*

$$D := \{e \in R \mid e | d \text{ and } \mathcal{L}_e \neq 0\}.$$

By Lemma 7, $D \neq \emptyset$ and we may define $E := \operatorname{lcm}(e, e \in D) \neq 0$. Then we have

$$\mathcal{P} \cap K(h) = \frac{1}{E} \cdot \mathcal{L}_E = \langle \frac{l_E}{E} \rangle_{K[h]}.$$

In particular, $\mathcal{P} \cap K(h)$ is a principal ideal over $K[h]$.

Proof First note that, by the definition of the least common multiple, we have $E | d$. Then the inclusion "$\supseteq$" follows from Lemma 7, so it is enough to prove "$\subseteq$". We start by showing $\langle E \rangle_R \cap K[h] \neq 0$. There exists $0 \neq r \in R$ with $\prod_{e \in D} e = rE$ and for each $e \in D$ there is $0 \neq r_e$ satisfying $r_e e \in K[h]$ (otherwise $\mathcal{L}_e$ would be zero and thus $e \notin D$). Then $\rho := \prod_{e \in D} r_e \neq 0$ fulfills

$$\rho r E = \rho \prod_{e \in D} e = \prod_{e \in D} r_e e \in K[h] \setminus \{0\},$$

which implies

$$E\langle r_E \rangle_{K[h]} = \langle E \rangle_R \cap K[h] \supseteq \langle \rho r E \rangle_{K[h]} \neq 0.$$

We conclude $r_E \neq 0$. Now let $0 \neq q \in \mathcal{P} \cap K(h) \overset{\text{Lem.7}}{=} \langle \frac{l_e}{e} \mid e \in D \rangle_{K[h]}$. We write

$$q = \sum_{e \in D} a_e \cdot \frac{l_e}{e} = \sum_{e \in D} \frac{a_e \cdot \frac{E}{e} \cdot l_e}{E},\ l_e \in \mathcal{L}_e \subseteq \mathcal{C}_e$$

for some $a_e \in K[h]$. If $p := \sum_{e \in D} a_e \cdot \frac{E}{e} \cdot l_e \in \mathcal{L}_E$, then $q = \frac{p}{E} \in \frac{1}{E} \cdot \mathcal{L}_E$ and we are done.

Since $\frac{d}{e} \cdot l_e \in \langle c_1, \ldots, c_k \rangle_R$ for all $e \in D$, we have

$$\frac{d}{E} \cdot p = \frac{d}{E} \cdot \sum_{e \in D} a_e \cdot \frac{E}{e} \cdot l_e = \sum_{e \in D} a_e \cdot \frac{d}{e} \cdot l_e \in \langle c_1, \ldots, c_k \rangle_R,$$

thus $p \in \mathcal{C}_E$. Now $0 \neq q \in K(h)$ implies the existence of $z, n \in K[h] \setminus \{0\}$ with

$$\frac{z}{n} = q = \frac{p}{E} = \frac{r_E p}{r_E E}.$$

Using Lemma 6.3 and $n r_E p = z r_E E \in K[h]$, we conclude $r_E p \in K[h]$. This shows $r_E p \in (r_E \mathcal{C}_E) \cap K[h] = r_E \mathcal{L}_E$ and finally $p \in \mathcal{L}_E$. ■

Example 6 For $\mathcal{P} = \frac{1}{d} \cdot I$, where $d = x^2yz^3$, $I = \langle y, z \rangle_R$ and $h = xz^2$, we look for the intersection $\mathcal{P} \cap K(h)$.

```
ring R=0,(x,y,z),dp;
ideal I=y,z;
poly d=x2yz3;
ideal h=xz2;
```

Using Lemma 7, we consider some divisors of d. At first let $e_1 = y$. We use the library *ncdecomp.lib* of SINGULAR:

```
LIB "ncdecomp.lib";
poly e1=y;
IntersectWithSub(e1,h);
```

This yields $\langle e_1\rangle_R \cap K[h] = \langle y\rangle_R \cap K[h] = 0$ and thus $\mathcal{L}_{e_1} = 0$. Now we take care of $e_2 = x^2$:

```
poly e2=x2;
IntersectWithSub(e2,h);
ideal Ce2=quotient(I,d/e2);
```

SINGULAR gives $\mathcal{C}_{e_2} = \langle 1\rangle_R$ and $\langle e_2\rangle_R \cap K[h] = \langle x^2z^4\rangle_{K[h]} = x^2\langle z^4\rangle_{K[h]}$, thus we put $r_{e_2} = z^4$ and compute

```
IntersectWithSub(z4,h);
```

to get $(r_{e_2}\mathcal{C}_{e_2}) \cap K[h] = \langle x^2z^4\rangle_{K[h]} = z^4\langle x^2\rangle_{K[h]}$, which finally yields $\mathcal{L}_{e_2} = \langle x^2\rangle_{K[h]}$. Analogous computations give $\mathcal{L}_{e_3} = \langle z^3\rangle_{K[h]}$. In particular, this shows $\mathcal{P} \cap K(h) \neq 0$, so we may use Theorem 7 and set $E = \operatorname{lcm}(e_2, e_3) = x^2z^3$. Going through the same procedure as with e_2, e_3, we get the final result

$$\mathcal{P} \cap K(h) = \frac{1}{E} \cdot \mathcal{L}_E = \frac{1}{x^2z^3} \cdot \langle xz\rangle_{K[h]} = \frac{1}{xz^2} \cdot K[h].$$

Now we come back to the more general setting, where $\mathcal{N}$ is an R-submodule of R^q. Again, we may write

$$Y\mathcal{N} = \frac{1}{d} \cdot \langle c_1, \ldots, c_k\rangle_R \subseteq Q^m,$$

for some $c_i \in R^m$ and $d \in R$. Let $P = (c_1, \ldots, c_k) \in R^{m\times k}$. The next lemma describes a way to compute $(Y\mathcal{N}) \cap K(h)^m = (\frac{1}{d} \cdot \operatorname{im}_R(P\cdot)) \cap K(h)^m$.

Lemma 8 *For*

$$P = \begin{pmatrix} p_{11} & \ldots & p_{1k} \\ \vdots & & \vdots \\ p_{m1} & \ldots & p_{mk} \end{pmatrix} \in R^{m\times k},$$

consider

$$\mathcal{P}_i := \langle p_{i1}, \ldots, p_{ik}\rangle_R \text{ and } \left(\frac{1}{d} \cdot \mathcal{P}_i\right) \cap K(h) = \langle \frac{l_i}{e_i}\rangle_{K[h]},$$

where the elements e_i are chosen such that e_i divides d. Further define

$$t_i := d \cdot \frac{l_i}{e_i} \in R \text{ and } T := \begin{pmatrix} t_1 & & \\ & \ddots & \\ & & t_m \end{pmatrix} \in R^{m \times m}.$$

Then we have

$$(Y\mathcal{N}) \cap K(h)^m = \left(\frac{1}{d} \cdot \mathrm{im}_R(P\cdot)\right) \cap K(h)^m = \frac{1}{d} \cdot T \cdot (\pi(\ker_R((T, P)\cdot)) \cap K[h]^m),$$

where π denotes the projection onto the first m components.

Proof The first equality is clear, we will prove the second.

"$\subseteq$": If $\frac{1}{d} \cdot Py \in (\frac{1}{d} \cdot \mathrm{im}_R(P\cdot)) \cap K(h)^m$, we have for $i = 1, \ldots, m$

$$\frac{1}{d} \cdot (Py)_i = \left(\frac{1}{d} \cdot Py\right)_i = \frac{1}{d} \cdot \left(p_{i1}, \ldots, p_{ik}\right) \cdot y \in \left(\frac{1}{d} \cdot \mathcal{P}_i\right) \cap K(h).$$

Thus, we can write

$$(Py)_i = d \cdot \frac{l_i}{e_i} \cdot a_i = t_i a_i,$$

for an element $a_i \in K[h]$. Stacking the a_i in a vector $a = (a_1, \ldots, a_m)^{tr} \in K[h]^m$ yields

$$Py = \begin{pmatrix} t_1 a_1 \\ \vdots \\ t_m a_m \end{pmatrix} = \begin{pmatrix} t_1 & & \\ & \ddots & \\ & & t_k \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = Ta \in \mathrm{im}_R(P\cdot),$$

and thus $a \in \pi(\ker_R((T, P)\cdot)) \cap K[h]^m$. This finally shows

$$\frac{1}{d} \cdot Py = \frac{1}{d} \cdot Ta \in \frac{1}{d} \cdot T \cdot (\pi(\ker_R((T, P)\cdot)) \cap K[h]^m),$$

which implies "$\subseteq$".

"$\supseteq$": Let $a = (a_1, \ldots, a_m)^{tr} \in \pi(\ker_R((T, P)\cdot)) \cap K[h]^m, a_i \in K[h]$. Then $Ta \in \mathrm{im}_R(P\cdot)$ and thus $\frac{1}{d} \cdot Ta \in \frac{1}{d} \cdot \mathrm{im}_R(P\cdot)$. Further we have

$$t_i a_i = d \cdot \frac{l_i}{e_i} \cdot a_i \in dK(h),$$

which yields

$$\frac{1}{d} \cdot Ta = \frac{1}{d} \cdot \begin{pmatrix} t_1 a_1 \\ \vdots \\ t_m a_m \end{pmatrix} \in K(h)^m.$$

Putting things together, we showed that $\frac{1}{d} \cdot Ta \in (\frac{1}{d} \cdot \mathrm{im}_R(P\cdot)) \cap K(h)^m$. ■

For a characterisation of controlled and conditioned invariance in the rational feedback case, let us shortly recapitulate some definitions made earlier in this chapter: Let the R-module $\mathcal{M} = \mathrm{im}_R(M\cdot) \subseteq R^n$ contain all admissible vector fields and $\tilde{g} \in R^{n\times r}$ be such that $\ker_R(k\cdot) = \mathrm{im}_R(\tilde{g})$, where k satisfies $\mathrm{im}_R(\cdot k) = \ker_R(\cdot g)$. Then there is $Y \in Q^{m\times r}$ with $\tilde{g} = gY$ (see Lemma 2) and we defined $\mathcal{N} = \pi(\ker_R((M, -\tilde{g})\cdot))$. We wish to find $\alpha^* \in \mathcal{O} = (Y\alpha + Y\mathcal{N}) \cap K(h)^m$, where $\alpha \in R^r$ is such that $\bar{\alpha} = Y\alpha \in Q^m$ is admissible for (10.1).

Example 7 Let $\bar{\alpha} = \frac{y+rx^2z^2}{x^2yz^4}$ with $r \in R \setminus \langle y\rangle_R$ arbitrary, $d = x^2yz^3$, $h = xz^2$ and $Y\mathcal{N} = \frac{1}{d} \cdot \langle y, z\rangle_R$ (see Example 6, $Y\mathcal{N} = \mathcal{P}$). We claim $\bar{\alpha} \notin K(h)$: If $\bar{\alpha} = \frac{p}{q}$ for some $p, q \in K[h]$, then $rx^2z^2q = y(pz - q)$. Since $y \nmid r$, we must have $y \mid q$ and thus $q \notin K[h]$, which is a contradiction. But we have

$$\bar{\alpha} - \underbrace{\frac{z}{x^2yz^3} \cdot rx^2}_{\in Y\mathcal{N}} = \frac{y + rx^2z^2 - rx^2z^2}{x^2yz^4} = \frac{1}{x^2z^4} = \frac{1}{h^2} \in K(h) \cap (\alpha + Y\mathcal{N}).$$

Thus, we may choose $\alpha^* = \frac{1}{x^2z^4}$ to get

$$\mathcal{O} = (\bar{\alpha} + Y\mathcal{N}) \cap K(h) = \alpha^* + Y\mathcal{N} \cap K(h) \overset{Ex.6}{=} \frac{1}{x^2z^4} + \frac{1}{xz^2} \cdot K[h].$$

The following notations and constructions will be useful to find α^* in any situation:

(a) Let $\mathcal{N} = \langle n_1, \ldots, n_k\rangle_R \subseteq R^r$.
(b) We may write $Y = \frac{1}{d} \cdot U$ for some $d \in R,\ U \in R^{m\times r}$.
(c) Use Lemma 8 to compute

$$(\langle Y\alpha\rangle_R + Y\mathcal{N}) \cap K(h)^m = \left(\frac{1}{d} \cdot (\langle U\alpha\rangle_R + U\mathcal{N})\right) \cap K(h)^m = \frac{1}{e} \cdot \langle l_1, \ldots, l_t\rangle_{K[h]},$$

where $l_i \in R^m$, $e \in R$ with $e|d$. Further define

$$S_\alpha := \left(U\alpha, Un_1, \ldots, Un_k, \frac{d}{e} \cdot l_1, \ldots, \frac{d}{e} \cdot l_t \right)$$

and $\mathcal{S}_\alpha := \ker_R(S_\alpha \cdot)$.

Theorem 8 *The following statements are equivalent:*

(a) There exists $\tilde{\alpha} \in K(h)^m$ *with* $f + g\tilde{\alpha} \in \mathcal{M}$.
(b) There is $\alpha \in R^r$ *with* $f + \tilde{g}\alpha \in \mathcal{M}$ *and* $\emptyset \neq (Y\alpha + Y\mathcal{N}) \cap K(h)^m$.
(c) There is $\alpha \in R^r$ *with* $f + \tilde{g}\alpha \in \mathcal{M}$ *and* $a \in R^k, b \in K[h]^t$ *such that* $\begin{pmatrix} 1 \\ a \\ b \end{pmatrix} \in \mathcal{S}_\alpha$.

Proof 1. $\Rightarrow$ 2. : Let $\tilde{\alpha} \in K(h)^m$ with $f + g\tilde{\alpha} \in \mathcal{M}$. Using Lemma 3, there is $\alpha \in R^r$ satisfying $f + \tilde{g}\alpha \in \mathcal{M}$. Then we have

$$\mathcal{M} \ni f + g\tilde{\alpha} - f - \tilde{g}\alpha = g\tilde{\alpha} - gY\alpha = g(\tilde{\alpha} - Y\alpha),$$

and since $\ker_Q(g\cdot)$ is assumed to be zero, Lemma 4 implies

$$\tilde{\alpha} - Y\alpha \in \{\bar{\alpha} \in Q^m \mid g\bar{\alpha} \in \mathcal{M}\} = Y\mathcal{N}.$$

This shows $\tilde{\alpha} \in (Y\alpha + Y\mathcal{N}) \cap K(h)^m$.
2. $\Rightarrow$ 3. : If

$$\begin{aligned} \tilde{\alpha} \in (Y\alpha + Y\mathcal{N}) \cap K(h)^m &= \left(\frac{1}{d} \cdot U\alpha + \frac{1}{d} \cdot U\mathcal{N} \right) \cap K(h)^m \\ &\subseteq \left(\frac{1}{d} (\langle U\alpha \rangle_R + U\mathcal{N}) \right) \cap K(h)^m = \frac{1}{e} \cdot \langle l_1, \ldots, l_t \rangle_{K[h]}, \end{aligned}$$

then there are $a_i \in R, b_i \in K[h]$ with

$$\tilde{\alpha} = \frac{1}{d} \cdot U\alpha + \frac{1}{d} \cdot U \sum_{i=1}^{k} a_i n_i = \frac{1}{e} \sum_{i=1}^{t} b_i l_i.$$

Setting $a := (a_1, \ldots, a_k), b := (b_1, \ldots, b_t)$, this is equivalent to

$$0 = U\alpha + U \sum_{i=1}^{k} a_i n_i - \frac{d}{e} \sum_{i=1}^{t} b_i l_i = S_\alpha \begin{pmatrix} 1 \\ a \\ b \end{pmatrix},$$

and thus $\begin{pmatrix} 1 \\ a \\ b \end{pmatrix} \in \ker_R(S_\alpha \cdot) = \mathcal{S}_\alpha$.

3. $\Rightarrow$ 1. : Let $\alpha \in R^r$ with $f + \tilde{g}\alpha \in \mathcal{M}$. For $a \in R^k$, $b_i \in K[h]^t$ such that

$$0 = S_\alpha \begin{pmatrix} 1 \\ a \\ b \end{pmatrix} = U\alpha + U \sum_{i=1}^{k} a_i n_i + \frac{d}{e} \sum_{i=1}^{t} b_i l_i, \tag{10.19}$$

we define

$$\tilde{\alpha} := \frac{1}{d} \cdot U\alpha + \frac{1}{d} \cdot U \sum_{i=1}^{k} a_i n_i.$$

Then (10.19) implies

$$K(h)^m \ni -\frac{1}{e} \sum_{i=1}^{t} b_i l_i = \tilde{\alpha} = \frac{1}{d} \cdot U\alpha + \frac{1}{d} \cdot U \sum_{i=1}^{k} a_i n_i \in Y\alpha + Y\mathcal{N}.$$

It remains to show $f + g\tilde{\alpha} \in \mathcal{M}$. Since $\sum_{i=1}^{k} a_i n_i \in \mathcal{N} = \pi(\ker_R((M, -\tilde{g})\cdot))$ there is x with entries in R such that

$$0 = (M, -\tilde{g}) \begin{pmatrix} x \\ \sum_{i=1}^{k} a_i n_i \end{pmatrix} = Mx - \tilde{g} \sum_{i=1}^{k} a_i n_i,$$

and thus $\tilde{g} \sum_{i=1}^{k} a_i n_i \in \mathcal{M}$. This yields

$$\begin{aligned} f + g\tilde{\alpha} &= f + g \cdot \frac{1}{d} \cdot U\alpha + g \cdot \frac{1}{d} \cdot U \sum_{i=1}^{k} a_i n_i = f + gY\alpha + gY \sum_{i=1}^{k} a_i n_i \\ &= \underbrace{f + \tilde{g}\alpha}_{\in \mathcal{M}} + \underbrace{gY \sum_{i=1}^{k} a_i n_i}_{\in \mathcal{M}} \in \mathcal{M}, \end{aligned}$$

which finishes the proof. ■

Finally, we describe a way how to find elements in $\mathcal{S}_\alpha$ satisfying condition 3. in Theorem 8 (if this is possible at all).

Consider an R-module

$$\mathcal{S} = \mathrm{im}_R \left(\begin{pmatrix} q \\ A \\ B \end{pmatrix} \cdot \right),$$

where $q \in R^{1\times k}$, $A \in R^{l\times k}$, $B \in R^{m\times k}$. We wish to find the set

$$\mathcal{T} := \left\{ \begin{pmatrix} 1 \\ a \\ b \end{pmatrix} \in \mathcal{S} \mid a \in R^l, b \in K[h] \right\}.$$

We use the following construction to do the task:

(a) If $\{u \in R^k \mid qu = 1\} = \emptyset$, then $\mathcal{T} = \emptyset$ (see Lemma 9 below).
Otherwise, let $u^* \in R^k$ with $qu^* = 1$ and $C \in R^{k\times n}$ satisfy $\mathrm{im}_R(C\cdot) = \ker_R(q\cdot)$.
Then

$$\{u \in R^k \mid qu = 1\} = u^* + \mathrm{im}_R(C\cdot). \tag{10.20}$$

Proof If $u = u^* + Cx$, then $qu = qu^* + qCx = 1 + 0 = 1$.
For the other inclusion let $u \in R^k$ with $qu = 1 = qu^*$. Then $q(u - u^*) = 0$ implies $u - u^* \in \ker_R(q\cdot) = \mathrm{im}_R(C\cdot)$ and thus $q \in u^* + \mathrm{im}_R(C\cdot)$. ■

(b) Let

$$\mathcal{O} := (Bu^* + \mathrm{im}_R(BC\cdot)) \cap K[h]^m.$$

If $\mathcal{O} = \emptyset$, then $\mathcal{T} = \emptyset$ (see Lemma 9 below). Otherwise let $b^* = Bu^* + BCx^* \in \mathcal{O}$ for some $x^* \in R^n$. Then we have

$$\mathcal{O} = b^* + (\mathrm{im}_R(BC\cdot) \cap K[h]^m) = b^* + \mathrm{im}_{K[h]}(H\cdot), \tag{10.21}$$

for some $H \in K[h]^{m\times r}$. There exists $D \in R^{n\times r}$ with $H = BCD$.

(c) Let $E \in R^{n\times s}$ with $\mathrm{im}_R(E\cdot) = \ker_R(BC\cdot)$.

Lemma 9 *The following statements are equivalent:*

(a) $\mathcal{T} \neq \emptyset$.
(b) There exist $u^* \in R^k$ *with* $qu^* = 1$ *and* $b^* \in \mathcal{O}$ *and*

$$\mathcal{T} = \begin{pmatrix} 1 \\ Au^* + ACx^* \\ b^* \end{pmatrix} + \mathrm{im}_{K[h]} \left(\begin{pmatrix} 0 \\ ACD \\ H \end{pmatrix} \cdot \right) + \mathrm{im}_R \left(\begin{pmatrix} 0 \\ ACE \\ 0 \end{pmatrix} \cdot \right).$$

Proof 2. $\Rightarrow$ 1.: We have

$$\mathcal{S} \ni \begin{pmatrix} q \\ A \\ B \end{pmatrix} (u^* + Cx^*) = \begin{pmatrix} qu^* + qCx^* \\ Au^* + ACx^* \\ Bu^* + BCx^* \end{pmatrix} = \begin{pmatrix} 1 \\ Au^* + ACx^* \\ b^* \end{pmatrix} \in \mathcal{T}.$$

1. $\Rightarrow$ 2.: Let $\begin{pmatrix} 1 \\ a \\ b \end{pmatrix} = \begin{pmatrix} q \\ A \\ B \end{pmatrix} x \in \mathcal{T}$. Then $qx = 1$, so u^* exists and according to (10.20)

$$\emptyset \neq \{u \in R^k \mid qu = 1\} = u^* + \mathrm{im}_R(C\cdot) \ni x,$$

i.e. $x = u^* + Cy$ for some $y \in R^n$. Now $b \in K[h]$ implies

$$b = B(u^* + Cy) = Bu^* + BCy \in K[h] \cap (Bu^* + \mathrm{im}_R(BC\cdot)) = \mathcal{O} \neq \emptyset,$$

and thus $b^* = Bu^* + BCx^* \in \mathcal{O}$ exists, where $x^* \in R^n$. From (10.21) above we conclude $b \in \mathcal{O} = b^* + \mathrm{im}_{K[h]}(H\cdot)$, so there is $z \in K[h]^r$ with

$$Bx = Bu^* + BCy = b = b^* + Hz = Bu^* + BCx^* + BCDz.$$

This yields $0 = BC(y - x^* + Dz)$, so

$$y \in x^* + \ker_R(BC\cdot) + \mathrm{im}_{K[h]}(D\cdot) = x^* + \mathrm{im}_R(E\cdot) + \mathrm{im}_{K[h]}(D\cdot).$$

We may write $y = x^* + Ew - Dz$ for some $w \in R^s$. Finally, we obtain

$$\begin{pmatrix} 1 \\ a \\ b \end{pmatrix} = \begin{pmatrix} q \\ A \\ B \end{pmatrix} x = \begin{pmatrix} q \\ A \\ B \end{pmatrix} (u^* + Cy) = \begin{pmatrix} qu^* + qC(x^* + Ew - Dz) \\ Au^* + ACx^* + ACEw - ACDz \\ Bu^* + BCx^* + BCEw - BCDz \end{pmatrix}$$
$$= \begin{pmatrix} 1 \\ Au^* + ACx^* \\ b^* \end{pmatrix} - \begin{pmatrix} 0 \\ ACD \\ H \end{pmatrix} z + \begin{pmatrix} 0 \\ ACE \\ 0 \end{pmatrix} w.$$

This shows "$\subseteq$".

For the other inclusion let $x \in K[h]^r$ and $y \in R^s$. We have

$$\begin{pmatrix} 1 \\ Au^* + ACx^* \\ b^* \end{pmatrix} + \begin{pmatrix} 0 \\ ACD \\ H \end{pmatrix} x + \begin{pmatrix} 0 \\ ACE \\ 0 \end{pmatrix} y = \begin{pmatrix} 1 \\ Au^* + ACx^* + ACDx + ACEy \\ b^* + Hx \end{pmatrix}$$
$$= \begin{pmatrix} q \\ A \\ B \end{pmatrix} \cdot (u^* + Cx^* + CDx + CEy) \in \mathcal{S},$$

and thus, since $b^* + Hx \in K[h]$, the term above is also in $\mathcal{T}$. ■

Concluding Remarks

Given a variety V and a polynomial control system, the methods in this chapter allow us to decide whether V is controlled invariant for the system with polynomial and rational feedback laws. The controlled and conditioned invariance is fully characterised in the polynomial feedback case, but only for single output systems in the rational setting. The general case with arbitrary many outputs is, up to now, an open problem. Nevertheless, we expect the algebraic tools of intersecting an affine ideal with a subalgebra (or an affine fractional ideal with a subfield with one generator, resp.) to be useful in other algebraic disciplines as well. Further, the investigation of invariant varieties for rational vector fields is another topic for future research.

References

1. Adams, W.W., Loustaunau, P.: An Introduction to Gröbner Bases. American Mathematical Society, Providence (1994)
2. Bourbaki, N.: Commutative Algebra. Addison-Wesley/Hermann, Reading/Paris (1972)
3. Chyzak, F., Quadrat, A., Robertz, D.: Effective algorithms for parametrizing linear control systems over Ore algebras. Appl. Algebra Eng. Commun. Comput. **16**, 319–376 (2005)
4. Greuel, G.-M., Pfister, G.: A Singular Introduction to Commutative Algebra. Springer, Berlin (2002)
5. Isidori, A., Krener, A.J., Gori-Giorgi, C., Monaco, S.: Nonlinear decoupling via feedback: a differential geometric approach. IEEE Trans. Autom. Control **26**, 331–345 (1981)
6. Levandovskyy, V.: Non-commutative computer algebra for polynomial algebras: Gröbner bases, applications and implementation. Ph.D. thesis, TU Kaiserslautern (2005)
7. Matsumura, H.: Commutative Ring Theory. Cambridge University Press, Cambridge (1987)
8. Quadrat, A.: On a generalization of the Youla-Kučera parametrization. Part I: the fractional ideal approach to SISO systems. Syst. Control Lett. **50**, 135–148 (2003)
9. Quadrat, A.: A lattice approach to analysis and synthesis problems. Math. Control Signals Syst. **18**, 147–186 (2006)

10. Quadrat, A., Robertz, D.: Computation of bases of free modules over the Weyl algebras. J. Symb. Comput. **42**, 1113–1141 (2007)
11. Zerz, E., Walcher, S.: Controlled invariant hypersurfaces of polynomial control systems. Qual. Theory Dyn. Syst. **11**, 145–158 (2012)

Chapter 11
A Note on Controlled Invariance for Behavioral nD Systems

Ricardo Pereira and Paula Rocha

Abstract In this chapter we extend the notion of invariance of nD behaviors introduced in Pereira and Rocha (European Control Conference 2013, ECC'13. ETH Zurich, Switzerland, pp. 301–305, 2013) [4], Rocha and Wood (Int. J. Appl. Math. Comput. Sci. 7(4):869–879, 1997) [7] to the controlsetting. More concretely, we introduce a notion which is the behavioral counterpart of classical controlled invariance, using the framework of partial interconnections. In such interconnections, the variables are divided into two sets: the variables to-be-controlled and the variables on which it is allowed to enforce restrictions (called control variables). In particular we focus on regular partial interconnection, i.e., interconnections in which the restrictions of the controller do not overlap with the ones already implied by the laws of the original behavior. For some particular cases, complete characterizations of controlled invariance and controller construction procedures are derived for both 1D and nD behaviors.

Keywords Autonomous behavior · Controllable · Controlled-invariant · Implementable · Invariant · Minimal left annihilator · Partial interconnection

11.1 Introduction

In this chapter we deal with the concept of controlled invariance. Similar to what happens for state space systems, controlled invariance means "invariance after control". In the behavioral approach, control is nothing but interconnecting (intersecting) a given behavior with a suitable controller behavior in order to obtain a desired controlled behavior. This can be done essentially in two ways, namely by full inter-

R. Pereira (✉)
CIDMA, Department of Mathematics, University of Aveiro, Aveiro, Portugal
e-mail: ricardopereira@ua.pt

P. Rocha
SYSTEC and Faculty of Engineering, University of Porto, Porto, Portugal
e-mail: mprocha@fe.up.pt

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5_11

connection (where all the system variables are available for control) or by partial interconnection (where the variables are divided into *to-be-controlled variables* and *control variables*, [2, 12]). Here we only consider the case of control by partial interconnection.

In particular, we are interested in *regular controllers* which are characterized by imposing restrictions on the control variables that do not overlap with the ones already implied by the laws of the original behavior.

Invariance in the behavioral context was introduced in [4]. Here we use a sightly different definition which is equivalent to the original one in the 1D case. Roughly speaking, a sub-behavior $\mathcal{V}$ of a behavior $\mathcal{B}$ is said to be $\mathcal{B}$-invariant if the freedom of the trajectories of $\mathcal{B}$ is "captured" by $\mathcal{V}$. Here we extend this notion of invariance to the control setting.

The chapter is organized as follows: in Sect. 11.2 we introduce the relevant preliminaries on behaviors, and the problem of control by partial interconnection is addressed in Sect. 11.3. Finally, in Sect. 11.4, we define and characterize the controlled invariance of a given behavior.

11.2 Preliminaries

In the behavioral approach [11] a dynamical system is defined as a quadruple $\Sigma = (\mathcal{T}, \mathcal{W}, \mathcal{U}, \mathcal{B})$, where $\mathcal{T}$ is the time axis, $\mathcal{W}$ is the signal space, and the *behavior* $\mathcal{B}$ is a subset of a *universe* $\mathcal{U} \subset \mathcal{W}^{\mathcal{T}} = \{f : \mathcal{T} \to \mathcal{W}\}$. The elements of $\mathcal{B}$ are called trajectories.

In this chapter we will consider *nD behaviors* $\mathcal{B}$ defined over the continuous nD domain $\mathbb{R}^n$ that can be described by a set of linear partial differential equations, i.e.,

$$\mathcal{B} = \ker H(\underline{\partial}) := \left\{ z \in \mathcal{U} : H(\underline{\partial}) z = 0 \right\},$$

where $\mathcal{U} = \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R}^{\mathfrak{q}})$, for some $\mathfrak{q} \in \mathbb{N}$, $\underline{\partial} = (\partial_1, \ldots, \partial_n)$, the ∂_i's are the elementary partial differential operators and $H(\underline{s})$, with $\underline{s} = (s_1, \ldots, s_n)$, is an nD polynomial matrix, i.e, $H(\underline{s})$ belongs to the set $\mathbb{R}^{\bullet \times \mathfrak{q}}[\underline{s}]$ of $\bullet \times \mathfrak{q}$ matrices with entries in the ring $\mathbb{R}[\underline{s}]$ of nD polynomials. Such matrix is known as a *(kernel) representation* of $\mathcal{B}$. We shall refer to these behaviors as *kernel behaviors* or simply as *behaviors*. For short, whenever the context is clear we omit the indeterminate $\underline{s}$ and the operator $\underline{\partial}$.

Note that different representations may give rise to the same behavior. In particular $\ker H = \ker UH$ for any unimodular nD polynomial matrix U. Moreover, $\mathcal{B}_1 = \ker H_1 \subseteq \mathcal{B}_2 = \ker H_2$ if and only if there exists an nD polynomial matrix $\bar{H}$ such that $H_2 = \bar{H} H_1$.

Instead of characterizing $\mathcal{B}$ by means of a representation matrix H, it is also possible to characterize it by means of its *orthogonal module* $\mathrm{Mod}(\mathcal{B})$, which consists of all the nD polynomial rows r such that $\mathcal{B} \subset \ker r$, and can be shown to coincide with the polynomial module generated by the rows of H, i.e., $\mathrm{Mod}(\mathcal{B}) = \mathrm{RM}(H)$, where RM stands for row module, see [13] for details.

In this paper, the notion of autonomy plays an important role. Although there are several (equivalent) ways of defining this property, here we simply define autonomy as the absence of free variables. Given a behavior $\mathcal{B}$ in the universe $\mathcal{U} = \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R}^q)$ and trajectories w with components w_i, $i \in \{1, \ldots, q\}$, w_i is said to be a *free variable* of $\mathcal{B}$ if

$$\forall w_i^* \in \mathcal{C}^\infty\left(\mathbb{R}^n, \mathbb{R}\right), \ \exists w \in \mathcal{B} \text{ s.t. } w_i = w_i^*.$$

Definition 1 An nD behavior $\mathcal{B}$ with a kernel representation is called *autonomous* if $\mathcal{B}$ has no free variables.

The next proposition provides a characterization of autonomy in terms of kernel representations. This was proven in [7, 15] for the discrete domain case, but the proofs are also valid in the case of continuous domains.

Proposition 1 *Given an nD behavior $\mathcal{B} = \ker H$, then $\mathcal{B}$ is autonomous if and only if the nD polynomial matrix H has full column rank.*

The following notion introduced in [16] is helpful to characterize the property of controllability for behaviors.

Definition 2 An nD polynomial matrix $H(\underline{s})$ is called *generalized factor left prime (GFLP)*, if the existence of a factorization $H = DH_1$ (D not necessarily square) with $rank(H) = rank(H_1)$ implies the existence of an nD polynomial matrix E such that $H_1 = EH$.

Roughly speaking, a system over $\mathbb{R}^n$ is controllable if its trajectories can be independently specified on any two open subsets of the domain $\mathbb{R}^n$ with disjoint closures [5, 8]. In the 1D case, this corresponds to the possibility of linking any known past trajectory to any desired future trajectory [11]. In [16], controllability was characterized as stated next.

Proposition 2 *Consider an nD behavior $\mathcal{B} = \ker H$. Then, $\mathcal{B}$ is controllable if and only if H is GFLP*

Another important concept introduced in [16] is the one of left-coprime factorization.

Definition 3 Two nD polynomial matrices D and N with the same number of rows are said to be *left-coprime* if the block matrix $X = \begin{bmatrix} N & -D \end{bmatrix}$ is GFLP.

Definition 4 A pair of nD polynomial matrices (D, N) is a *left-coprime factorization* of the nD rational matrix G if $DG = N$, D has full column rank and D and N are left-coprime.

Minimal left annihilators will be relevant in the sequel and are defined as follows [16].

Definition 5 Let $H \in \mathbb{R}^{g \times q}[\underline{s}]$. Then $X \in \mathbb{R}^{m \times g}[\underline{s}]$ is called a *minimal left annihilator (MLA)* of H if the following conditions hold:

(a) X is a left annihilator of H, i.e., $XH = 0$.

(b) If $X_1 H = 0$, with $X_1 \in \mathbb{R}^{p\times g}[\underline{s}]$, then $X_1 = MX$, for some nD polynomial matrix M.

It is shown in [16, Theorem 9] that.

Theorem 1 *Given an nD polynomial matrix* $M = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}$ *with* M_1 *square and non-singular,* (D, N) *is a left-coprime factorization of* $G = M_2 M_1^{-1}$ *if and only if* $X = \begin{bmatrix} N & -D \end{bmatrix}$ *is an MLA of* M.

In [3], Oberst showed that the *quotient of two behaviors* admits the structure of a behavior (see also [13, 14]). Indeed, if $\mathcal{B}$ and $\mathcal{B}'$ are behaviors such that $\mathcal{B}' \subseteq \mathcal{B}$, choosing a kernel representation H' of $\mathcal{B}'$ the following isomorphism holds:

$$\mathcal{B}/\mathcal{B}' \cong H'(\mathcal{B}).$$

The kernel representation of the quotient of two behaviors can be related with the kernel representations of the latter as stated in the following result, [8].

Proposition 3 *Let* $\mathcal{B}' \subseteq \mathcal{B}$ *be two nD behaviors, where* $\mathcal{B}' = \ker H'$ *and* $\mathcal{B} = \ker EH'$, *for some nD polynomial matrices* H' *and* E. *Let* C *be an MLA of* H', *and set*

$$L = \begin{bmatrix} E \\ C \end{bmatrix}.$$

Then $\mathcal{B}/\mathcal{B}' \cong \ker L$. *In the case where* H' *has full row rank,* $\mathcal{B}/\mathcal{B}' \cong \ker E$.

The *sum of two behaviors* $\mathcal{B}_1$ and $\mathcal{B}_2$ is defined as

$$\mathcal{B}_1 + \mathcal{B}_2 := \{z : \exists z_1 \in \mathcal{B}_1, \exists z_2 \in \mathcal{B}_2 : z = z_1 + z_2\},$$

and is clearly the smallest behavior containing both $\mathcal{B}_1$ and $\mathcal{B}_2$. A kernel representation of the sum of two behaviors was derived in [8], (see also [10]).

Proposition 4 *Let* $\mathcal{B} = \ker H$ *and* $\overline{\mathcal{B}} = \ker \overline{H}$ *be two nD behaviors and* $\begin{bmatrix} C & \overline{C} \end{bmatrix}$ *be an MLA of* $\begin{bmatrix} H \\ \overline{H} \end{bmatrix}$. *Then* $CH = -\overline{C}\,\overline{H}$ *is a kernel representation of* $\mathcal{B} + \overline{\mathcal{B}}$.

11.3 Control by Partial Interconnection

In the behavioral approach, to control a behavior we should impose suitable restrictions to its variables in order to obtain a new desired behavior. This is achieved by interconnecting (intersecting) the given behavior with another behavior called controller. Two situations can be considered, namely full interconnection (where all the system variables are available for control) and partial interconnection (where the

variables are divided into *to-be-controlled variables* and *control variables*, [2, 12]). Here we only consider the case of control by partial interconnection.

In the sequel we denote the to-be-controlled variables by w and the control variables by c. We assume that the joint behavior of these variables, i.e., the (w, c)-behavior, is given as:

$$\mathcal{B}_{(w,c)} := \left\{(w, c) \in \mathcal{U}^{\mathrm{w}} \times \mathcal{U}^{\mathrm{c}} \mid R(\underline{\partial})w = M(\underline{\partial})c\right\}, \tag{11.1}$$

where, for $\mathrm{q} \in \mathbb{N}, \mathcal{U}^{\mathrm{q}} := \mathcal{C}^{\infty}(\mathbb{R}^n, \mathbb{R}^{\mathrm{q}})$ and $R(\underline{s}) \in \mathbb{R}^{g\times \mathrm{w}}[\underline{s}]$, $M(\underline{s}) \in \mathbb{R}^{g\times \mathrm{c}}[\underline{s}]$ are nD polynomial matrices.

The w-behavior induced by $\mathcal{B}_{(w,c)}$, i.e, $\mathcal{B}_w = \pi_w\left(\mathcal{B}_{(w,c)}\right)$, where π_w denotes the projection into $\mathcal{U}^{\mathrm{w}}$, is obtained by eliminating c from the equation $R(\underline{\partial})w = M(\underline{\partial})c$, which is achieved by applying to both sides of the equation a minimal left annihilator $L(\underline{\partial})$ of $M(\underline{\partial})$. This yields $\mathcal{B}_w = \ker(LR)$.

The control action then consists in restricting the behavior of the control variables c in order to obtain a desired effect on w, this is, given a behavior to be controlled $\mathcal{B}_{(w,c)} \subset \mathcal{U}^{\mathrm{w}} \times \mathcal{U}^{\mathrm{c}}$ and a desired behavior $\mathcal{D}_w \subset \mathcal{U}^{\mathrm{w}}$, a controller behavior $\mathcal{C}_c \subset \mathcal{U}^{\mathrm{c}}$ (given by $\mathcal{C}_c = \{c \in \mathcal{U}^{\mathrm{c}} \,:\, C(\underline{\partial})c = 0\} = \ker C$, for some adequate nD polynomial matrix $C(\underline{s})$) has to be determined such that

$$\mathcal{D}_w = \pi_w\left(\mathcal{B}_{(w,c)} \cap \mathcal{C}^*_{(w,c)}\right), \tag{11.2}$$

where $C^*_{(w,c)}$ stands for the lifted behavior

$$\mathcal{C}^*_{(w,c)} := \{(w, c) \in \mathcal{U}^{\mathrm{w}} \times \mathcal{U}^{\mathrm{c}} \mid w \text{ is free and } c \in \mathcal{C}_c\}.$$

If (11.2) holds, we say that $\mathcal{D}_w$ is *implementable* by partial interconnection from $\mathcal{B}_{(w,c)}$.

Regular controllers play an important role in this context. They are characterized by imposing restrictions on the control variables that do not overlap with the ones already implied by the laws of the original behavior $\mathcal{B}_{(w,c)}$.

Partial interconnection with a regular controller is called *regular partial interconnection*. In terms of the nD polynomial matrices $R(\underline{s})$, $M(\underline{s})$ and $C(\underline{s})$ that describe the to-be-controlled behavior $\mathcal{B}_{(w,c)}$ and the controller $\mathcal{C}_c$, the regularity of the corresponding partial interconnection is equivalent to the following condition:

$$\operatorname{rank}\begin{bmatrix} R(\underline{s}) & M(\underline{s}) \\ 0 & C(\underline{s}) \end{bmatrix} = \operatorname{rank}\begin{bmatrix} R(\underline{s}) & M(\underline{s}) \end{bmatrix} + \operatorname{rank}\begin{bmatrix} 0 & C(\underline{s}) \end{bmatrix}.$$

In terms of modules, the previous equation is equivalent to

$$\operatorname{Mod}(\mathcal{B}_{(w,c)}) \cap \operatorname{Mod}(\mathcal{C}^*_{(w,c)}) = \{0\}.$$

Thus, every controller $\mathcal{C}_c = \ker C$ is regular if the nD polynomial matrix $R(\underline{s})$ has full row rank. In turn, this condition means that all the control variables are free in

the to-be-controlled behavior $\mathcal{B}_{(w,c)}$. The case where $R(\underline{s})$ is not full row rank will also be treated, but leads to more cumbersome computations.

It is not difficult to see that only sub-behaviors $\mathcal{D}_w$ of $\mathcal{B}_w$ are implementable from $\mathcal{B}_{(w,c)}$ by partial interconnection. Moreover, the smallest sub-behavior of $\mathcal{B}_w$ implementable by partial interconnection is clearly obtained by setting all the control variables to be zero. This gives rise to the behavior

$$\mathcal{N}_w := \left\{ w \in \mathcal{U}^w \mid (w, 0) \in \mathcal{B}_{(w,c)} \right\},$$

whose kernel representation is $\mathcal{N}_w = \ker R$, known as *hidden behavior*, [12]. As the following result shows, $\mathcal{N}_w$ plays an important role in the characterization of (the possibility of) implementation by partial interconnection (see [2, 6, 8, 9] for details).

Proposition 5

(a) An nD behavior $\mathcal{D}_w$ is implementable from $\mathcal{B}_{(w,c)}$ by partial interconnection if and only if $\mathcal{N}_w \subset \mathcal{D}_w \subset \mathcal{B}_w$.

(b) A 1D behavior $\mathcal{D}_w$ is implementable from $\mathcal{B}_{(w,c)}$ by regular partial interconnection if and only if

$$\mathcal{N}_w \subset \mathcal{D}_w \subset \mathcal{B}_w \text{ and } \mathcal{B}_w/\mathcal{D}_w \text{ is controllable.}$$

11.4 Behavioral Controlled-Invariance

Before introducing the notion of behavioral controlled-invariance, following [4, 7] we adopt the next definition for behavioral invariance.

Definition 6 Given an nD behavior $\mathcal{B}_w$, a sub-behavior $\mathcal{V}_w$ of $\mathcal{B}_w$ is said to be $\mathcal{B}_w$-*invariant* if the quotient behavior $\mathcal{B}_w/\mathcal{V}_w$ is autonomous.

Since autonomy is the absence of free variables, this intuitively means that all the freedom of the trajectories of $\mathcal{B}_w$ is captured by $\mathcal{V}_w$.

By Propositions 1 and 3 the following corollary is immediate.

Corollary 1 *Let $\mathcal{V}_w \subseteq \mathcal{B}_w$ be two nD behaviors, where $\mathcal{V}_w = \ker V$ and $\mathcal{B}_w = \ker EV$, for some nD polynomial matrices V and E, with V full row rank. Then $\mathcal{V}_w$ is $\mathcal{B}_w$-invariant if and only if E is full column rank.*

Example 1 Consider the 2D behaviors $\mathcal{B}_w = \ker R$ and $\mathcal{V}_w = \ker V$ with

$$R = \begin{bmatrix} 1 - s_1^2 & s_2 - s_1 \end{bmatrix} \text{ and } V = \begin{bmatrix} 1 + s_1 & s_2 \\ 0 & s_1 \end{bmatrix}.$$

Note that $R = EV$, with $E = \begin{bmatrix} 1 - s_1 & s_2 - 1 \end{bmatrix}$, and so $\mathcal{V}_w \subset \mathcal{B}_w$. Since E is not fcr, by Corollary 1, $\mathcal{V}_w$ is not $\mathcal{B}_w$-invariant.

Example 2 Consider the 2D behaviors $\mathcal{B}_w = \ker R$ and $\mathcal{V}_w = \ker V$ with

$$R = \begin{bmatrix} 1 - s_1^2 & (1 - s_1)(1 + s_2) \\ (1 + s_1)(1 - s_2) & 1 - s_2^2 \end{bmatrix} \text{ and } V = \begin{bmatrix} 1 + s_1 & 1 + s_2 \end{bmatrix}.$$

We have that $\mathcal{V}_w \subset \mathcal{B}_w$, because $R = EV$ with $E = \begin{bmatrix} 1 - s_1 \\ 1 - s_2 \end{bmatrix}$. Since E is fcr, by Corollary 1, $\mathcal{V}_w$ is $\mathcal{B}_w$-invariant.

In the previous setting, controlled-invariance is defined as follows.

Definition 7 Let $\mathcal{B}_{(w,c)} \subset \mathcal{U}^{\mathrm{w}} \times \mathcal{U}^{\mathrm{c}}$ be an nD behavior. A sub-behavior $\mathcal{V}_w$ of the induced w-behavior $\mathcal{B}_w \subset \mathcal{U}^{\mathrm{w}}$ is said to be $\mathcal{B}_{(w,c)}$-*controlled-invariant* if there exists a behavior $\mathcal{D}_w$ implementable by partial interconnection from $\mathcal{B}_{(w,c)}$, such that $\mathcal{V}_w \subset \mathcal{D}_w$ and $\mathcal{V}_w$ is $\mathcal{D}_w$-invariant.

As expected, not every sub-behavior of $\mathcal{B}_w$ is controlled-invariant. Indeed, if $\mathcal{D}_w$ is implementable by partial interconnection from $\mathcal{B}_{(w,c)}$, then, by Proposition 5, $\mathcal{D}_w$ must contain $\mathcal{N}_w$.

Now, if in particular $\mathcal{V}_w \subset \mathcal{N}_w \subset \mathcal{B}_w$ is a sub-behavior which is not $\mathcal{N}_w$-invariant, then $\mathcal{N}_w/\mathcal{V}_w$ is not autonomous, and hence neither is $\mathcal{D}_w/\mathcal{V}_w$, as it contains $\mathcal{N}_w/\mathcal{V}_w$. Therefore, by Definition 6, the following result holds.

Proposition 6 *Let $\mathcal{B}_{(w,c)} \subset \mathcal{U}^{\mathrm{w}} \times \mathcal{U}^{\mathrm{c}}$ be an nD behavior. Then, if $\mathcal{V}_w \subset \mathcal{N}_w$:*

$$\mathcal{V}_w \textit{ is } \mathcal{B}_{(w,c)}\textit{-controlled-invariant} \Leftrightarrow \mathcal{V}_w \textit{ is } \mathcal{N}_w\textit{-invariant}.$$

The general case, i.e., when $\mathcal{V}_w$ is not necessarily a subset of $\mathcal{N}_w$, will be treated in the sequel by considering two distinct cases.

11.4.1 The Case Where R Is Full Row Rank

In this section we assume that the matrix $R(\underline{s})$ of the (w, c)-behavior description (11.1) is a full row rank polynomial matrix. Recall that, in this case, every partial controller is regular and hence implementability by partial regular interconnection simply reduces to implementability by partial interconnection. The following characterization of controlled invariance then holds for nD behaviors.

Proposition 7 *Let $\mathcal{B}_{(w,c)} \subset \mathcal{U}^{\mathrm{w}} \times \mathcal{U}^{\mathrm{c}}$ be an nD behavior and $\mathcal{V}_w \subset \mathcal{B}_w$. Then:*

$$\mathcal{V}_w \textit{ is } \mathcal{B}_{(w,c)}\textit{-controlled-invariant} \Leftrightarrow \overline{\mathcal{B}}_w/\mathcal{V}_w \textit{ is autonomous},$$

where $\overline{\mathcal{B}}_w := \mathcal{N}_w + \mathcal{V}_w$.

Remark 1 Note that $\overline{\mathcal{B}}_w := \mathcal{N}_w + \mathcal{V}_w$ is the smallest implementable behavior by partial interconnection from $\mathcal{B}_{(w,c)}$ containing $\mathcal{V}_w$. Thus, in this case, $\mathcal{V}_w$ is $\mathcal{B}_{(w,c)}$-controlled-invariant if and only if $\mathcal{V}_w$ is invariant with respect to the smallest implementable behavior that contains it.

Proof of Proposition 7. "$\Leftarrow$" Assume that $\overline{\mathcal{B}}_w/\mathcal{V}_w$ is autonomous. Then, by Definition 6, $\mathcal{V}_w$ is $\overline{\mathcal{B}}_w$-invariant. On the other hand, since $\mathcal{N}_w \subset \overline{\mathcal{B}}_w := \mathcal{N}_w + \mathcal{V}_w$, $\overline{\mathcal{B}}_w$ is implementable from $\mathcal{B}_{(w,c)}$ by partial interconnection. By Definition 7 this means that $\mathcal{V}_w$ is $\mathcal{B}_{(w,c)}$-controlled-invariant.

"$\Rightarrow$" Assume that $\mathcal{V}_w$ is $\mathcal{B}_{(w,c)}$-controlled-invariant. This implies, by definition, that there exists a behavior $\mathcal{D}_w \supset \mathcal{V}_w$ which is implementable by partial interconnection from $\mathcal{B}_{(w,c)}$ and such that $\mathcal{V}_w$ is $\mathcal{D}_w$-invariant. So, by Definition 6, $\mathcal{D}_w/\mathcal{V}_w$ is autonomous. Moreover, the achievability of $\mathcal{D}_w$ implies that $\mathcal{N}_w \subset \mathcal{D}_w$. Thus $\overline{\mathcal{B}}_w := \mathcal{N}_w + \mathcal{V}_w \subset \mathcal{D}_w$ and $\overline{\mathcal{B}}_w/\mathcal{V}_w \subset \mathcal{D}_w/\mathcal{V}_w$. Since $\mathcal{D}_w/\mathcal{V}_w$ is autonomous then the same applies to $\overline{\mathcal{B}}_w/\mathcal{V}_w$. □

Given an nD behavior $\mathcal{B}_{(w,c)}$, the following procedure shows how to construct a controller $\mathcal{C}_c$ that implements $\overline{\mathcal{B}}_w$ by partial interconnection from $\mathcal{B}_{(w,c)}$. Let $\mathcal{B}_{(w,c)}$ be given by (11.1) and $\mathcal{C}_c = \ker C$ such that $\overline{\mathcal{B}}_w = \pi_w \left(\mathcal{B}_{(w,c)} \cap \mathcal{C}^*_{(w,c)} \right)$. Let also $\mathcal{V}_w = \ker V$, $\mathcal{N}_w = \ker R$. Note that, since $\overline{\mathcal{B}}_w = \mathcal{N}_w + \mathcal{V}_w = \ker R + \ker V$, it follows from Proposition 4 that $\overline{\mathcal{B}}_w = \ker F$ with $F = AR = BV$ and $\begin{bmatrix} -A & B \end{bmatrix}$ an MLA of $\begin{bmatrix} R \\ V \end{bmatrix}$.

We show next that $C = AM$ yields the desired controller, i.e., that $\overline{\mathcal{B}}_w$ is the w-behavior described by

$$\begin{bmatrix} R \\ 0 \end{bmatrix} w = \begin{bmatrix} M \\ AM \end{bmatrix} c. \tag{11.3}$$

It is not difficult to see that $\begin{bmatrix} L & 0 \\ A & -I \end{bmatrix}$, where L is an MLA of M, is an MLA of $\begin{bmatrix} M \\ AM \end{bmatrix}$ and so the w-behavior corresponding to (11.3) is

$$\ker \begin{bmatrix} L & 0 \\ A & -I \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} = \ker \begin{bmatrix} LR \\ AR \end{bmatrix} = \ker LR \cap \ker AR = \mathcal{B}_w \cap \overline{\mathcal{B}}_w = \overline{\mathcal{B}}_w,$$

since $\overline{\mathcal{B}}_w \subset \mathcal{B}_w$.

The next results provide a characterization of controlled invariance in terms of the matrix representations associated with the relevant behaviors, in the case where, besides R, also the representation matrix V of $\mathcal{V}_w$ has full row rank.

Proposition 8 *Consider the nD behaviors $\mathcal{V}_w = \ker V$, $\mathcal{N}_w = \ker R$, with R and V full row rank, and $\overline{\mathcal{B}}_w = \mathcal{N}_w + \mathcal{V}_w$. Then*

$$\text{rank}\begin{bmatrix} R \\ V \end{bmatrix} = \text{rank}\, R \Leftrightarrow \overline{\mathcal{B}}_w/\mathcal{V}_w \textit{ is autonomous.}$$

Proof As noticed before, $\overline{\mathcal{B}}_w = \ker F$ with $F = AR = BV$ and $\begin{bmatrix} -A & B \end{bmatrix}$ an MLA of $\begin{bmatrix} R \\ V \end{bmatrix}$. Moreover, by Proposition 3, $\overline{\mathcal{B}}_w/\mathcal{V}_w \simeq \ker B$.

"$\Leftarrow$" Assume that $\overline{\mathcal{B}}_w/\mathcal{V}_w$ is autonomous. Since, by hypothesis, R has full row rank, there exists a square nonsingular nD rational matrix U such that $RU = \begin{bmatrix} R_1 & 0 \end{bmatrix}$, with R_1 square and non singular. Thus

$$\begin{bmatrix} R \\ V \end{bmatrix} U = \begin{bmatrix} R_1 & 0 \\ V_1 & V_2 \end{bmatrix},$$

for a suitable partition of VU. Since $\begin{bmatrix} -A & B \end{bmatrix}$ is an MLA of $\begin{bmatrix} R \\ V \end{bmatrix}$ then $BV_2 = 0$. But the fact that $\overline{\mathcal{B}}_w/\mathcal{V}_w$ is autonomous implies that B is full column rank and so $V_2 = 0$. Hence, taking into account that R_1 is square and nonsingular

$$\text{rank}\begin{bmatrix} R \\ V \end{bmatrix} = \text{rank}\begin{bmatrix} R_1 \\ V_1 \end{bmatrix} = \text{rank}\, R_1 = \text{rank}\begin{bmatrix} R_1 & 0 \end{bmatrix} = \text{rank}\, RU = \text{rank}\, R.$$

"$\Rightarrow$" Assume now that $\text{rank}\begin{bmatrix} R \\ V \end{bmatrix} = \text{rank}\, R$. Then if U is a square nonsingular nD rational matrix such that $RU = \begin{bmatrix} R_1 & 0 \end{bmatrix}$ with R_1 square and nonsingular, VU must be of the form $VU = \begin{bmatrix} V_1 & 0 \end{bmatrix}$. Moreover, it is possible to define U in such a way that both R_1 and V_1 are polynomial matrices. Now, $\text{rank}\begin{bmatrix} R \\ V \end{bmatrix} = \text{rank}\begin{bmatrix} R_1 & 0 \\ V_1 & 0 \end{bmatrix}$ and obviously $\begin{bmatrix} -A & B \end{bmatrix}$ is an MLA of $\begin{bmatrix} R \\ V \end{bmatrix}$ if and only if it is an MLA of $\begin{bmatrix} R_1 \\ V_1 \end{bmatrix}$. Now, it follows from Theorem 1 that $\begin{bmatrix} A & -B \end{bmatrix}$ is an MLA of $\begin{bmatrix} R_1 \\ V_1 \end{bmatrix}$ if and only if the pair (A, B) is a left coprime factorization of $G_1 = V_1 R_1^{-1}$ which in turn implies that B has full column rank. Finally, since $\overline{\mathcal{B}}_w/\mathcal{V}_w \simeq \ker B$ we conclude that $\overline{\mathcal{B}}_w/\mathcal{V}_w$ is autonomous. □

The following corollary is an immediate consequence of the previous results.

Corollary 2 *Consider the nD behavior $\mathcal{B}_{(w,c)}$ described by $Rw = Mc$, let $\mathcal{B}_w = \pi_w\left(\mathcal{B}_{(w,c)}\right)$ and $\mathcal{V}_w = \ker V \subset \mathcal{B}_w$. Moreover assume that R and V have full row rank. Then*

$$\mathcal{V}_w \subset \mathcal{B}_w \textit{ is } \mathcal{B}_{(w,c)}\textit{-controlled-invariant} \Leftrightarrow \text{rank}\begin{bmatrix} R \\ V \end{bmatrix} = \text{rank}\, R.$$

11.4.2 The Case Where R Is Not Full Row Rank

When the matrix $R(\underline{s})$ of the (w, c)-behavior description (11.1) is not full row rank, a characterization of controlled-invariance for general nD systems is still a subject under investigation. However, in the 1D case it is possible to obtain some results taking advantage of Proposition 5(b), which allows to convert the problem of implementability by regular partial interconnection, involving the variables w and c, to a problem stated only in terms of w-behaviors.

Proposition 9 *Let $\mathcal{B}_{(w,c)}$ be a 1D behavior described by $Rw = Mc$, $\mathcal{B}_w = \pi_w\left(\mathcal{B}_{(w,c)}\right)$ and $\mathcal{V}_w = \ker V \subset \mathcal{B}_w$. Then*

$$\mathcal{V}_w \text{ is } \mathcal{B}_{(w,c)}\text{-controlled-invariant} \Leftrightarrow \exists\, \mathcal{D}_w \subset \mathcal{B}_w \text{ such that:}$$

- $\mathcal{N}_w + \mathcal{V}_w \subset \mathcal{D}_w$;
- $\mathcal{B}_w/\mathcal{D}_w$ *controllable;*
- $\mathcal{D}_w/\mathcal{V}_w$ *autonomous.*

Proof "$\Leftarrow$" The existence of $\mathcal{D}_w \subset \mathcal{B}_w$ such that $\mathcal{N}_w + \mathcal{V}_w \subset \mathcal{D}_w$ and $\mathcal{B}_w/\mathcal{D}_w$ controllable implies, by Proposition 5(b), that $\mathcal{D}_w$ is implementable from $\mathcal{B}_{(w,c)}$ by regular partial interconnection. Moreover, by hypothesis, $\mathcal{D}_w/\mathcal{V}_w$ autonomous and so, by Definition 6, $\mathcal{V}_w$ is $\mathcal{D}_w$-invariant. Hence, $\mathcal{V}_w$ is $\mathcal{B}_{(w,c)}$-controlled-invariant by definition.

"$\Rightarrow$" The proof is analogous. □

Note that when w is *observable* from c (i.e., if $c \equiv 0$ implies $w \equiv 0$), $\mathcal{N}_w = \{0\}$ and the conditions of the previous proposition amounts to the existence of an "intermediate" behavior $\mathcal{D}_w$, with $\mathcal{V}_w \subset \mathcal{D}_w \subset \mathcal{B}_w$, such that $\mathcal{B}_w/\mathcal{D}_w$ is controllable and $\mathcal{D}_w/\mathcal{V}_w$ is autonomous.

Example 3 Consider the state space system

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$$

with

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 2 & 3 \\ 1 & 0 & 4 \end{bmatrix},\ B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix},$$

and let $\mathcal{B}_{(w,c)}$ be described by $Rw = Mc$, where

$$R = \begin{bmatrix} \frac{d}{dt}I - A \\ C \end{bmatrix},\ w = x,\ M = \begin{bmatrix} B & 0 \\ 0 & 1 \end{bmatrix} \text{ and } c = \begin{bmatrix} u \\ y \end{bmatrix}.$$

It is easy to check that

$$\mathcal{B}_w = \ker \left[\tfrac{d}{dt} \; -1 \; 1 \; -1 \right] \text{ and } \mathcal{N}_w = \{0\}.$$

Let $\mathcal{V}_w \subset \mathcal{B}_w$ be described by $\mathcal{V}_w = \ker V$ with

$$V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix}.$$

Is follows that $\mathcal{V}_w$ is $\mathcal{B}_{(w,c)}$-controlled-invariant. Indeed, considering the behavior

$$\mathcal{D}_w = \ker \begin{bmatrix} \frac{d}{dt} - 1 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix} = \ker \begin{bmatrix} \frac{d}{dt} - 1 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix},$$

clearly $\mathcal{V}_w \subset \mathcal{D}_w \subset \mathcal{B}_w$ and moreover

$$\mathcal{B}_w / \mathcal{D}_w \cong \ker \begin{bmatrix} 1 & 0 \end{bmatrix} \text{ is controllable}$$

and

$$\mathcal{D}_w / \mathcal{V}_w \cong \ker \begin{bmatrix} \frac{d}{dt} - 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ is autonomous.}$$

Recall that in the classical state-space setting [1] a subspace $\mathcal{V}$ of the state space $\mathcal{X}$ is said to be (A, B)-*controlled invariant* if

$$A\mathcal{V} \subseteq \mathcal{V} + \operatorname{im} B.$$

Considering the matrices A, B and V of the previous example, straightforward calculations show that $\mathcal{V} = \ker V$ is also (A, B)-controlled invariant as a subspace of the state space $\mathcal{X} = \mathbb{R}^3$.

11.5 Conclusions and Future Work

In this chapter, controlled invariance in the context of nD behavioral systems has been introduced and characterized. The case where every controller is regular was completely characterized, including the construction of the controllers that achieve invariance, while in the case where the controllers are not necessarily regular only preliminary results for 1D systems were given. The overall problem of behavioral control invariance for nD systems is under current investigation. In this case the situation is considerably more involved, as the regularity of the partial interconnection cannot in general be assumed without loss of generality.

Acknowledgements This work is supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology, references UIDB/04106/2020 and UIDP/04106/2020 and also by UIDB/00147/2020—SYSTEC—Research Center for Systems and Technologies funded by national funds through the FCT/MCTES (PIDDAC). The authors thank Diego Napp for his helpful comments.

References

1. Basile, G., Marro, G.: Controlled and conditioned invariant subspaces in linear system theory. J. Optim. Theory Appl. **3**(5), 306–315 (1969)
2. Belur, M., Trentelman, H.L.: Stabilization, pole placement, and regular implementability. IEEE Trans. Autom. Control **47**(5), 735–744 (2002)
3. Oberst, U.: Multidimensional constant linear systems. Acta Appl. Math. **20**, 1–175 (1990)
4. Pereira, R., Rocha, P.: A remark on conditioned invariance in the behavioral approach. In: European Control Conference 2013, ECC'13, pp. 301–305. ETH Zurich, Switzerland (2013)
5. Pillai, H., Shankar, S.: A behavioral approach to control of distributed systems. SIAM J. Control Optim. **37**(2), 388–408 (1998)
6. Rocha, P.: Canonical controllers and regular implementation of nD behaviors. In: Proceedings of the 16th IFAC World Congress. Czech Republic, Prague (2005)
7. Rocha, P., Wood, J.: A new perspective on controllability properties for dynamical systems. Int. J. Appl. Math. Comput. Sci. **7**(4), 869–879 (1997)
8. Rocha, P., Wood, J.: Trajectory control and interconnection of 1D and nD systems. SIAM J. Control Optim. **40**(1), 107–134 (2001)
9. Trentelman, H.L., Avelli, D.N.: On the regular implementability of nD systems. Syst. Control Lett. **56**(4), 265–271 (2007)
10. Valcher, M.: Characteristic cones and stability properties of two-dimensional autonomous behaviors. IEEE Trans. Circuits Syst. I **47**, 290–302 (2000)
11. Willems, J.C.: Paradigms and puzzles in the theory of dynamical systems. IEEE Trans. Autom. Control **36**(3), 259–294 (1991)
12. Willems, J.C., Trentelman, H.L.: Synthesis of dissipative systems using quadratic differential forms, Part I. IEEE Trans. Autom. Control. **47**(1), 53–69 (2002)
13. Wood, J.: Modules and behaviors in nD systems theory. Multidimens. Syst. Signal Process. **11**, 11–48 (2000)
14. Wood, J., Oberst, U., Rogers, E., Owens, D.H.: A behavioral approach to the pole structure of one-dimensional and multidimensional linear systems. SIAM J. Control Optim. **38**, 627–661 (2000)
15. Wood, J., Rogers, E., Owens, D.H.: Controllable and autonomous nD linear systems. Multidimens. Syst. Signal Process. **10**, 33–69 (1999)
16. Zerz, E.: Topics in Multidimensional Linear System Theory. Springer Lecture Notes in Control and Information Theory, vol. 256. Springer, London (2000)

Index

A. Quadrat and E. Zerz (eds.), *Algebraic and Symbolic Computation Methods in Dynamical Systems*, Advances in Delays and Dynamics 9,
https://doi.org/10.1007/978-3-030-38356-5