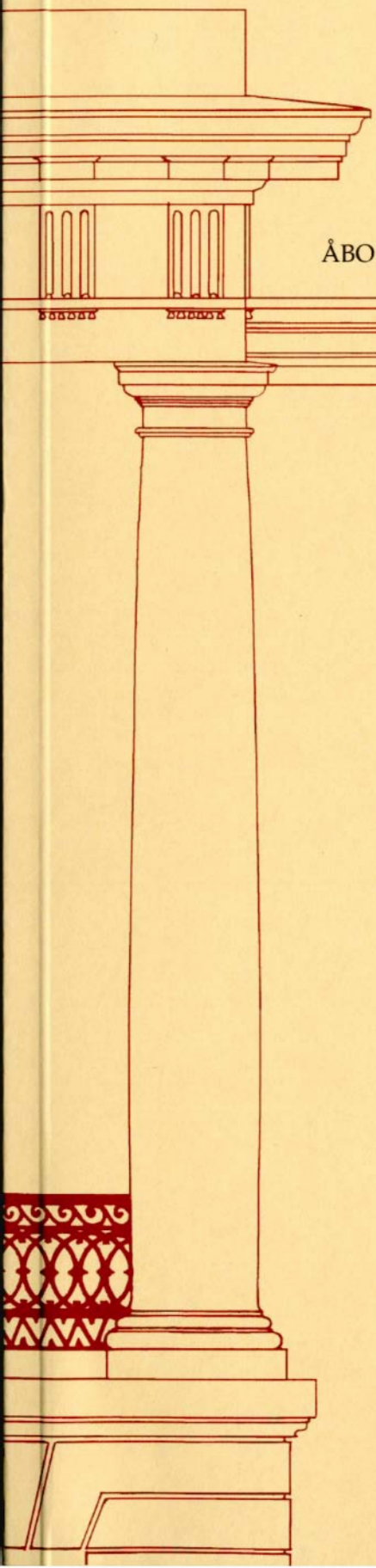


ÅBO AKADEMIS FÖRLAG – ÅBO AKADEMI UNIVERSITY PRESS

# CADE - 2007

## Computer Algebra and Differential Equations

A. Mylläri, V. Edneral and N. Ourusoff
Editors

ACTA ACADEMIAE ABOENSIS, SER. B

Mathematica et physica

Mathematics, science, engineering

Edited by Göran Högnäs

**Vol. 67, no. 2**

CADE - 2007

# CADE - 2007

## Computer Algebra and Differential Equations

A. Mylläri, V. Edneral and N. Ourusoff
Editors

**ÅBO 2007**

**ÅBO AKADEMIS FÖRLAG – ÅBO AKADEMI UNIVERSITY PRESS**

**CIP Cataloguing in Publication**

**CADE** -2007 : computer algebra and differential equations / A. Mylläri, V. Edneral, and N. Ourusoff. – Åbo : Åbo Akademi University Press, 2007. – (Acta Academiae Aboensis, Ser. B, Mathematica et physica, ISSN 0001-5105; vol. 67, no.2)
ISBN 978-951-765-402-9

ISSN 0001-5105
ISBN 978-951-765-402-9
ISBN 978-951-765-403-6 (digital)
UNIPRINT
Åbo 2007

# Contents

The workshop "Computer algebra and differential equations / CADE -2007" continues a series of International workshops on computer algebra and its applications held in Dubna in 1979, 1982, 1985, 1990 and 2001. This series of conferences covered all areas of computer algebra. Until the early 90's this research area was called symbolic and algebraic computation, and the main application fields were theoretical physics and mechanics. In addition to the above-mentioned workshops, an additional workshop entitled "Computer algebra in differential equations", organized by the Technical University of Prague, was held in Prague in 1997.

The workshop series was discontinued until this year when the Åbo Akademi University kindly agreed to resume the series with the application of computer algebra to differential equations as the topic. This initiative was supported by the Department of Information Technology (University of Turku) and the Turku Center for Computer Science (TUCS). CADE - 2007 convened participants from Russia, Finland, Spain, Germany, France, UK, and Norway. Unofficially, this Workshop was dedicated to the $60^{th}$ birthday of Vladimir Gerdt, who was one of the pioneers of Computer Algebra in the former Soviet Union and is one of the main moving forces in organizing this series of conferences.

The conference schedule included twenty-five talks, of which six were invited. In addition to the talks, there were numerous and fruitful discussions stimulated by multiple coffee-breaks, informal parties and visits to the aqua park and sauna.

The appearance of this volume as well as the organization of the conference itself would not have been possible without financial support from the Academy of Finland, the Magnus Ehrnrooth Foundation, and the Research Institute of Åbo Akademi University Foundation. We also appreciate the financial and organizational support from the Turku Center for Computer Science (TUCS). The Department of Information Technology of the University of Turku provided copy, printing and other facilities. We give special thanks to Irmeli Laine, secretary of TUCS, whose help with practical arrangements was invaluable and also to Maja Sundberg, whose expert help with the lodging of participants is appreciated.

Victor Edneral  Aleksandr Mylläri

## Program Committee

V. Edneral (Moscow State University, Russia) - chair
V. Gerdt (Joint Institute for Nuclear Research, Dubna, Russia)
J. Hietarinta (University of Turku, Finland)
G. Högnäs (Åbo Akademi University, Finland)
J. Karhumäki (University of Turku, Finland)
R. Kragler (University of Applied Sciences, Weingarten, Germany)
A. Mylläri (Åbo Akademi University, Finland)
N. Vasiliev (PDMI RAS, St.Petersburg, Russia)
S. Walcher (RWTH Aachen University, Germany)

## Local Organizing Committee

A. Mylläri - chair
G. Högnäs
T. Mylläri

## Conference participants.

**Third row:** M. Zinin, S. Evlakhov, J. Tuomela, J. Palacián, V. Kornyak, Yu. Blinkov, L. Floría, T. Mylläri, A. Mylläri
**Second row:** N. Vassiliev, P. Burovskiy, T. Sturm, A. Elipe, V. Keränen, A. Gusev, V. Rostovtsev, V. Lychagin, D. Stefănescu
**First row:** V. Edneral, I. Edneral, Hoang Ngoc Minh, S. Walcher, V. Gerdt, P. Yanguas, A. Zobnin, A. Semenov

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# On Completion to Involution Based on Janet Division

Vladimir P.Gerdt

*Laboratory of Information Technologies*
*Joint Institute for Nuclear Research*
*141980 Dubna, Russia*
*E-mail: gerdt@jinr.ru*

**Abstract.** In this paper we outline a general algorithmic approach to systems of algebraic, differential and (linear) difference equations which depend polynomially on unknowns with rational function coefficients. For pure algebraic or linear differential / difference systems the approach is based on their completion to involution. We present a completion algorithm based on the Janet monomial division. In the case of polynomial-nonlinear systems of differential type they can be decomposed into a finite number of involutive subsystems.

## 1 Completion to Involution

Among the most brilliant results in theory of analytical partial differential equations (PDEs) is the Cauchy-Kowalevskaya theorem which establishes a class of quasi-linear PDEs which admit posing an initial-value problem providing existence and uniqueness of the analytical solution.

The main obstacle in investigating other classes of PDE systems of some given order $q$ is existence of *integrability conditions*, that is, such differential but not pure algebraic consequences of equations in the system whose derivatives have (total) order $\leq q$. A *formally integrable* system of PDEs has all its integrability conditions incorporated in the system. This means that *prolongations* of the system do not reveal integrability conditions. An *involutive system* is a formally integrable one of special type [1, 2, 3]. The quasi-linear systems of Cauchy-Kowalevskaya type form a particular family of involutive systems.

Extension of a system by its integrability conditions is called *completion*. From the completion point of view the linear homogeneous PDE systems with constant coefficients can be associated with pure polynomial systems [2, 3, 4], and polynomial involutive systems [5] provide a fruitful algorithmic tool in commutative algebra [6].

The most general involutive approach is based on the concept of *involutive monomial division* [5] which is defined for a finite monomial set. Every particular

division provides, for each monomial in the set, the self-consistent separation of variables into multiplicative and non-multiplicative. Completion to involution is performed by combining of non-multiplicative prolongations with multiplicative reductions.

In differential / difference algebra [7], in addition to their role as canonical bases of differential ideals, linear and quasi-linear (orthonomic) involutive systems allow one to construct the Hilbert polynomial and thereby to determine the structure of arbitrariness in general solution [4, 6]. Recently, it was established that in linear difference algebra, involutive bases as well as Gröbner bases form a useful algorithmic tool to construct finite difference schemes for linear PDEs [8] and to reduce multiloop Feynman integrals [9].

Involutive bases of polynomial ideals [5, 10] as well as of linear differential [4] / difference ideals [8] are Gröbner bases [11, 12] of special form. Wu Wen-tsün was the first who pointed out this fact [13]. Though involutive bases are generally redundant as Gröbner ones, their use makes more accessible the structural information of ideals. Janet and Pommaret bases may be cited as typical representatives of involutive bases [5, 6, 10].

# 2 Involutive Bases

## 2.1 Polynomial Bases

The basic algorithmic ideas go back to M.Janet [2] who invented a constructive approach to study PDEs in terms of corresponding monomial sets based on the following association between derivatives and monomials:

$$\frac{\partial^{\mu_1+\cdots+\mu_n} u^\alpha}{\partial x_1^{\mu_1} \cdots \partial x_n^{\mu_n}} \Longleftrightarrow [x_1^{\mu_1} \cdots x_n^{\mu_n}]_\alpha \,. \tag{1}$$

The monomials associated with the different dependent variables $u^\alpha$ are to be considered as belonging to different monomial sets.

The association (1) allows one to reduce the involution analysis of linear homogeneous systems of PDEs or difference equations with constant coefficients to pure algebraic systems [2, 3, 4, 9]. Having this fact in mind, we first consider involutive algebraic systems.

Let $\mathbb{R} = \mathbb{K}[x_1, \ldots, x_n]$ be a ring of multivariate polynomials over a zero characteristic coefficient field $\mathbb{K}$. Then a finite set $F = \{f_1, \ldots, f_m\} \subset \mathbb{R}$ of polynomials in $\mathbb{R}$ is *a basis of the ideal*

$$< F >=< f_1, \ldots, f_m >= \{ \sum_{i=1}^{m} h_i f_i \mid h_j \in \mathbb{R} \} \,.$$

In the involutive approach to commutative (polynomial) algebra [5, 10], which is a mapping of the involutivity analysis of linear PDEs [3, 4], for every polynomial in a finite set $F$, variables $x_1, \ldots, x_n$ are separated into the disjoint subsets of *multiplicative* and *nonmultiplicative* variables.

To be self-consistent such a separation must satisfy certain axioms [5], and every appropriate separation generates an *involutive monomial division* in the following sense. Fix a linear *admissible monomial order* $\succ$, that is, one satisfying

$$m \neq 1 \Longrightarrow m \succ 1\,, \quad m_1 \succ m_2 \iff m_1 m \succ m_2 m\,, \quad \forall\, m, m_1, m_2 \in \mathbb{R} \qquad (2)$$

where $m$, $m_1$, $m_2$ are monomials – power products of the variables with integer exponents. Then for every polynomial $f \in F$ one can select its *leading monomial* $\mathrm{lm}(f)$ (with respect to $\succ$). All leading monomials in $F$ form a finite monomial set $U$. If $u \in U$ divides a monomial $w$ such that all the variables which occur in $w/u$ are multiplicative for $u$, then $u$ is called an *involutive divisor* of $w$. We shall denote by $\mathcal{L}$ an involutive division, which specifies a set of multiplicative (resp. nonmultiplicative) variables for every monomial $u$ in any given finite monomial set $U$ and write $u|_{\mathcal{L}} w$ if $u$ is an ($\mathcal{L}-$)involutive divisor of $w$. In the latter case we shall also write $w = u \times v$ where, by the above definition, monomial $v = w/u$ contains only multiplicative variables.

As an example of involutive division we consider here one named after M.Janet [5], one of the founders of the involutive approach to PDEs who devised the related separation of variables [2].

Given a finite set $U$ of monomials in $\{x_1, \ldots, x_n\}$ and a monomial $u = x_1^{d_1} \cdots x_n^{d_n} \in U$, a variable $x_i$ ($i > 1$) is *Janet multiplicative* for $u$ if its degree $d_i$ in $u$ is maximal among all the monomials in $U$ having the same degrees in variables $x_1, \ldots, x_{i-1}$. As for $x_1$, it is Janet multiplicative for $u$ if $d_1$ takes the maximal value among degrees in $x_1$ of monomials in $U$. If a variable is not Janet multiplicative for $u$ in $U$ it is considered as *Janet nonmultiplicative*.

Consider, for example, a monomial set $U = \{u = x_1^2 x_2, v = x_2 x_3^2, w = x_3^3\}$. This gives the following Janet multiplicative and nonmultiplicative variables for monomials in $U$:

| Monomial | Variables | |
|---|---|---|
| | Multiplicative | Nonmultiplicative |
| $x_1^2 x_2$ | $x_1, x_2, x_3$ | $-$ |
| $x_2 x_3^2$ | $x_2, x_3$ | $x_1$ |
| $x_3^2$ | $x_3$ | $x_1, x_2$ |

Given a finite polynomial set $F$, a noetherian [5, 10] involutive division $\mathcal{L}$, for instance, Janet division, and an admissible monomial order $\succ$, one can algorithmically construct [5, 10] a *minimal $\mathcal{L}-$involutive basis* or $\mathcal{L}-$basis $G \subset \mathbb{R}$ of ideal $< F >=< G >$ such that for any polynomial $f$ in the ideal there is a polynomial $g$ in $G$ satisfying $\mathrm{lm}(g)|_{\mathcal{L}} \mathrm{lm}(f)$, and every polynomial $g$ in $G$ does not contain monomials having involutive divisors among the leading monomials of other polynomials in $G$.

If $F = \{f_1, \ldots, f_m\} \subset \mathbb{R}$ is a polynomial set, $\mathcal{L}$ is an involutive division, and $\succ$ is an admissible monomial order, then any polynomial $p$ in $R$ can be rewritten

(reduced) modulo ideal $< F >$ as

$$p = h - \sum_{i=1}^{m} \sum_{j} a_{ij} f_i \times u_{ij} \tag{3}$$

where $a_{ij}$ are elements ( coefficients ) of the ground field $\mathbb{K}$, $u_{ij}$ are $\mathcal{L}-$multiplicative monomials for $\mathrm{lm}(f_i)$ such that $\mathrm{lm}(f)\, u_{ij} \preceq \mathrm{lm}(p)$ for all $i, j$, and there are no monomials occurring in $h$ which have $\mathcal{L}-$ involutive divisors among $\mathrm{lm}(f_i)$ $(1 \leq i \leq m)$. In this case $h$ is said to be in the $\mathcal{L}-$*normal form* modulo $F$ and written as $h = NF_{\mathcal{L}}(p, F)$.
If $G$ is an $\mathcal{L}-$basis, then $NF_{\mathcal{L}}(p, G)$ is uniquely defined[1] for any polynomial $p$. In this case $NF_{\mathcal{L}}(p, G) = 0$ if and only if $p$ belongs to the ideal $< G >$ generated by $G$. Moreover, if the ideal is *radical* for which any its elements (polynomial) vanish at the common roots of all the polynomials in $G$ if and only if this polynomial belongs to the ideal, then it follows that the condition $NF_{\mathcal{L}}(p, G) = 0$ is necessary and sufficient for vanishing $p$ on those common roots.

It is important to emphasize that any involutive basis is a Gröbner basis, generally redundant, and can be used in the same manner as the reduced Gröbner basis [11, 12].

## 2.2 Linear Differential/Difference Bases

Let $\mathbb{N}_{>0}$ be the set of positive integers, $\mathbb{N}_{\geq 0}$ be the set of nonnegative integers, $Y := \{\, y^j(x_1, \ldots, x_n)\} \mid j = 1, \ldots, m;\ \ m, n \in \mathbb{N}_{>0}\}$ be a set of *dependent variables*, i.e., functions in $n$-variables, and $\theta_i$ be the derivation operator for the $i$-th variable in the differential case

$$\theta_i \circ y^j(x_1, \ldots, x_n) := \frac{\partial y^j(x_1, \ldots, x_n)}{\partial x_i}$$

or the right-shift operator in the difference case:

$$\theta_i \circ y^j(x_1, \ldots, x_n) := y^j(x_1, \ldots, x_i + 1, \ldots, x_n)\,.$$

For the power products $\theta_1^{\mu_1} \cdots \theta_n^{\mu_n}$ of derivation/shift operators we shall use the multiindex notation $\theta^\mu$ where $\mu := \{\mu_1, \ldots, \mu_n\}$ $(\mu \in \mathbb{N}^n_{\geq 0})$ with $|\mu| := \sum \mu_i$. The set of all such (operator) products will be denoted by $\Theta$.

Then the most general form of a system of $K \in \mathbb{N}_{>0}$ partial $(n > 1)$ and multivariate $(m > 1)$ linear differential/difference equations is given by

$$a_0 + \sum_{j=1}^{m} \sum_{\nu} a_{kj;\,\nu} \vartheta_k^\nu \circ y^j = 0\,, \qquad k = 1, \ldots, K\,, \qquad \vartheta_k^\nu \in \Theta\,, \tag{4}$$

where all sums are finite and coefficients $a_0, a_{kj;\,\nu}$ may depend on the variables in $X := \{x_1, \ldots, x_n\}$ and on a finite set of parameters in $C := \{c_1, \ldots\}$. Hereafter we shall assume that all coefficients in (4) are rational functions of the

[1] For other properties of involutive bases, proofs and illustrating examples see [5, 10].

(independent) variables and parameters with integer coefficients:

$$a_0, a_{kj;\,\nu} \in \mathbb{Q}(X \cup C)\,. \tag{5}$$

As well as for the polynomial case of Sect.2.1, the Gröbner basis form of system (4) is defined by a linear order on terms $\theta^\mu \circ y^j$, in this case called *a ranking* [7], and such that for all $i, j, k, \mu, \nu$ the following analogue of (2) holds for all $\mu, \nu, i, j, k$:

$$\theta_i\theta^\mu \circ y^j \succ \theta^\mu \circ y^j\,, \quad \theta^\mu \circ y^j \succ \theta^\nu \circ y^k \iff \theta_i\theta^\mu \circ y^j \succ \theta_i\theta^\nu \circ y^k\,. \tag{6}$$

When $|\mu| \succ |\nu|$ implies $\theta^\mu \circ y^j \succ \theta^\nu \circ y^k$ the ranking is called *orderly*. When $j > k$ implies $\theta^\mu \circ y^j \succ \theta^\nu \circ y^k$ the ranking is called *elimination*.

Denote by $f_k$ the left-hand side of the $k$-th equation in (4) and by $F := \{f_1, \ldots, f_K\}$ the set of all the left-hand sides in the system. Then fixing a ranking $\succ$ provides every $f \in F$ with the *leading term* $\mathrm{lt}(f) := a\,\vartheta \circ y^j$ ($\vartheta \in \Theta$, $a \neq 0$), *leading monomial* $\mathrm{lm}(f) := \vartheta \circ y^j$, and *leading coefficient* $\mathrm{lc}(f) := a$. Furthermore, denote by $R \supset F$ the set of all right-hand sides $f \neq 0$ for linear difference equations $f = 0$ which are consequences of system (4-5). $F$ is called *a generating set or basis* of $R$ (denotation: $R =< F >$). In what follows we shall assume that, given a ranking $\succ$, all $f \in R$ are normalized, that is, divided by their leading coefficients. Furthermore, $\mathrm{lm}(F)$ will denote the set of leading monomials and $\mathrm{lm}_j(F)$ will denote its subset for function $y^j$. Therefore,

$$\mathrm{lm}(F) = \cup_{j=1}^{m} \mathrm{lm}_j(F)\,. \tag{7}$$

Now we are ready to define a *Gröbner basis*, given a set $F$ and a ranking $\succ$, as a finite subset $G \subset R =< F >$ such that $R =< G >$ and

$$\forall f \in R\,, \exists\, g \in G,\, \theta \in \Theta \;:\; \mathrm{lm}(f) = \theta \circ \mathrm{lm}(g)\,. \tag{8}$$

It follows that the leading term of every $f \in R$ is *reducible modulo* $G$ and yields *the head reduction*:

$$f \underset{g}{\to} f' := f - \theta \circ g, \quad f' \in R\,.$$

If $f' \neq 0$, then its leading term is again reducible modulo $G$, and, by repeating the reduction finitely many times [7, 11, 12], we obtain $f \underset{G}{\longrightarrow} 0$. Generally, if a linear differential/difference expression (not necessarily from $R$) contains a term $u$ with coefficient $c \neq 0$ such that $u = \vartheta \circ \mathrm{lm}(f)$ for some $\vartheta \in \Theta$ and $f \in F \subset \mathbb{R}$, then $h$ can be reduced:

$$h \underset{g}{\to} h' := h - c\,\vartheta \circ f\,. \tag{9}$$

By applying the reduction finitely many times, one obtains a polynomial $\bar{h}$ which is either zero or such that all its (nonzero) terms are *irreducible modulo* $F$. In both cases $\bar{h}$ is said to be in the *normal form modulo* $F$ (denotation: $\bar{h} = NF(h, F)$). A Gröbner basis $G$ is called *reduced* if $g = NF(g, G \setminus \{g\})$ for every $g \in G$.

With the use of association (1) and its difference analogue one can define multiplicative and nonmultiplicative differential/difference operators (prolongations) $\theta_j$ for any involutive monomial division [5]. In doing so, this separation must be done for every indexed group of the leading terms (monomials) in (7). Below we can consider the Janet division as it defined in Sect.2.1.

A finite set $G \in R =< F >$ is called a *Janet basis* (cf. [10]) if

$$\forall\, f \in R\,, \exists\, g \in G\,, \theta \in J(g,G) \;:\; \mathrm{lm}(f) = \theta \circ \mathrm{lm}(g) \tag{10}$$

where $J(g,G)$ is the *monoid* of $J-$multiplicative operators for $g \in G$. Similarly to (9), a $J-$*reduction* is defined as

$$h \underset{g}{\rightarrow} h' := h - c\,\vartheta \circ f \tag{11}$$

for a polynomial $h \in R$ containing a term $u$ with coefficient $c \neq 0$ satisfying $u = \vartheta \circ \mathrm{lm}(f)$ for some $f \in F$ and $\vartheta \in J(f,F)$.

Since $J-$reducibility (11) implies the Gröbner reducibility (9), a Janet basis satisfying (10) is also a Gröbner basis. The converse is generally not true, that is, not every Gröbner basis is a Janet one. The algorithmic characterization of a Janet basis $G$ is the following condition (cf. [10]):

$$\forall\, g \in G\,, \vartheta \in NM_J(g,G) :\; NF_J(\vartheta \circ g, G) = 0\,. \tag{12}$$

where $NM_J(g,G)$ is the set of nonmultiplicative derivation / difference operators for $g \in G$. Condition (12) is a cornerstone of the following algorithm for construction of Janet bases.

**Algorithm: *JanetBasis*$(F, \succ)$**

```
1: choose f ∈ F with the lowest lm(f) w.r.t. ≻
2: G := {f};  Q := F \ {f}
3: do
4:    h := 0
5:    while Q ≠ ∅ and h = 0 do
6:       choose p ∈ Q with the lowest lm(p) w.r.t. ≻
7:       Q := Q \ {p};  h :=NF_J(p, G)
8:    od
9:    if h ≠ 0 then
10:      if lm(h) ≠ lm(p) then
11:         for all g ∈ G such that lm(g) = θ^μ ∘ lm(h), |μ| > 0 do
12:            Q := Q ∪ {g};  G := G \ {g}
13:         od
14:      fi
15:      G := G ∪ {h/lc(h)};  Q := Q ∪ { θ^β ∘ g | g ∈ G, θ^β ∈ NM_J(g, G) }
16:   fi
17: od while Q ≠ ∅
18: return G
```

This algorithm, including the subalgorithm $\boldsymbol{NF}_J$ in line 7 which computes the Janet normal form, constructs a minimal and reduced differential/difference Janet basis for a linear input set $F$ and also a pure polynomial Janet basis if one replaces the operators $\theta^\mu$ by the corresponding products of variables. In its improved form without the repeated prolongations and with avoidance of some useless zero reductions the algorithm has been implemented in *Maple* [14, 15] and, for the polynomial case, in C/C++ (see references in [10]).

# 3 Nonlinear Differential Systems

For a finite system $F$ of polynomial-nonlinear PDEs one can also use a ranking $\succ$ satisfying (6) to select the *leading derivative (leader)* in any equations of the system and thereby to separate derivative operators $\theta_i$ into multiplicative and nonmultiplicative. However, the concept (11) of reduction admits a natural generalization only to quasi-linear (orthonomic) equations, i.e. those that are linear with respect to their highest ranking derivative. Moreover, in the course of completion to involution a non-orthonomic integrability condition may arise that makes impossible a straightforward generalization of the above algorithm ***JanetBasis*** to nonlinear systems. Even worse, differential ideals generated by a finite set of polynomials may not have finite Gröbner [16], and, thereby, involutive bases. This means that there is no hope of generalizing the algorithm to nonlinear differential polynomials. Instead, one can use a decomposition of the initial differential system into finitely many involutive subsystems. Such a decomposition was suggested by J.F.Thomas [17] who applied his own separation of variables into multiplicative and nonmultiplicative sets that generate an involutive division called Thomas division [5]. Thomas also gives an algorithm for explicit construction of the decomposition.

In so doing, Thomas considers, first, an input system of differential polynomials $F := \{f_1, \ldots, f_m\}$ as an algebraic system in the dependent variables $y^j$ and their derivatives $\{\theta y^j \mid \theta \in \Theta, j = 1, \ldots, k\}$ which occur in $F$, whereas independent variables $\{x_1, \ldots, x_n\}$ are treated as parameters. In accordance with the ranking $\succ$ chosen, every $f \in F$ is a polynomial in its *leader* (leading derivative with denotation: $\mathrm{ld}(f)$) whose coefficient in the term in $f$ of the maximal degree in $\mathrm{ld}(f)$ is called *initial* and will be denoted by $\mathrm{init}(f)$.

Thomas introduced a concept of *simple algebraic system* which has a triangular form. Some of the underlying simplicity properties enumerated below were also used in [18, 19] for decomposition of algebraic systems and in [20] for decomposition of PDEs. A system $S := \{P \cup Q\}$ of polynomials $P := \{p_1, \ldots, p_s\}$ for equations $(p_i = 0)$ and $Q := \{q_1, \ldots, q_t\}$ for inequations $(q_j \neq 0)$ is called *simple* if the following holds

1. $\mathrm{ld}(p_1) \prec \ldots \prec \mathrm{ld}(p_s)$, $\mathrm{ld}(q_1) \prec \ldots \prec \mathrm{ld}(q_t) \wedge \forall p \in P, q \in Q : \mathrm{ld}(p) \neq \mathrm{ld}(q)$;
2. $\forall r \in S, \forall \bar{\mathbf{x}} \in \mathrm{Zero}(P_{\prec r}/Q_{\prec r}) : \mathrm{init}(r)(\bar{\mathbf{x}}) \neq 0 \wedge r(\mathrm{ld}(r), \bar{\mathbf{x}})$ is *squarefree*;
3. elements in $P$ and $Q$ are *primitive* as polynomials in their leaders;

4. set $P$ is (algebraically) *autoreduced* and set $Q$ is *reduced* modulo $P$.

Here $\mathrm{Zero}(P/Q) := \{\bar{\mathbf{x}} \in \mathrm{Zero}(p = 0)\ \forall p \in P \wedge q(\bar{\mathbf{x}}) \neq 0,\ \forall q \in Q\}$ is the set of solutions in an algebraically closed extension of the ground field of characteristic zero and $F_{\prec r} := \{f \in F \mid \mathrm{ld}(f) \prec \mathrm{ld}(r)\}$.

In his book [17] Thomas designed an algorithm (cf. [18]) for decomposition of an algebraic polynomial system $H := \{h_1 = 0, \ldots, h_r = 0\}$ into a finite set of simple systems $\{P_i, Q_i\}$ such that

$$\mathrm{Zero}(H) = \cup_i \mathrm{Zero}(P_i/Q_i)\,. \tag{13}$$

He also described a procedure for completion of (algebraically) simple differential systems to involution based on his separation of variables (Thomas division [5]). Generally, such a completion violates simplicity of the system. In this case the system is further decomposed into simple subsystems and the completion procedure is further processed. Finally, the initial differential system is decomposed into a finite set of involutive and algebraically simple subsystems.

Motivated by efficiency of the above algorithm ***JanetBasis*** in its improved form [10], we suggest another completion algorithm[2] for systems decomposed into the simple subsystems. Our algorithm starts with the subsystem $\{P_j, Q_j\}$ in (13) whose highest ranking leader occurring in $P_j$ is minimal (w.r.t.$\succ$) among all the subsystems in the right-hand side of (13).

**Algorithm: *JanetCompletion*$(\{P, Q\}, \succ)$**

**choose** $f \in P$ with the lowest $\mathrm{ld}(f)$ w.r.t. $\succ$
$G := \{f\};\ T := P \setminus \{f\}$
**do**
  $h := 0$
  **while** $T \neq \emptyset$ and $h = 0$ **do**
    **choose** $p \in T$ with the lowest $\mathrm{ld}(p)$ w.r.t. $\succ$
    $T := T \setminus \{p\};\ h :=$***PREM***$_J(p, G)$
  **od**
  **if** $h \neq 0$ **then**
    **if** $\mathrm{ld}(h) = \mathrm{ld}(p)$ **then**
      $h :=$***Primitive***$(h)$
    **else**
      ***UpdateSimplicity***
      **for all** $g \in G$ such that $\mathrm{ld}(g) = \theta^\mu \circ \mathrm{ld}(h),\ |\mu| > 0$ **do**
        $T := T \cup \{g\};\ G := G \setminus \{g\}$
      **od**
    **fi**
    $G := G \cup \{h\};\ T := T \cup \{\, \theta^\beta \circ g \mid g \in G,\ \theta^\beta \in NM_J(g, G)\,\}$
  **fi**
**od while** $T \neq \emptyset$
**return** $\{P :=$***Autoreduce***$(G), Q :=$***Reduce***$(Q, P)\}$

[2] A detailed description of the algorithm will be given elsewhere.

Subalgorithm $\boldsymbol{PREM}_J$ of line 7 computes the *Janet pseudo-reminder* of $p$ w.r.t. $G$ by means of *pseudo-reductions*. The output $h$ of $\boldsymbol{PREM}_J(p, G)$ does not contain proper $J-$multiplicative derivatives of elements in $\{\mathrm{ld}(g) \mid g \in G\}$. This means that $h$ is *partially* $J-$*reduced w.r.t.* $G$ and is computed by performing the *pseudo-division* of $p$ by elements $g \in G$:

$$\mathrm{sep}(g)^j p = \sum_\alpha A_\alpha \theta^\alpha g + r\,,\ \alpha \in \mathbb{N}^n_{\geq 0},\ |\alpha| > 0,\ j \in \mathbb{N}_{\geq 0},\ \theta^\alpha \in J(g, G)\,,\ \forall \alpha\,,$$

where the *pseudo-reminder* $r$ is partially $J-$reduced w.r.t. $g$ and $\mathrm{sep}(g)$ denotes the *separant* of $g$. For $g$ written as a polynomial $g := \sum_i u_i\, \mathrm{ld}(g)^i$ in its leader (here $\forall i$ $u_i$ does not contain $\mathrm{ld}(q)$) $\mathrm{sep}(g) := \sum_i i\, u_i\, ld(g)^{i-1}$.
Subalgorithm $\boldsymbol{Primitive}(h)$ invoked in line 11 computes the *primitive part* of $h$ as a polynomial in its leader. Subalgoritm ***UpdateSimplicity*** invoked in line 13 verifies for $G$, enlarged with $h$, the simplicity conditions 2 and 3. If they are violated, then, in accordance with Thomas's algorithm [17], the subalgorithm updates system $\{G, Q\}$ to provide validity of the conditions. Generally, the update may split the system and thus affect the decomposition (13). The condition 1 of simplicity is fulfilled by the appropriate sorting whereas condition 4 for equations is provided by subalgorithms ***Autoreduce*** and ***Reduce*** that are called in line 21 and performs algebraic autoreduction and reduction, respectively, for the system constructed in the loop 3-20 whose $G$-part is involutive under the inequality conditions in $Q$.

After completion to involution of the given system, algorithm ***JanetCompletion*** is applied to the next (w.r.t. to the maximal ranking of its leaders) system in (13), and so on. As a result, the decomposition into the *regular* (cf. [20]), involutive and algebraically simple systems is obtained. This gives

$$\mathrm{dZero}(H) = \cup_i\, \mathrm{dZero}(P_i/Q_i),\ \mathrm{dZero}(P/Q) := \left\{ \begin{array}{c} \bar{\mathbf{x}} \mid p(\bar{\mathbf{x}}) = 0, q(\bar{\mathbf{x}}) \neq 0, \\ \forall p \in P,\, q \in Q \end{array} \right\},$$

where $\bar{\mathbf{x}}$ belongs to the universal differential field of characteristic zero.

**Acknowledgements.** The author thanks W.Plesken for useful remarks on Thomas's algorithm. This research was supported in part by grants 05-02-17645, 07-01-00660 from the Russian Foundation for Basic Research and by grant 5362.2006.2 from the Ministry of Education and Science of the Russian Federation.

# References

[1] Riquier C.: *Les Systèmes d'Equations aux Dérivées Partielles*, Gauthier-Villars, Paris, 1910.

[2] Janet M.: *Leçons sur les Systèmes d'Equations aux Dérivées Partielles*, Cahiers Scientifiques, IV, Gauthier-Villars, Paris, 1929.

[3] Pommaret J.F.: *Partial Differential Equations and Group Theory. New Perspectives for Applications*, Kluwer, Dordrecht, 1994.

[4] Gerdt V.P.: *Completion of Linear Differential Systems to Involution*. Computer Algebra in Scientific Computing / CASC 1999, Springer-Verlag, Berlin, 1999, pp. 115–137.

[5] Gerdt V.P. and Blinkov Yu. A.: *Involutive Bases of Polynomial Ideals*, Mathematics and Computers in Simulation **45** (1998) 519-542; *Minimal Involutive Bases*, ibid., 543–560.

[6] Seiler W.M.: *A Combinatorial Approach to Involution and Delta-Regularity I: Involutive Bases in Polynomial Algebras of Solvable Type; II: Structure Analysis of Polynomial Modules with Pommaret Bases*, Preprint, Universität Mannheim, 2002.

[7] Kondratieva M.V., Levin A.B., Mikhalev A.V. and Pankratiev E.V.: *Differential and Difference Dimension Polynomials*, Kluwer, 1999.

[8] Gerdt V.P., Blinkov Yu.A. and Mozzhilkin V.V.: *Gröbner bases and generation of difference schemes for partial differential equations*, Symmetry, Integrability and Geometry: Methods and Applications, 2:26, 2006.

[9] Gerdt V.P.: *Gröbner Bases Applied to Systems of Linear Difference Equations*, Physics of Particles and Nuclei, Letters, to appear. arXiv:cs.SC/0611041

[10] Gerdt V.P.: *Involutive Algorithms for Computing Gröbner Bases*, Computational Commutative and Non-Commutative Algebraic Geometry. NATO Science Series, IOS Press, 2005, pp. 199–225.

[11] Becker T., Weispfenning V. and Kredel H.: *Gröbner Bases. A Computational Approach to Commutative Algebra*, Graduate Texts in Mathematics **141**, Springer-Verlag, New York, 1993.

[12] Buchberger B. and Winkler F. (Eds.): *Gröbner Bases and Applications*, Cambridge University Press, 1998.

[13] Wen-tsün Wu: *On the Construction of Gröbner Basis of a Polynomial Ideal Based on Riquier-Janet Theory*. System Science and Mathematical Sciences, 4/3, 1991, 193–207.

[14] Blinkov Yu.A., Gerdt V.P., Cid C.F., Plesken W. and Robertz D.: *The Maple Package "Janet": I. Polynomial Systems*. Computer Algebra in Scientific Computing / CASC 2003, Institute of Informatics, Technical University of Munich, Garching, 2003, pp.31–40.; *II. Linear Partial Differential Equations*, ibid., pp.41–54.

[15] Gerdt V.P. and D.Robertz D.: *A Maple package for computing Gröbner bases for linear recurrence relations*. Nuclear Instruments and Methods in Physics Research, A559, 215–219, 2006.

[16] Gallo G., Mishra B. and Ollivier F.: *Some Constructions in Rings of Differential Polynomials*, Lecture Notes in Computer Science, Vol. 539, 1991, 171–182.

[17] Thomas J.F.: *Differential Systems*, AMS Publication, New York, 1937.

[18] Wang D.: *Elimination Methods*, Springer, Wien, 2000.

[19] Wang D.: *Elimination Theory, Methods and Practice*, Mathematics and Mathematics-Mechanization, Shandong Education Publishing House, Jinan, 2001, pp. 91–137.

[20] Li Z. and Wang D.: *Coherent, Regular and Simple Systems in Zero Decompositions of Partial Diffrential Systems*, System Science and Mathematical Sciences 12 (Suppl.), 1999, 43–60.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# *Mathematica* in Word Pattern Avoidance Research

Veikko Keränen

*Rovaniemi University of Applied Sciences (RAMK),*
*School of Technology*
*Jokiväylä 11, 96300 Rovaniemi, Finland*
veikko.keranen@ramk.fi, http://south.rotol.ramk.fi

**Abstract.** We discuss our long-lasting extensive research from 1990 to the present on abelian pattern avoidance in words and the use of *Mathematica* as a programming tool and computing environment which has been crucial in interactive development of the code, in visualizations, and in extensive distributed computations. *Mathematica* has enabled us to discover phenomena that otherwise would have been inaccessible or would have been regarded as unbelievable.

Our recent research findings include a powerful abelian square-free substitution over four letters. The key characteristics of the new substitution derive from delicate mutations in the image words that merely emerged from the computations. We do not think it would have been feasible to create them by any design. Another quite recent finding concerns unfavourable factors that can be used to explain, at least partly, both the highly nonlinear behavior in our earlier computations and the extreme difficulty that has dominated the search for abelian square-free endomorphisms and substitutions over four letters.

This paper contains a number of visualizations of the structures and processes.

## 1 Introduction

The systematic study of word structures, i.e., combinatorics on words, was started by Axel Thue (1863−1922) in [37] at the beginning of the 20th century. One of his discoveries was that consecutive repetitions of non-empty factors (squares) can be avoided in infinite words over a three-letter alphabet. As a simple example of the square concept, consider the words $abacaba$ and $ab\,\underline{cd}\,\underline{cd}\,ab$. The first word does not contain any square, i.e., it is square-free, whereas the second word contains the underlined square $\underline{cd}\,\underline{cd}$ as a factor.

The above-mentioned square-freeness property of words is not trivial to prove. The tool which Thue invented for constructing square-free, and other repetition-free, words, namely the concept of a repetition-free morphism, is still a basic technique in the study

of avoidable patterns in words. Repetition-free morphisms are mappings between free monoids that preserve the repetition-freeness of words. The iteration of a non-trivial repetition-free endomorphism or substitution (that maps a letter to more than one word) produces repetition-free words of any length. Dealing with substitutions somewhat later, we point out that repetition-free morphisms have been sharply characterized in [5, 10, 19, 20, 27, 28, 34, 38]. The results therein concern different types of repetitions (*k*-repetitions for a given integer $k \geq 2$) and alphabet sizes. Informally speaking, most of the characterizations mean that it is possible of test the repetition-freeness of a given morphism just by checking whether the image words of short repetition-free words are also repetition-free. A general survey of these and related results, achieved before 1984, is given in [3]. For a short survey of Thue's results concerning repetition-free words and their applications, see [16]. Fundamental topics are discussed in [29, 36].

In a paper from 1961, see [15, p. 240], Paul Erdös (1913−1996) raised the question whether abelian squares can be avoided in infinitely long words, i.e., whether there exist infinitely many abelian square-free words over a given alphabet. Here, an abelian square means a non-empty word $uv$, where $u$ and $v$ are permutations (anagrams) of each other. For example, $\underline{abc}\,\underline{acb}$ is an abelian square. A word is called abelian square-free, if it does not contain any abelian square as a factor. For example, the word $abacaba$ is abelian square-free, while $ab\,\underline{cabdc}\,\underline{bcacd}\,ac$ is not.

Later, in a 1970 paper, Pleasants [33] showed that there exists an infinite abelian square-free word over five letters. Finally, in 1991, see Keränen [21] the year after, we managed to show that the same holds true also in the case of four letters. It is easily seen that abelian squares cannot be avoided over a three-letter alphabet. Indeed, in this alphabet, each word of length 8 contains an abelian square. In [14] Entringer et al. showed that every infinite word over a binary alphabet contains arbitrarily long abelian squares. Dekking [12] in turn proved that abelian repetitions to the fourth power can be avoided in infinite words over two letters, and abelian repetitions to the third power (cubes) can be avoided in infinite words over three letters. For a generalization of abelian squares, see Avgustinovich and Frid [2]. Abelian fractional powers were studied by Cassaigne and Currie [8]. In [11], Currie showed that the number of binary words avoiding abelian fourth powers grows exponentially, and in [1], Aberkane, Currie, and Rampersad showed that the number of ternary words avoiding abelian cubes grows exponentially as well.

An application of Dekking's result was given by Justin et al. in [17], where it was shown that a finitely-generated semigroup is uniformly repetitive if and only if it is finite. In [32], Pirillo et al. used similar reasoning when proving, among other results, that the additive semigroup $\mathbb{N}^+$ is not uniformly 4-repetitive. It seems to be an open problem whether $\mathbb{N}^+$ is uniformly 2-repetitive or 3-repetitive. In all these considerations the use of van der Waerden's theorem has been very central. In Lothaire [29, pp. 55−62] van der Waerden's theorem was used to show that every morphism from a free semigroup $A^+$, where $A$ is finite, to $\mathbb{N}^+$ is repetitive. This means that every long enough sequence on a finite set of integers contains two adjacent segments (not necessarily of the same length) that have the same sum.

The original problem of abelian squares has also attracted attention in the study of free partially-commutative monoids, see for instance [9, 13]. Moreover, abelian square-free words have aroused interest in algorithmic music (Laakso [26]) and quite

recently in cryptography (Rivest [35], Bouillaguet et al. [4]).

In 1993, Carpi [5] gave sufficient conditions for morphisms to preserve abelian *k*th power-freeness of words. A conjecture is that these conditions yield an effective characterization also for abelian square-free endomorphisms on a four-letter alphabet $\Sigma_4 = \{a, b, c, d\}$. However, new examples of relatively short abelian square-free endomorphisms $g$ of ${\Sigma_4}^*$ have turned out to be extremely hard to find – and the same difficulty applies to every systematic attempt for constructing long abelian square-free words over 4 letters. Before our current findings, we were not at all optimistic that it would be possible to find more examples of abelian square-free endomorphisms – not to speak of proper substitutions of ${\Sigma_4}^*$. Thus far, since 1992 when we presented $g_{85}$ in [21], the only new abelian square-free endomorphisms and substitutions have been found by Carpi [6], cf. also [7, pp. 80−81]. However, his mappings are all based on the structure of $g_{85}$. Moreover, the size of these endomorphisms and substitutions are large. By using these substitutions, Carpi showed that the number of abelian square-free words of each length grows exponentially, and that the monoid of (uniform) abelian square-free endomorphisms of ${\Sigma_4}^*$ is not finitely generated.

Very recently, we succeeded in finding 200 new abelian square-free endomorphisms and some of them work as a starting point for a new powerful abelian square-free substitution. These endomorphisms have the property that the image words $g(x)$, $x \in \Sigma_4$ are all obtained by cyclically permutating the letters in $g(a)$. The image words $g(a)$ can be viewed and copied from the Internet [18]. The same cyclic permutation property is true for $g_{85}$ as well, and this method was already used by Pleasants [33] in connection with five letters. Consequently, all of these endomorphisms have a uniform modulus and the generated words grow uniformly. The size of Pleasants' endomorphism is $5 \times 15 = 75$. In our case, we have checked using computers that the size $4 \times 85 = 340$ for $g_{85}$, in spite of its largeness, is actually minimal, at least as far as cyclic permutation method is used. So far, the search for other kinds of abelian square-free endomorphisms of $\Sigma_4$* has not been very successful, even with extensive experimentation. However, we show later in (1) that 20724 (out of 20736) abelian square-free endomorphisms indeed possess a different structure.

Moreover, in 2002 we [22, 23] found a nice endomorphism $g_{98}$ of ${\Sigma_4}^*$ that can be used in iterations, and together with $g_{85}$ to produce infinite abelian square-free DT0L-languages (i.e., languages obtained by using compositions of morphims). This $g_{98}$ itself is not an abelian square-free endomorphism, as it does not preserve abelian square-freeness for all words (starting with already from the length 7). The structure of the image words of $g_{98}$ also partly differs from that of the other above-mentioned 201 (remembering also $g_{85}$) cyclic endomorphisms.

Quite recently, we have also gained insight as to why these abelian square-free structures are so rare. In [24] we explain, at least partly, this rareness of long words avoiding abelian squares by using the concept of unfavourable factor. We take an abelian square-free word and, using *Mathematica*, try to extend it in abelian square-free fashion to the right and to the left in all possible ways up to a given upper bound for the total length. At a time, increasing the length of the word by a given fixed length at each step. We extend alternately to right and left, and backtrack if necessary. If the given upper bounds are reached then the original word is a *so-far-favourable* one (it may still turn out to be unfavourable on later experiments). If there is no way to

reach the upper bounds, then the original word is classified, without any doubt, to be as unfavourable. Thus we obtain three kinds of words: *unfavourable* (*bad*), *so-far-favourable* (*so-far-so-good*), and *favourable* (*good*).

It is a remarkable phenomenon that sometimes relatively short so-far-favourable words turn out to be unfavourable factors after being 'safely' extendable (to the right and left) for quite a long distance and with a really huge number of branches. One might have expected the quite long buffers to guarantee the further growth. We suspect that the majority of abelian square-free words over four letters cannot occur as proper factors in the middle of very long (infinite) abelian square-free words. In a way, the experimental facts concerning unfavourable factors explain the highly non-linear behavior of our earlier computations and also the difficulty of finding abelian square-free endomorphisms of $\Sigma_4{}^*$ (not to speak of substitutions). At present we know that in the four letter case about 60 % of the abelian square-free words of length 24 are indeed unfavourable.

It will be interesting to study, in a similar way, the case of three letters, for which an exciting open problem was posed by Mäkelä [30], who allows repetitions $xx$ and $xxx$, for a letter $x$, but no other abelian squares (or cubes).

# 2 Preliminaries for combinatorics on words

In this section we present notation and terminology. Our terminology is quite standard in the field of combinatorics on words. Consequently, the reader may consult this section later, as needed.

An *alphabet* $\Sigma$ is a finite non-empty set of abstract symbols called *letters*. A *word* (*string*) over $\Sigma$ is a finite (unless otherwise indicated) string, or sequence, of letters belonging to $\Sigma$. The set of all words over $\Sigma$ is denoted by $\Sigma^*$, while the set of non-empty words is denoted by $\Sigma^+$. For words $u$ and $v$ in $\Sigma^*$, the associative binary operation of *catenation* is defined as the juxtaposition $uv$. The *empty word*, which is the neutral element of catenation, is denoted by $\lambda$. The algebraic structures $\Sigma^*$ and $\Sigma^+$ are called, respectively, the free monoid and the free semigroup generated by $\Sigma$.

Let $w = x_1 \cdots x_m$, $x_i \in \Sigma$. The *length* of the word $w$, denoted by $|w|$, is the number of occurrences of letters in $w$, i.e., $|w| = m$. Let $\Sigma = \{a_1, \ldots, a_n\}$. The number of occurrences of one letter $x \in \Sigma$ in the word $w$ is denoted by $|w|_x$, or simply by $|w|_i$ if *x* = $a_i$. The notation $\psi_\Sigma$(*w*) stands for the *Parikh vector* of $w$, i.e., $\psi_\Sigma$(*w*) = ($|w|_1$, ... , $|w|_n$). Usually we will omit the subscript $\Sigma$ and write simply $\psi$ instead of $\psi_\Sigma$.

A word $u$ is called a *factor* of a word $w$, if $w = p\,u\,s$ for some words $p$ and $s$. The notation FACT($w$) stands for the set of all factors of $w$. If $p$ (or $s$) $= \lambda$, then $u$ is called a *prefix* (or a *suffix*) of $w$.

Let $k \geq 2$ be a given integer. A *k-repetition* is a non-empty word of the form $R^k$. An abelian k-repetition is a non-empty word of the form $P_1 \cdots P_k$, where $\psi(P_\mu) = \psi(P_\nu)$ for all $1 \leq \mu < \nu \leq k$, i.e., $P_i$:s are *commutatively equivalent*, that is, they are permutations, or anagrams, of each other. Instead of [abelian] 2- and 3-repetitions, the terms [*abelian*] *squares* and *cubes* are often used. A word is called *k-repetition free*, or *k-free* for short, if it does not contain any *k*-repetition as a factor. A word sequence or a word set is *k-free*, if all words in it are *k*-free. Abelian analogs of these terms

and definitions also exist and are formed in a natural way by preceding any term with the word abelian, i.e., abelian square, abelian cube, abelian *k*-repetition free, etc. The abelian analog of the short term, *k*-free is a-*k*-free. If, for a fixed *k*, it is possible to construct arbitrarily long (infinite) a-*k*-free (or other pattern-free) words over a given alphabet $\Sigma$, then we say that abelian *k*-repetitions (or those patterns) are *avoidable* over $\Sigma$.

A *morphism* $h$ is a mapping between free monoids $\Sigma^*$ and $\Delta^*$ with $h(uv) = h(u)h(v)$ for every $u$ and $v$ in $\Sigma^*$. In particular, $h(\lambda) = \lambda$. A morphism $h : \Sigma^* \to \Delta^*$, being compatible with the catenation of words, is uniquely defined, if the word $h(x) \in \Delta^*$ is (effectively) given for each $x \in \Sigma$. If $\Delta = \Sigma$, we call $h$ an *endomorphism* (and usually write $g$ instead of $h$). For a morphism $h$ and a language $L$ we define $h(L) = \{h(w)|w \in L\}$. A morphism $h$ is called *uniformly growing*, or is said to have a *uniform modulus*, if $|h(x)| = |h(y)| \geq 2$ for every $x$ and $y \in \Sigma$.

A *substitution* $\sigma : \Sigma^* \to 2^{\Delta^*}$ is a monoid morphism of $\Sigma^*$ into a subset monoid of $\Delta^*$. The substitution $\sigma$ can be regarded as a multi-valued mapping between the free monoids $\Sigma^*$ and $\Delta^*$, and written $\sigma : \Sigma^* \to \Delta^*$. The substitution $\sigma$ is finite if $\sigma(\Sigma)$ is a finite subset of $\Delta^*$. Obviously, for a morphism $h : \Sigma^* \to \Delta^*$, it holds that Card($h(\Sigma)$) $\leq$ Card($\Sigma$), and thus a morphism is a special case of a finite substitution. Following the terminology of Carpi [7], a substitution $\sigma : \Sigma^* \to \Delta^*$ is called *commutatively functional*, if dom($\sigma$) = $\Sigma^*$, and, for all $x \in \Sigma$, $v' \in \sigma(x)$, it holds that $\psi(v) = \psi(v')$ (this is also written as $v \sim v'$). In other words, a substitution is termed commutatively functional, if the image words of a fixed letter are all commutatively equivalent. Moreover, for a commutatively functional substitution $\sigma$ and any word $w$ in $\Sigma^*$, all the words $\sigma(w)$ are commutatively equivalent.

For a given integer $k \geq 2$, a substitution (or a morphism) $\sigma : \Sigma^* \to \Delta^*$ is called *k-free* [*a-k-free*], if all the words $\sigma(w)$ are (or the word $\sigma(w)$ is) *k*-free [a-*k*-free] for every *k*-free [a-*k*-free] word $w \in \Sigma^*$.

# 3 The new a-2-free substitution $\sigma_{109}$ over 4 letters

Let $\Sigma_4 = \{a, b, c, d\}$. Define the substitution $\sigma_{109}$: ${\Sigma_4}^* \to {\Sigma_4}^*$ as follows. First let the 12 image words of $\sigma_{109}(a)$, say $\{A_1, A_2, \ldots, A_{12}\}$, have the form

$A_i = p_{16}\ w_4\ u_{27}\ w_3\ s_{59} = abcacdcbcdcadcdb\ w_4\ badacdadbdcdbdabdbcbabcbdcb\ w_3\ bdcdadcdbcbabcbdcbcacdcacbadabcbdcbcadbabcbabdbcdbdadbdcbca,$

with 12 different factor pairs ($w_4, w_3$), taken in the natural lexicographical order form $\{abcd, abdc, adbc, dabc\} \times \{acd, adc, cad\}$. The subscripts of the factors $p_{16}, w_4, u_{27}, w_3, s_{59}$ indicate their lengths. Note that all the words in $\{abcd, abdc, adbc, dabc\}$, and respectively in $\{acd, adc, cad\}$, are commutatively equivalent. The delicate mutations in words of $\sigma_{109}(a)$ can also be described and visualized as a movement of letters $d$ and $c$ in the invariant background:

____ **d**abc _____ **c**ad _____ , ____ ab**d**c _____ **c**ad _____ ,
____ **d**abc _____ a**c**d _____ , ____ ab**d**c _____ a**c**d _____ ,
____ **d**abc _____ ad**c** _____ , ____ ab**d**c _____ ad**c** _____ ,
____ a**d**bc _____ **c**ad _____ , ____ abc**d** _____ **c**ad _____ ,
____ a**d**bc _____ a**c**d _____ , ____ abc**d** _____ a**c**d _____ ,
____ a**d**bc _____ ad**c** _____ , ____ abc**d** _____ ad**c** _____ .

To complete the definition of $\sigma_{109}$, let $\sigma_{109}(\phi(x)) = \phi(\sigma_{109}(x))$ for all $x \in \{a, b, c, d\}$, where $\phi : {\Sigma_4}^* \to {\Sigma_4}^*$ is the circular letter-to-letter endomorphism defined by $\phi(a) = b$, $\phi(b) = c$, $\phi(c) = d$, $\phi(d) = a$. Thus, informally, the set of image words for $b$, $c$, $d$ are obtained, letter-by-letter, by cyclic permutation of letters of all the words in $\{A_1, A_2, \ldots, A_{12}\}$. Obviously, $\sigma_{109}$ is a commutatively functional substitution of ${\Sigma_4}^*$ (of uniform modulus 109). The Parikh vectors for the image words of letters are the rows of the matrix below:

$$\begin{pmatrix} \psi(A) \\ \psi(B) \\ \psi(C) \\ \psi(D) \end{pmatrix} = \begin{pmatrix} 21 & 31 & 29 & 28 \\ 28 & 21 & 31 & 29 \\ 29 & 28 & 21 & 31 \\ 31 & 29 & 28 & 21 \end{pmatrix},$$

whenever

$$A \in \sigma_{109}(a), B \in \sigma_{109}(b), C \in \sigma_{109}(c), D \in \sigma_{109}(d).$$

Using a computer, we checked the a-2-freeness of $\sigma_{109}$ in two (albeit not completely) different ways. The first way was a direct but long method similar to what we used previously in the paper [21] in 1992. There, the code development was done in LISP. In the present work, we used *Mathematica* to make most of the computational steps visible, thus providing a way to recheck the result. In these computations we benefitted greatly from *Mathematica*'s dynamic programming feature, which guarantees that functions remember the values they have found. The second method that we used for checking the a-2-freeness of $\sigma_{109}$, is an application of Carpi's [7] characterization. The details of that method are explained in [25] and lie outside the main topic of this paper, though, we may mention, in passing, that in this case as well, it was natural to develop the algorithms by using *Mathematica*.

In connection with the substitution $\sigma_{109}$, let us consider the $12^4$ = 20736 different endomorphisms $g_{109,ijkl}$ of ${\Sigma_4}^*$, defined by

$$g_{109,ijkl}(a) = A_i, \quad g_{109,ijkl}(b) = B_j = \phi(A_j), \quad g_{109,ijkl}(c) = C_k = \phi(B_k),$$
$$g_{109,ijkl}(d) = D_l = \phi(C_l), \quad i, j, k, l = 1, \ldots, 12. \tag{1}$$

Our checking of eachshow that all $12^4$ endomorphisms are indeed abelian square-free. This alone suggests that the substitution $\sigma_{109}$ might really be a-2-free. Although additional tests are needed, they pose no particular difficulty. Indeed, in [25] we justify the following conclusion:

**Proposition 1** *The substitution $\sigma_{109} : {\Sigma_4}^* \to {\Sigma_4}^*$ defined above is abelian square-free.*

Of course, the computational details should also be carried out independently of us. We hope that in the very near future people will accomplish this.

It is likely that new abelian square-free substitutions of ${\Sigma_4}^*$ can be constructed not only from $g_{109,ijkl}$, but also from other a-2-free endomorphisms that we have recently found. The image word $g(a)$ for all 200 a-2-free endomorphisms, $g_{85}$ (found in 1990), and $g_{98}$ (found in 2002), can be viewed and copied from the Internet [18].

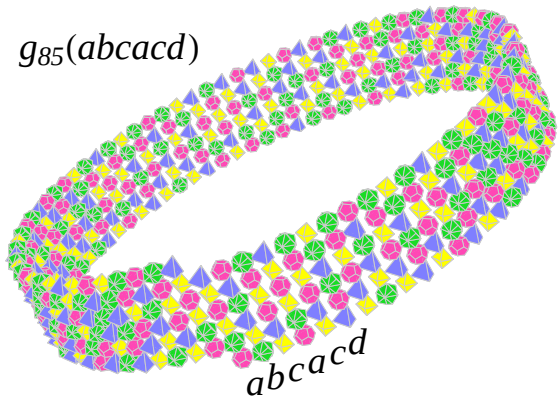



Figure 1: Six starting image words of a self-reading string for $g_{85}$.

The properties of $\sigma_{109}$ lead to a considerably sharper lower bound for the exponential growth of $c_n$, i.e., of the number of a-2-free words over 4 letters of length *n*. We find that $c_n > \beta^{-50}\beta^n$ with $\beta = 12^{1/m} \simeq 1.02306$. For details the reader is referred to [25]. The exponential growth of $c_n$ was first proved by Carpi [6], who showed that $c_n > \beta^{-t}\beta^n$ with $\beta = 2^{19/t} = 2^{19/(85^3-85)} \simeq 1.000021$, where $t = 85^3 - 85$ is the modulus of his substitution constructed from $g_{85}$ that we presented in [21].

The number of all a-2-free words over 4 letters up to the length 60 can be found on the Internet [18].

# 4 Visualizations of structures and processes

Some of the visualizations presented in this section can also be found at our web pages. All of them are created using *Mathematica*. In many cases, the *Mathematica* graphics is imported into interactive tools, such as the LiveGraphics3D Java applet or applications developed by C++.

The pictorial representation of Figure 1 contains the six image words of the word *abcacd* related to the iteration of the abelian square-free endomorphism $g_{85}$. In Figure 2, the visualization can be used to detect structures separately at even and at odd positions. In Figure 3, one finds the first twenty image words for the self-reading sequence associated with the endomorphism $g_{98}$ of $\Sigma_4{}^*$. Note the pink diagonal consisting of occurrences of the letter *a*. In Figure 4, the word $(g_{98})^2(a)$ is represented by using only the letters in {*b*, *d*} and leaving the occurrences of *a* and *b* white. In Figure 5, the directions for letters are indicated in a quad tree representation. Quad trees can be used to visualise large sets of words over 4 letters at a glance. In Figure 6, one sees how longer words can be represented by dividing the squares further. All abelian square-

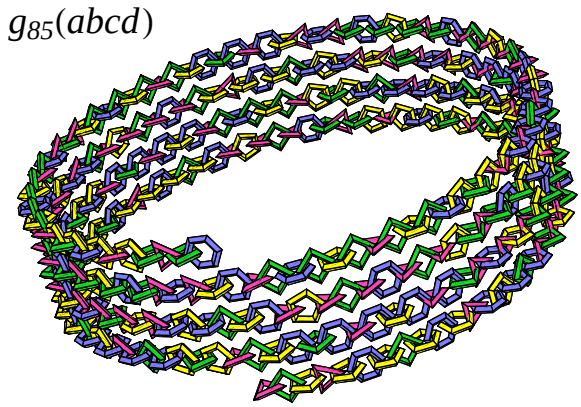



Figure 2: Even and odd positions separated.

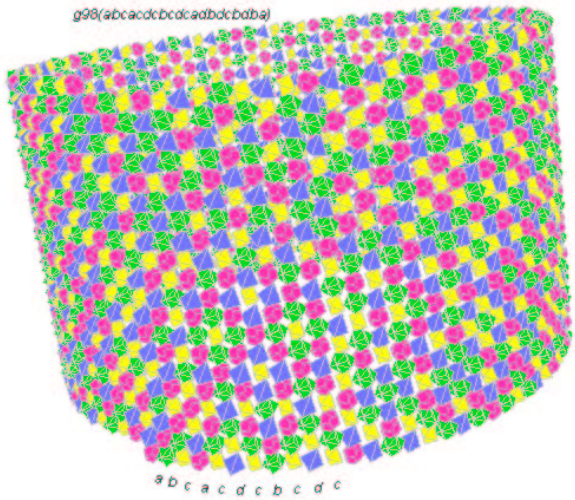



Figure 3: Twenty starting image words for $g_{98}$. Note the pink diagonal.

free words of length 2 over 4 letters are shown. Note the white corners representing the unfavourable words *aa*, *bb*, *cc*, *dd*. In Figure 7, the guad tree shows all the 3576 abelian square-free words of length 8 over 4 letters in one picture. In the Introduction we explained the concept of *unfavourable* (*bad*), *so-far-favourable* (*so-far-so-good*), and *favourable* (*good*) factors, and noted that it is a remarkable phenomenon that relatively short so-far-favourable words turn out to be unfavourable factors even being 'safely' extendable (to the right and left) for quite a long distance and sometimes with a really huge number of branches.

Most surprising in this respect is the behavior of *abcdacbabdabacdacbcdad*. For this word of length 22, we obtain a list of pairs (*x*, *y*), where *x* represents the length of the words in the so-far-favourable bi-directional tree, and *y* represents the number of all possible extensions of the length in question. The words, in the (somewhat abbreviated) list below, are extended by one letter at a time only, that is, all extensions of length 1 are tried.

{{22, 1}, {23, 2}, {24, 2}, {25, 5}, {26, 14}, {27, 23}, {28, 14}, {29, 26}, {30, 10}, {31, 16}, {32, 8}, {33, 9}, {34, 9}, {35, 16}, {36, 16}, {37, 27}, {38, 27}, {39, 54}, {40, 54}, {41, 68}, {42, 136}, {43, 194},{44, 291}, {45, 444}, {46, 296}, {47, 450}, {48, 225}, {49, 331}, {50, 331}, {51, 474}, {52, 948},..., {107, 840479}, {108, 1679287}, {109, 2301836}, {110, 2302465}, {111, 3157227}, {112, 3154210}, {113, 4306159}, {114, 8466798}, {115, 11575001}, {116, 5779271}, {117, 7866918}, {118, 0}}.

The death of all of the nearly 8 million branches of this bi-directional tree at length 117 looks dramatic. We remark that it was necessary to construct all the possible a-2-free words in the tree to be able to find the numbers and see the collapse. Consequently, this computation is quite a big, albeit a straightforward, one. Of course a much more massive search was needed to find this example in the first place. In the search for unfavourable factors, we have been using many features of *Mathematica*, including conversions from strings to symbols and to patterns, and further to cumulative integer lists. This part of the code uses a state-machine paradigm and the overall structure is actually quite complex. We do not think it would have been feasible for us to have developped the code and all the necessary pre-computational structures without the aid of *Mathematica*'s technical computing environment. In Figure 8, a part of the complex behavior of the above list for the unfavourable factor *abcdacbabdabacdacbcdad* is represented. In this Figure, we zoom into the middle behavior. The final peak of the list is omitted for clarity. Figure 9 is a visualization of $g_{85}(a)$ in the form of DNA. It would be desirable, if these kind of loops (hypohelix structures) were avoided in real DNA, since they may lead to diseases, see for example Mirkin [31]. It seems to be an open question to what extent the loops can be avoided over four letters (provided that the structure is not too trivial). This figure was designed by Erik Jensen, University of California, Berkeley, in 2005.

In Figures 10, 11, and 12, we represent 4-letter walks in the diamond lattice. The walks are obtained by first saving *Mathematica* graphics in a file and, then using Martin Kraus's LiveGraphics3D Java applet. The direction vectors for the letters *a*, *b*, *c*, *d* are shown in Figure 10. In Figure 11, the loops inside the walk represent factors containing an equal number of occurrences of each letter. The example word is the prefix *abc acdcbcdcadcdbdabacabadbabcbd bcb* of $g_{85}(a)$.

Our last image, Figure 12, is obtained from interactive experiments with LiveGraphics3D. Rotating the figure of $g_{98}(a)$ and looking at it from the direction pointed

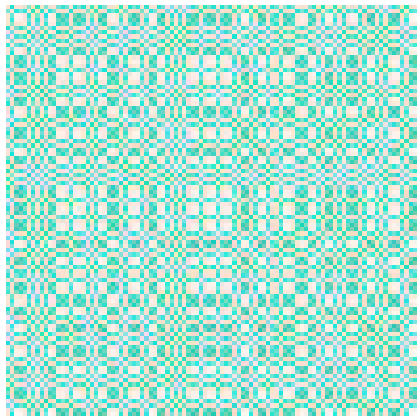


Figure 4: Occurrences of *b* and *d* in the word $(g_{98})^2(a)$.

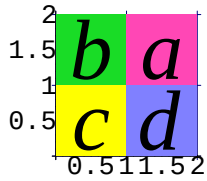



Figure 5: Directions for letters in a quad tree.

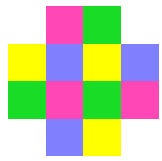


Figure 6: All abelian square-free words of length 2 over 4 letters.

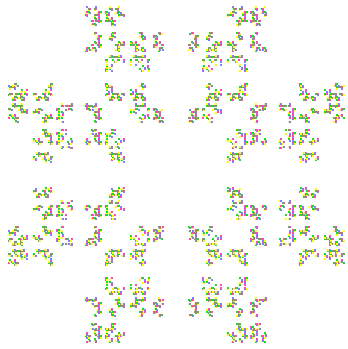


Figure 7: All abelian square-free words of length 8 over 4 letters.

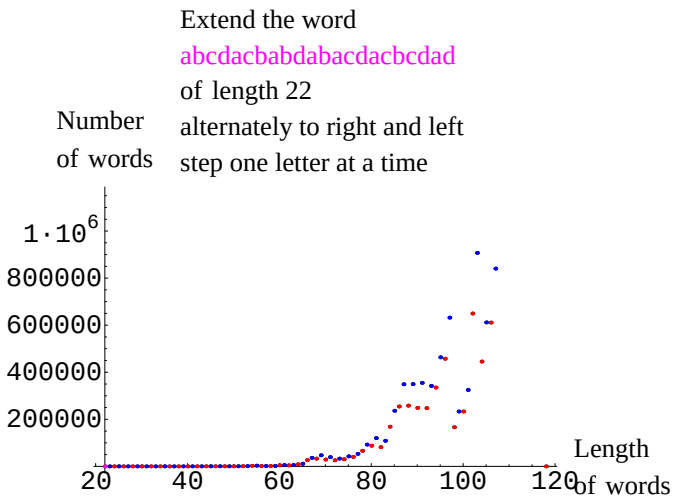



Figure 8: Unfavourable behavior of *abcdacbabdabacdacbcdad* when extended by one letter at a time.

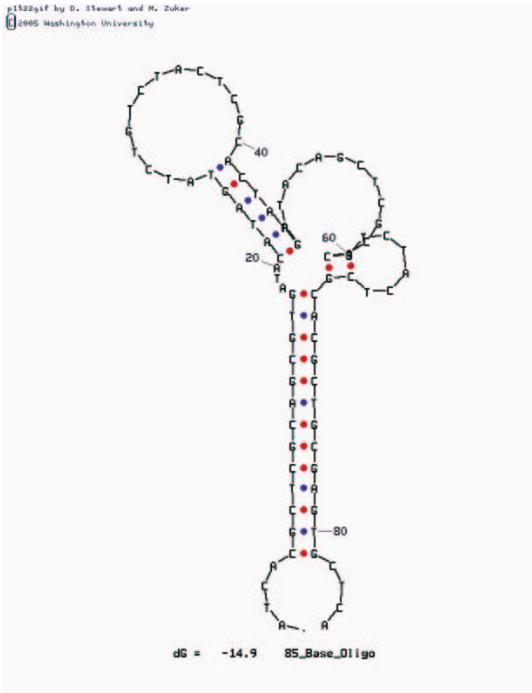



Figure 9: Hypohelix structures of $g_{85}(a)$.

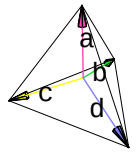



Figure 10: Direction vectors for letters *a*, *b*, *c*, *d*.

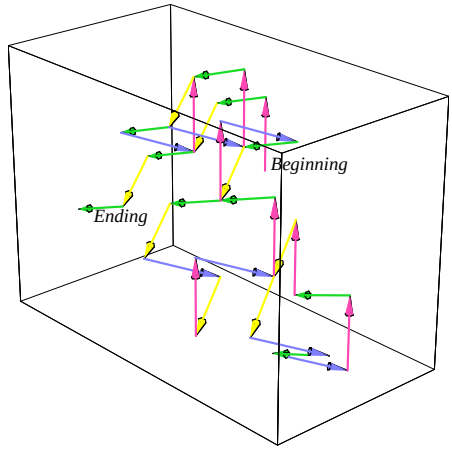



Figure 11: Loops represent factors containing an equal number of occurrences of each letter.

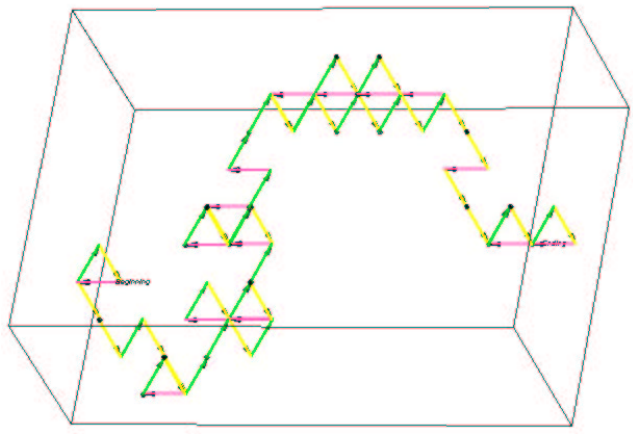


Figure 12: Semipalindrome structure of *dcdadbdcbdbabdbcbacbcdbabdc d bdcacdbcbacbcdcacdcbdcdadbd* inside $g_{98}(a)$.

by the vector for *d*, one suddenly detects a semipalindrome structure of a long factor *dcdadbdcbdbabdbcbacbcdbabdc d bdcacdbcbacbcdcacdcbdcdadbd* inside this word.

**Acknowledgements.** We gratefully acknowledge the participation of several of our students over the course of our study from 1990 to 2007. They were responsible for coding a number of computer programs for searching strings with desirable properties, and set up and worked with the distributed computing environments at RAMK. The participating students, with their starting year given in parentheses are: Kari Tuovinen (1990); Minna Iivonen, Anja Keskinarkaus, Marko Manninen (1993); Abdeljalil Chabani, Tomi Laakso (1994); Mika Moilanen, Juha Särestöniemi (1996); Juho Alfthan (1999); Olli-Pentti Saira (2000); Marja Kenttä, Ville Mattila (2001); Lauri Autio, Marianna Mölläri (2002); Antti Eskola (2003); Antti Karhu, Veli-Matti Lahtela, Olli-Pekka Siivola (2004); Esa Nyrhinen, Sami Vuolli (2005); Esa Taskila, Mikhail Kalkov, Antti Oja, Viet Pham Hoang (2006); Alena Mekhnina, Shijing Zhang, Jing Lin, and Irina Sekushina (2007).

For the most recent findings, we have used the code written by Kari Tuovinen, Ville Mattila, Mikhail Kalkov, and Viet Pham Hoang. Viet Pham Hoang was also responsible for building the grid and for the production runs executed from it.

# References

[1] A. Aberkane, J.D. Currie, and N. Rampersad. The number of ternary words avoiding abelian cubes grows exponentially. *J. of Integer Sequences*, 7, 2004. Article 04.2.7, 13 pages.

[2] S.V. Avgustinovich and A.E. Frid. Words avoiding abelian inclusions. *J. of Aut., Lang. and Combin.*, 7:3–9, 2002.

[3] J. Berstel. Some recent results on square-free words. In M. Fontet and K. Melhorn, editors, *Proc. STACS '84, Lecture Notes in Comp. Sci.*, volume 166, pages 14–25. Springer-Verlag, Berlin, 1984.

[4] C. Bouillaguet, P.A. Fouque, A. Shamir, and S. Zimmer. Second preimage attacks on dithered hash functions. Submitted to EUROCRYPT '08.

[5] A. Carpi. On abelian power-free morphisms. *Int. J. of Algebra and Comp.*, 3:151–167, 1993.

[6] A. Carpi. On the number of abelian square-free words on four letters. *Discrete Appl. Math.*, 81:155–167, 1998.

[7] A. Carpi. On abelian squares and substitutions. *Theor. Comp. Sci.*, 218:61–81, 1999.

[8] J. Cassaigne and J.D. Currie. Words strongly avoiding fractional powers. *Europ. J. Combinatorics*, 20:725–737, 1999.

[9] R. Cori and M.R. Formisano. Partially abelian square-free words. *RAIRO Inform. Théor. et Appl.*, 24:509–520, 1990.

[10] M. Crochemore. Sharp characterizations of square-free morphisms. *Theoret. Comp. Sci.*, 18:221–226, 1982.

[11] J.D. Currie. The number of binary words avoiding abelian fourth powers grows exponentially. *Theoret. Comp. Sci.*, 319:441–446, 2004.

[12] F.M. Dekking. Strongly non-repetitive sequences and progression-free sets. *J. Combin. Theory Ser. A*, 27:181–185, 1979.

[13] V. Dickert. Research topics in the theory of free partially commutative monoids. *Bull. Europ. Assoc. Theoret. Comp. Sci.*, 40:479–491, 1990.

[14] R.C. Entringer, D.E. Jackson, and J.A. Schatz. On non-repetitive sequences. *J. Combin. Theory Ser. A*, 16:159–164, 1974.

[15] P. Erdös. Some unsolved problems. *Magyar Tud. Kutató Int. Közl.*, 6:221–254, 1961.

[16] G.A. Hedlundd. Remarks on the work of Axel Thue on sequences. *Nordisk Mat. Tidskr.*, 15:148–150, 1967.

[17] J. Justin, G. Pirillo, and S. Varricchio. Unavoidable regularities and finiteness conditions for semigroups. In A. Bertoni, C. Bohm, and P. Miglioli, editors, *Proc. 3rd Italian Conf. on Theoret. Comp. Sci. '89*, pages 350–355. World Scientific, Singapore, 1989.

[18] V. Keränen. A-2-free endomorphisms and substitutions over 4 letters. Available online at http://south.rotol.ramk.fi/keranen/words2007/a2f.html.

[19] V. Keränen. *On the k-Freeness of Morphisms on Free Monoids*. Number 61 in Annales Academiæ Scientiarum Fennicæ, Ser. A. I. Mathematica Dissertationes. Finnish Science Academy, 1986. 55 pages.

[20] V. Keränen. On the k-freeness of morphisms on free monoids. In F.J. Brandenburg, G. Vidal-Naquet, and M. Wirsing, editors, *Proc. STACS '87, Lecture Notes in Comp. Sci.*, volume 274, pages 180–188. Springer-Verlag, Berlin, 1987.

[21] V. Keränen. Abelian squares are avoidable on 4 letters. In W. Kuich, editor, *Proc. ICALP'92, Lecture Notes in Comp. Sci*, volume 623, pages 41–52. Springer-Verlag, Berlin, 1992.

[22] V. Keränen. New abelian square-free DT0L-languages over 4 letters. In V. Demidov and V. Keränen, editors, *Proc. IAS 2002*. Murmansk State Pedagogical Institute and Rovaniemi Polytechnic, 2002. Available online at http://south.rotol.ramk.fi/keranen/ias2002/ias2002papers.html.

[23] V. Keränen. On abelian square-free DT0L-languages over 4 letters. In T. Harju and J. Karhumäki, editors, *Proc. WORDS'03, 4th International Conference on Combinatorics on Words*, pages 95–109. TUCS General Publication, Turku, 2003.

[24] V. Keränen. Suppression of unfavourable factors in pattern avoidance. In B. Autin and Y. Papegay, editors, *eProc. (CD) 8th International Mathematica Symposium*, 2006. To appear in the *Mathematica* Journal.

[25] V. Keränen. New abelian square-free endomorphisms and a powerful substitution over 4 letters. In P. Arnoux, N. Bédaride, and J. Cassaigne, editors, *Proc. WORDS'07, 6th International Conference on Combinatorics on Words. Institut de Mathématiques de Luminy*. Marseille, September 17–21, 2007.

[26] T. Laakso. Musical rendering of an infinite repetition-free string. In C. Gefwert, P. Orponen, and J. Seppänen, editors, *Logic, Mathematics and the Computer. Proc. Finnish Artificial Intelligence Society*, volume 14, pages 292–297. Hakapaino, Helsinki, 1996.

[27] M. Leconte. A characterization of power-free morphisms. *Theoret. Comp. Sci.*, 38:117–122, 1985.

[28] M. Leconte. Kth power-free codes. In M. Nivat and D. Perrin, editors, *Proc. Automata on Infinite Words'84. Lecture Notes in Comp. Sci.*, volume 192, pages 172–187. Springer-Verlag, Berlin, 1985.

[29] M. Lothaire. *Combinatorics on Words*. Addison-Wesley, Reading, Massachusetts, 1983.

[30] S. Mäkelä. *Patterns in Words (in Finnish)*. M.Sc. Thesis. Univ. Turku, 2002.

[31] S.M. Mirkin. Expandable DNA repeats and human disease. *Nature*, 447:932–940, 2007.

[32] G. Pirillo and S. Varricchio. On uniformly repetitive semigroups. *Semigroup Forum*, 49:125–129, Springer-Verlag, New York,1994.

[33] P.A.B. Pleasants. Non-repetitive sequences. *Proc. Cambridge Phil. Soc.*, 68:267–274, 1970.

[34] G. Richomme and F. Wlazinski. Some results on k-power-free morphisms. *Theoret. Comp. Sci.*, 273:119–142, 2002.

[35] R.L. Rivest. Abelian square-free dithering for iterated hash functions. MIT, 2005. Available online at http://theory.lcs.mit.edu/∼rivest/publications.html.

[36] A. Salomaa. *Jewels of Formal Language Theory*. Computer Science Press, Rockville, Maryland, 1981.

[37] A. Thue. Über unendliche Zeichenreihe. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania*, 7:1–22, 1906.

[38] F. Wlazinski. *Ensembles de Test et Morphismes sans Répétition*. Thèse. Université de Picardie Jules Verne − LaRIA, 2002.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# A *Mathematica* Package for Construction of Circuit Matrices in Quantum Computation

Vladimir P. Gerdt [1], Robert Kragler [2], Alexander N. Prokopenya [3]

[1] *Joint Institute for Nuclear Research*
*141980 Dubna, Russia* gerdt@jinr.ru
[2] *University of Applied Sciences*
*D-88241 Weingarten, Germany*
kragler@hs-weingarten.de
[3] *Brest State Technical University*
*Moskowskaya 267, 224017 Brest, Belarus*
prokopenya@brest.by

**Abstract.** In the present paper we briefly describe a *Mathematica* package which allows users to specify a quantum circuit, to draw it, and to construct the unitary matrix for quantum computation defined by the circuit. For circuits composed from the Toffoli and Hadamard gates the package can also output the corresponding multivariate polynomial system over $\mathbb{F}_2$ whose number of solutions in $\mathbb{F}_2$ determines the circuit matrix. So the matrix can also be constructed by applying to the polynomial system the Gröbner basis technique based on the corresponding built-in *Mathematica* functions. For circuits of more general form the matrix is computed by means of the built-in *Mathematica* linear algebra tools. We illustrate the circuit matrix construction by an example.

## 1 Introduction

In spite of the recent unveiling of the 16-qubit quantum computer *Orion* by the Canadian company D-WAVE [1] and its demonstration in a black box regime, many top world experts doubt the authenticity of that demo. Doubts may be cleared up soon when, as anticipated, *Orion* is demonstrated in a realistic situation after further development. However, until realistic multiqubit quantum computers become available, a computer simulation of quantum computation will be a prerogative of classical computers. At present, there are several different simulators [2], a few of which are written in *Mathematica* [3].

In our previous papers [4, 5] we presented one more *Mathematica* program intended to build quantum circuits and compute their unitary matrices. Unlike other *Mathematica*-based quantum simulators [2], our program provides a

user-friendly graphical interface for generating quantum circuits and implements special computer-algebra tools for constructing the multivariate polynomial system that is uniquely defined for a circuit composed from Toffoli and Hadamard gates only [6]. The number of roots of the polynomial system in $\mathbb{F}_2$ fully determines the circuit matrix and can be found by using the Gröbner basis built in *Mathematica* module.

In the given paper we illustrate the indicated facilities of our *Mathematica* program by an example of a three-qubit circuit with five Hadamard and three Toffoli gates. In contrast to the example from [6] considered also in [5], the polynomial system in the present paper is underdetermined – it consists of four equations in five unknowns.

Nevertheless, as we show by explicit computation based on construction of a pure lexicographical Gröbner basis, it also admits an easy counting of the number of solutions, and, hence, construction of the circuit matrix.

The paper is organized as follows: In Section 2 we briefly describe the basic features of our program. Then we illustrate (Section 3) these features by inputting the example and computing its circuit matrix by applying straightforward built-in *Mathematica* linear algebra. In Section 4 we outline the relation of commutative algebra to the circuits built from Hadamard and Toffoli gates and apply the *Mathematica* function `GroebnerBasis` to reconstruct the example circuit matrix by counting the number of common roots for the multivariate polynomial system associated with the circuit.

## 2 Basic features of program

The circuit model is a model of quantum computation based on application of quantum logical gates to the input qubits [7] similar to application of classical logical gates in the classical circuit model. The principal difference between classical and quantum circuit models is that a quantum gate performs certain unitary transformation on the state of its target qubits whereas a classical gate performs boolean logical operations on the target bits.

A fixed number $n$ of qubits may serve as input for building a quantum circuit. Our program shows the input qubit states as $\mathbf{a_j}$ $(j = 1, 2, ..., n)$ from which quantum "wires" start and, under possible effect of quantum gates, the output qubit states $\mathbf{b_j}$ $(j = 1, 2, ..., n)$ at the terminus of the wires..

In quantum computing one uses certain sets of single-qubit, two-qubit and three-qubit gates. Some of the them, for example, the single-qubit Hadamard gate and the three-qubit Toffoli gate form a universal gate basis [8]. The data base of gates in our program contains the following gates [7]:

- **one-qubit gates:** Hadamard, Pauli X, Pauli Y and Pauli Z, Phase S and $\pi/8$ $T$.
- **two-qubit gates:** Controlled-X (CNOT), Controlled-Z, Controlled-S, Controlled-T and Swap gate.
- **three-qubit gates:** Toffoli (CCNOT).

This set of basis gates can be easily extended by the user.

In our package a quantum circuit is represented as a rectangular table. Rows in the table correspond to qubits in the circuit and columns show the circuit quantum gates and their arrangement. Each column in the table can contain either one multi-qubit gate or single-qubit gates only provided that no neighboring column contains only single-qubit gates acting on different qubits.

Thinking of any quantum circuit as a table of elementary quantum gates, our program associates a certain matrix with the circuit. By invoking the function `matrixGenerating` a window appears in which the user is invited to input the number of rows and columns of the circuit matrix. As a result, a skeleton matrix is generated with interactive unit entries. Then the user can input one-,two- or three-qubit gates into the matrix by changing the entries in an appropriate way. For example, the Hadamard gate is input by typing the symbol **H** in the corresponding entry whereas the Toffoli gate is prompted in a column by typing **C** for the desired locations of control qubits and **N** for the location of the target qubit.

Having all the gates given, the circuit can be depicted by invoking the function `circuit` whose argument is the output of function `matrixGenerating`. The unitary $2^n \times 2^n$ matrix of an $n-$qubit circuit is computed by the function `matrixU`. This function in turn calls the *Mathematica* built-in *LinearAlgebra* package.

For a circuit containing the Hadamard and Toffoli gates only, our package has special functions intended to construct the multivariate polynomial over $\mathbb{F}_2$ whose number of roots uniquely defines the circuit matrix [6]. This relation between quantum circuits and polynomial systems can also be used for computation of the circuit unitary matrices as we briefly describe in Section 5. Function `circuitPol` draws the circuit with explicit indication of the polynomial variables associated with the Hadamard gates, and function `polynomials` outputs the polynomial system. As an argument both functions have the matrix form of the circuit constructed by the function `matrixGenerating`.

# 3 Illustrative example

In this section we consider an example of a three-qubit circuit that contains five Hadamard and three Toffoli gates. First of all, as outlined in the previous section, we call the function which opens a window and asks the user to enter

**`matrixGenerating`**

Figure 1: Generating matrix for a quantum circuit

the number of rows (qubits) and the number of columns for the circuit. After entering 3 and 6, respectively, the program outputs the skeleton table (Fig.2). The unit entries of the skeleton matrix can now be interactively upgraded to specify the circuit as shown in Fig.3. The command of Fig. 3 outputs the circuit

$$\texttt{mat} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \texttt{; circuit[mat]}$$

Figure 2: Cell produced for matrix $3 \times 6$

$$\texttt{mat} = \begin{pmatrix} C & H & N & H & C & 1 \\ N & H & C & 1 & C & H \\ C & 1 & C & H & N & 1 \end{pmatrix} \texttt{; circuit[mat]}$$

Figure 3: Specification of entries in the circuit matrix of Fig.2

with five Hadamard gates and three Toffoli gates (Fig.4).

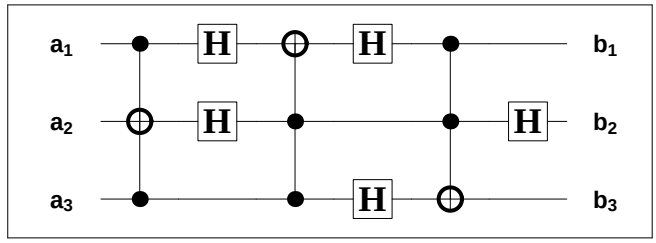



Figure 4: The circuit generated

The unitary $8 \times 8$ matrix determined by the circuit of Fig.4 is computed by calling function `matrixU` (see Fig.5) with the same argument as used in Fig.2 and Fig.3, and the output is shown in Fig.6.

# 4 Polynomial equations for circuits built from Hadamard and Toffoli gates

For circuits composed from the Toffoli and Hadamard gates that form a a universal gate basis [8], one can construct the circuit unitary matrix by counting the number of solutions for a multivariate polynomial system associated with the circuit [5, 6].

The systems of multivariate polynomials associated with such quantum circuits are obtained by applying the quantum-mechanical Feynman's sum-over-paths approach [6]. In so doing, the classical gate for the quantum Haramard gate outputs the path variable $x \in \mathbb{F}_2$ [6] irrespective of the input. Its value determines one of two possible paths of computation. Thereby, the classical Hadamard gate acts at qubit $a$ as

$$a \mapsto x\,, \qquad a, x \in \mathbb{F}_2\,.$$

```
matrixU[mat] // MatrixForm
```

Figure 5: Computation of unitary circuit matrix

$$\begin{pmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\
0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} \\
0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
0 & 0 & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0
\end{pmatrix}$$

Figure 6: The unitary matrix for circuit of Fig.4

Classically, the Toffoli gate acts on the triple of qubits with control qubits $a_1, a_2$ and target qubit $a_3$ in the following way

$$(a_1, a_2, a_3) \mapsto (a_1, a_2, a_3 \oplus a_1 a_2)$$

where $\oplus$ denotes addition modulo 2.

In Feynman's sum-over-paths approach, action of a quantum circuit is given as a sum over all possible classical paths. A classical path is defined by a sequence of classical bit strings $\mathbf{a} = \mathbf{s_1}, \mathbf{s_2}, \ldots, \mathbf{s_m} = \mathbf{b}$ obtained from action of the classical gates. Each set of values of the path variables $x_i$ gives a sequence of classical bit strings which form an admissible classical path.

For the circuit of Fig. 4 the path variables and, thus, all admissible classical paths, can be explicitly shown by invoking the function `circuitPol` (Fig.7). As Fig.8 shows, this function depicts the circuit together with the path variables and the related classical bit strings.

The sequence of classical bit strings for this circuit is given by $\mathbf{a} = \{a_1, a_2, a_3\}$ $= \mathbf{s_1}$, $\mathbf{s_2} = \{a_1, a_2 \oplus a_1 a_3, a_3\}$, $\mathbf{s_3} = \{x_1, x_2, a_3\}$, $\mathbf{s_4} = \{x_1 \oplus x_2 a_3, x_2, a_3\}$, $\mathbf{s_5} = \{x_3, x_2, x_4\}$, $\mathbf{s_6} = \{x_3, x_2, x_4 \oplus x_2 x_3\}$, $\mathbf{s_7} = \{x_3, x_5, x_4 \oplus x_2 x_3\} = \mathbf{b}$.

Each admissible classical path has a phase factor. The phase is determined in terms of the Hadamard gates applied [6] and is changed only if both input and output of the Hadamard gate are equal to 1. It yields the formula

$$\varphi(\mathbf{x}) = \sum_{\text{Hadamard gates}} input \bullet output \tag{1}$$

**`circuitPol[mat]`**

Figure 7: Function showing the path variables for circuit of Fig.4

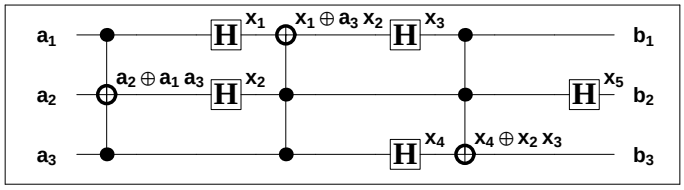



Figure 8: Path variables for circuit of Fig.4

where summation is done in $\mathbb{F}_2$. For all that Toffoli gates do not affect the phase.

In the example of Fig. 4 the phase of the path $\mathbf{x}$ is given by expression (cf. Fig.9)

$$\varphi(\mathbf{x}) = a_1x_1 \oplus a_2x_2 \oplus a_1a_3x_2 \oplus x_1x_3 \oplus a_3x_2x_3 \oplus a_3x_4 \oplus x_2x_5\,.$$

The Feynman's sum-over-paths method derives the following representation for matrix elements of a circuit matrix $U$ as sums over all the allowed paths from the initial classical state $\mathbf{a}$ to the final classical state $\mathbf{b}$ [6]

$$\langle \mathbf{b}|\, U\, |\mathbf{a}\rangle = \frac{1}{\sqrt{2^h}} \sum_{\mathbf{x}:\mathbf{b}(\mathbf{x})=\mathbf{b}} (-1)^{\varphi(\mathbf{x})}\,.$$

The sum is evaluated over $h$ Hadamard gates which are contained in the circuit.

Let $N_0$ be the number of positive terms in the sum and $N_1$ be the number of negative terms:

$$N_0 = |\,\{\ \mathbf{x}\mid \mathbf{b}(\mathbf{x}) = \mathbf{b} \quad \textbf{and} \quad \varphi(\mathbf{x}) = 0\ \}\,|\,, \tag{2}$$
$$N_1 = |\,\{\ \mathbf{x}\mid \mathbf{b}(\mathbf{x}) = \mathbf{b} \quad \textbf{and} \quad \varphi(\mathbf{x}) = 1\ \}\,|\ . \tag{3}$$

Thus, $N_0$ and $N_1$ count, respectively, the number of solutions in $\mathbb{F}_2^h$ for systems (2) and (3) of $n+1$ polynomials in $h$ variables over $\mathbb{F}_2$. Thereby the matrix element of the circuit unitary matrix $U$ may be written as the difference

$$\langle \mathbf{b}|\, U\, |\mathbf{a}\rangle = \frac{1}{\sqrt{2^h}}\,(N_0 - N_1)\ . \tag{4}$$

Our *Mathematica* package contains the function `polynomials[mat_?MatrixQ]` which constructs the set of polynomials over $\mathbb{F}_2$ which follows from the bit string of the form $\mathbf{b}(\mathbf{x}) = \mathbf{b}$ that relates the output classical qubit states $\mathbf{b}$ with the path variables. Here $\mathbf{b}(\mathbf{x})$ denotes the last bit string $\mathbf{s_m}$ in the admissible path set which depends polynomially on the path variables $\mathbf{x} = \{x_1, \ldots, x_h\}$.

```
x3⊕ b1
x5⊕ b2
x4⊕ x2 x3⊕b3
a1 x1⊕a2 x2⊕a1 a3 x2⊕x1 x3⊕a3 x2 x3⊕a3 x4⊕x2 x5
```

Figure 9: Polynomial system (5) for the circuit of Fig.4

To determine the circuit unitary matrix we have to count the number of solutions for polynomial systems (2) and (3) in $\mathbb{F}_2^h$ with the input and output bit variables $a_i, b_i$ taking values in $\mathbb{F}_2$. For this purpose the function `polynomials[mat_?MatrixQ]` of our *Mathematica* program outputs polynomials in the form $\mathbf{b}(\mathbf{x}) \oplus \mathbf{b} = 0$ and adds the phase polynomial (1) to the system.

For the circuit of Fig.4 the call of function `polynomials[mat]` outputs the system shown in Fig.9. The first three polynomial in Fig.9 are those generated by the output bit string relating the input and output qubit values for admissible paths coded in terms of the variables $\{x_1, x_2, x_3, x_4, x_5\}$. The last (bottom) polynomial is the phase polynomial defined by formula (1).

# 5 Solving circuit polynomial systems

To count the total number of solutions for the polynomial systems (2) and (3) when the variables take their values in $\mathbb{F}_2$ we apply formula (4) and rewrite the polynomial systems into the form

$$F_0 = \{\mathbf{b}(\mathbf{x}) \oplus \mathbf{b}, \phi(\mathbf{x})\}\,, \tag{5}$$

$$F_1 = \{\mathbf{b}(\mathbf{x}) \oplus \mathbf{b}, \phi(\mathbf{x}) \oplus 1\}\,. \tag{6}$$

Here $F_0$ denotes the output of the function `polynomials[mat_?MatrixQ]` in our *Mathematica* package. It is convenient to transform the system into the canonical Gröbner basis form [9]. The Gröbner basis method invented in [10] is the most universal algorithmic tool for investigating and solving multivariate polynomial systems. The pure lexicographical Gröbner bases are most suitable for our purposes, and the built-in module of *Mathematica* can be used for construction of such Gröbner bases.

For the system of polynomials $F_0$ in (5) shown in Fig.6 the lexicographical Gröbner basis for the ordering on the variables $x_5 \succ x_4 \succ x_3 \succ x_2 \succ x_1$ is given by

$$G_0 : \begin{cases} g_1 = a_1x_1 \oplus b_1x_1 \oplus a_2x_2 \oplus a_1a_3x_2 \oplus b_2x_2 \oplus a_3b_3\,, \\ g_2 = x_3 \oplus b_1\,, \\ g_3 = b_1x_2 \oplus x_4 \oplus b_3\,, \\ g_3 = x_5 \oplus b_2\,, \end{cases} \tag{7}$$

The Gröbner basis (7) can easily be obtained with *Mathematica*. To do this it is sufficient to define the polynomial sets (5) and (6) as *Mathematica* polynomial

```
F0 = {x3 + b1, x5 + b2, x4 + x2 x3 + b3, a1 x1 + a2 x2 +
      a1 a3 x2 + x1 x3 + a3 x2 x3 + a3 x4 + x2 x5};
```

Figure 10: Input *Mathematica* form for $F_0$ in (5)

lists by the command shown in Fig.10 and invoke the *Mathematica* function `GroebnerBasis` as shown in Fig.11. The last option in Fig.11 specifies the

```
GB0 = GroebnerBasis[F0, {x5, x4, x3, x2, x1},
  MonomialOrder → Lexicographic, Modulus → 2]
  /. y_^2 → y
```

Figure 11: *Mathematica* command for computation of (7)

coefficient field as $\mathbb{F}_2$. As a result, *Mathematica* will output the Gröbner basis (7) shown in Fig.12. Here the *Mathematica* substitution rule $x^2 \to x$ for all

```
{a3 b3 + a1 x1 + b1 x1 + a2 x2 + a1 a3 x2 + b2 x2, b1 + x3,
 b3 + b1 x2 + x4, b2 + x5}
```

Figure 12: The *Mathematica* output for the command of Fig.11

variables $x$ is used in the output. This substitution rule takes into account that we are looking for the number of roots for variables in (5) taking values in $\mathbb{F}_2$.

Similarly, for the system $F_1$ in (6) the Gröbner basis is

$$G_1 : \begin{cases} g_1 = a_1x_1 \oplus b_1x_1 \oplus a_2x_2 \oplus a_1a_3x_2 \oplus b_2x_2 \oplus a_3b_3 \oplus 1\,, \\ g_2 = x_3 \oplus b_1\,, \\ g_3 = b_1x_2 \oplus x_4 \oplus b_3\,, \\ g_3 = x_5 \oplus b_2\,, \end{cases} \tag{8}$$

and is computed with *Mathematica* exactly in the same way as (7). The lexicographical Gröbner bases (7) and (8) immediately yield the following relations between the parameters and the numbers $N_0$ and $N_1$ in (2) and (3):

$$G_0 : \begin{cases} a_1 \oplus b_1 = 1 \quad \to \quad N_0 = 2\,, \\ a_2 \oplus b_2 \oplus a_1a_3 = 1 \quad \to \quad N_0 = 2\,, \\ a_2 \oplus b_2 \oplus a_1a_3 = 0 \ \wedge \ a_3b_3 = 0 \quad \to \quad N_0 = 4\,, \\ a_2 \oplus b_2 \oplus a_1a_3 = 0 \ \wedge \ a_3b_3 = 1 \quad \to \quad N_0 = 0\,. \end{cases} \tag{9}$$

$$G_1 : \begin{cases} a_1 \oplus b_1 = 1 \quad \to \quad N_1 = 2\,, \\ a_2 \oplus b_2 \oplus a_1a_3 = 1 \quad \to \quad N_1 = 2\,, \\ a_2 \oplus b_2 \oplus a_1a_3 = 0 \ \wedge \ a_3b_3 = 0 \quad \to \quad N_1 = 0\,, \\ a_2 \oplus b_2 \oplus a_1a_3 = 0 \ \wedge \ a_3b_3 = 1 \quad \to \quad N_1 = 4\,. \end{cases} \tag{10}$$

In that way, the $8 \times 8$ matrix for the circuit of Fig.4 is easily determined by the formula (4) with $h = 5$ where $N_0$ and $N_1$ are defined, respectively, in (9) and (10). As a result, the matrix of Fig.6 is obtained.

For an $n-$qubit circuit with $h-$Hadamard gates the polynomial systems (5) and (6) contains $n + 1$ polynomials in $h-$variables $\mathbf{x} = \{x_1, x_2, \ldots, x_h\}$ and $2n-$parameters $\mathbf{a} = \{a_1, a_2, \ldots, a_n\}$, $\mathbf{b} = \{b_1, b_2, \ldots, b_n\}$. These parameters determine the values of the input and output qubits, respectively. To apply formula (4) for computing the circuit matrix by the Gröbner bases method, as we emphasized above, one needs to take into account that both variables and parameters are elements in the finite field $\mathbb{F}_2$. For this reason, generally, to increase the efficiency of computation with the use of the *Mathematica* function `GroebnerBasis`, instead of the substitution rule used in Fig.11, one should add to each of the polynomial systems (5) and (6) binomials of the form

$$x_i^2 + x_1 \quad (i = 1, \ldots, h) \tag{11}$$

and also take into account the restrictions

$$a_j^2 + a_j = 0, \quad b_j^2 + b_j = 0 \quad (j = 1, \ldots, n)\,.$$

Due to the last restrictions all the intermediate polynomials arising in the Gröbner basis construction by Buchberger's algorithm [9, 10] admit substantial simplification.

It turns out that if one uses another algorithmic approach for the construction of Gröbner bases called involutive (see [11] and references therein), then one can avoid handling extra binomials (11). In doing so, one can work with variables directly as with elements in $\mathbb{F}_2$. The first implementation in C++ of an involutive algorithm for computation of Gröbner basis over $\mathbb{F}_2$ with polynomial variables from $\mathbb{F}_2$ is described in [12].

In addition to quantum computation, solving systems of multivariate polynomial equations variables over $\mathbb{F}_2$ whose variables take values in $\mathbb{F}_2$ is also of interest for cryptoanalysis [13]. In particular, $n$ quadratic polynomials in $n \geq 80$ variables over field $\mathbb{F}_2$ were recommended as a public key, and $n = 80$ was suggested as the first challenge. In [14] this challenge was met by computing the Gröbner basis with a C program implementing author's algorithm [15]. This remarkable computational result gives hope that construction of circuit matrices by means of polynomial systems (5) and (6) may be computationally superior to the straightforward linear algebra based method for circuits with $n \gg h$ where $n$ and $h$ are, as above, the numbers of qubits and Hadamard gates, respectively.

**Acknowledgements.** The contribution of one of the authors (V.P.G.) was supported in part by grant 07-01-00660 from the Russian Foundation for Basic Research and by grant 5362.2006.2 from the Ministry of Education and Science of the Russian Federation.

# References

[1] `http://www.dwavesys.com/`

[2] `http://www.quantiki.org/wiki/index.php/List_of_QC_simulators`

[3] *Wolfram S.* The Mathematica book. Wolfram Media/Cambridge University Press (1999)

[4] *Gerdt V.P., Kragler R. and Prokopenya A.N.* On Simulation of Quantum Circuits with *Mathematica*. In: "Computer Algebra Systems in Teaching and Research / CASCR 2007", University of Podlasie, Scieldce, Poland, 2007, pp.135–144

[5] *Gerdt V.P., Kragler R. and Prokopenya A.N.* On computer algebra application to simulation of quantum computation. Proceedings of the UNISA-JINR Symposium "Models and Methods in Few- and Many-Body Systems" (Skukuza, Kruger National Park, South Africa, 6-9 February 2007), to appear

[6] *Dawson C.M., Nielsen M.A. et al.* Quantum computing and polynomial equations over the finite field $Z_2$. arXiv:quant-ph/0408129

[7] *Nielsen M. and Chuang I.* Quantum Computation and Quantum Information. Cambridge University Press (2000)

[8] *Aharonov D.* A Simple Proof that Toffoli and Hadamard Gates are Quantum Universal. arXiv:quant-ph/0301040

[9] *Buchberger B. and Winkler F.* (eds.) Gröbner Bases and Applications. Cambridge University Press, 1998

[10] *Buchberger B.* An Algorithm for Finding a Basis for the Residue Class Ring of a Zero-Dimensional Polynomial Ideal. PhD Thesis, University of Innsbruck, 1965 (in German)

[11] *Gerdt V.P.* Involutive Algorithms for Computing Gröbner Bases. In: "Computational commutative and non-commutative algebraic geometry", IOS Press, Amsterdam, 2005, pp. 199–225. arXiv:math.AC/0501111

[12] *Gerdt V.P. and Zinin M.V.* On Gröbner bases over $\mathbb{F}_2$. These Proceedings

[13] *Patarin J.* Hidden fields equations (HFE) and isomorphisms of polynomials (IP): Two new families of asymmetric algorithms. In: EUROCRYPT'96, volume 1070 of LNCS 1070, Springer-Verlag, 1996, pp.33–48

[14] *Faugère J.C. and Joux A.* Algebraic cryptanalysis of Hidden Field Equations (HFE) Using Gröbner Bases. LNCS 2729, Springer-Verlag, 2003, pp. 44–60

[15] *Faugère J.C.* A new efficient algorithm for computing Gröbner bases without reduction to zero ($F_5$). In: Proceedings of Issac'2002, ACM Press, New York (2002), pp.75–83

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# Constructivity of Involutive Divisions: Facts and Examples

Alexander S. Semenov and Petr A. Zyuzikov

*Moscow State University,*
*Department of Mechanics and Mathematics*
semyonov1980@mail.ru, petrzyuzikov@gmail.com

**Abstract.** This work considers involutive divisions and the property of constructivity, which assures the existence of a minimal involutive base for every commutative algebraic system. The main result is classification of constructive involutive $\succ$-divisions in three variables.

## 1 Introduction

The theory of involutive divisions [1, 2, 3] is the core of the approach to Gröbner base computing for commutative algebraic systems via an involutive base algorithm. The involutive base of a polynomial ideal occurs to as a variant of its Gröbner base, in most cases redundant but efficient in the sense of computational speed for a wide range of polynomial systems. However, the involutive approach uses an entirely different notation then the Buchberger algorithm, as the involutivity concept comes from PDE theory, where completion of a PDE system to involution has been used for many years as a common tool. The possibility of transferring the involutive PDE approach to commutative algebraic systems was suggested in [8, 9]. The algebraic framework for the involutive technique was developed by V. P. Gerdt and Yu. A. Blinkov [2, 3] and J. Apel [1]. In their works the concept of involutive divisions was explicitly introduced. An overview of involutive division theory can be found in [5] (in Russian).

An involutive division is a rule to separate the variables into multiplicative and non-multiplicative sets, as in the Pommaret, Janet, and Thomas approach to PDE systems. The description of algorithmically acceptable involutive divisions and their properties is a current problem in computer algebra theory. Investigating the process of involutive division researchers found that an involutive division appropriate for algorithmic use should possess the special properties of noetherianity, continuity, and constructivity [2, 3, 4]. In addition, monotonicity [4] is also an important property that simplifies the computational process, making it more "smooth".

There is a consensus among specialists that Janet division, well-known for its applications to PDEs, may be the best method to use in an involutive base computing algorithm for commutative algebraic ideals. This view is supported by

the good performance of the Janet base computing algorithm on some algebraic systems that are considered difficult traditionally and are used as benchmarks in Gröbner base computations. In spite of that, there is almost no hint of a rigorous proof of the conjecture that Janet division is excellent. In this situation the question of Janet division generalizations naturally arises. This paper presents some results in this direction based on the pair property of involutive divisions.

A generalization of Janet division is the class of $\succ$-divisions, each corresponding to an admissible monomial ordering. They can be explicitly described on monomial sets consisting of no more than two distinct elements, and on other monomial sets they can be defined according to the pair property. Janet division occurs to be the $\succ$-division, corresponding to lexicographic ordering.

The main result of this paper is the classification of constructive $\succ$-divisions for three variables together with the description of techniques useful in the classification process. Some of these techniques could work for a greater number of variables. The last topic will be the theme for future research by the authors.

The authors are grateful to their scientific advisor Dr. E.V. Pankrat'ev, as well as to Dr. V.P. Gerdt, for the help, remarks, and useful ideas which influenced the work. The work was partly supported by the Russian Foundation for Basic Research, project no. 05-01-00671.

## 2 Basic definitions

By $\mathbb{N}$ we denote the set of nonnegative integers. Then $\mathbb{M} = \{x_1^{d_1} \dots x_n^{d_n} | d_i \in \mathbb{N}\}$ is the set of all possible monomials in $n$ variables.

By $\deg(u)$ and $\deg_i(u)$, we denote the total degree of monomial $u$ and the degree of $u$ with respect to variable $x_i$. For the least common multiple and the greatest common divisor of two monomials $u$ and $v$, we use the notation $\mathrm{lcm}(u, v)$ and $\gcd(u, v)$.

In this work $U$ and $V$ denote finite monomial sets with distinct elements.

We say that an *involutive division* $L$ is specified, if, for any $u \in U$, a submonoid $L(u, U)$ of $\mathbb{M}$ is defined such that the following axioms hold [2, 3]:

- if $w \in L(u, U)$ and $v|w$, then $v \in L(u, U)$,
- if $u, v \in U$ and $uL(u, U) \cap vL(v, U) \neq \emptyset$, then $u \in vL(v, U)$ or $v \in uL(u, U)$,
- if $v \in U$ and $v \in uL(u, U)$, then $L(v, U) \subseteq L(u, U)$,
- if $U \subseteq V$, then $L(u, V) \subseteq L(u, U)$ (filter axiom).

Elements of $L(u, U)$ are *multiplicative* for $u$. If $w \in uL(u, U)$, then $u$ is an *involutive divisor* of $w$. This is denoted as $u|_L w$. The monomial $w$ is an *involutive multiple* of $u$. The monomial $v = w/u$ is *multiplicative* for $u$, and the equality $w = uv$ is written as $w = u \times v$. If $u$ is an ordinary divisor of $w$, but not an involutive one, then this equality is written as $w = u \cdot v$. In this case, the monomial $v$ is *non-multiplicative* for $u$.

For any $u$ in $U$, there exists a partition of the set of all variables into two disjoint sets, namely, *multiplicative variables* $M_L(u,U) \subset L(u,U)$ and *non-multiplicative variables* $NM_L(u,U) \notin L(u,U)$.

The submonoids $L(u,U)$ may be naturally interpreted geometrically. Consider a set $uL(u,U)$ for an involutive division $L$. Denote this set by $C_L(u,U)$. One can easily verify that the image of the set $C_L(u,U)$ under a bijective mapping from $\mathbb{M}$ onto $\mathbb{Z}^n$ is a discrete cone. The first three axioms are equivalent to the following two geometric facts:

- the set $C_L(u,U)$ is a discrete cone,
- $C_L(u,U) \cap C_L(v,U) \neq \emptyset \Longrightarrow C_L(u,U) \subseteq C_L(v,U) \vee C_L(v,U) \subseteq C_L(u,U)$

The notation $C_L(U) = \cup_{u \in U} C_L(u,U)$ is used further.

**Definition 1.** *An involutive division on a finite monomial set $U$ is disjoint, if $\nexists u, v \in U$, $v \neq u$, $v \in uL(u,U)$.*

**Example 1.** *(Thomas division) Consider a finite monomial set $U$ with distinct elements. Variable $x_i$ is treated as multiplicative for $u \in U$, if $\deg_i(u) = \max\{\deg_i(v) |\ v \in U\}$ and non-multiplicative otherwise.*

**Example 2.** *(Janet division) Consider a finite monomial set $U$ with distinct elements. For any $1 \leq i \leq n$, the set $U$ may be partitioned into subsets labeled with nonnegative integers $d_1, \ldots, d_i$ as follows:*

$$[d_1, \ldots, d_i] = \{u \in U | d_j = \deg_j(u), 1 \leq j \leq i\}.$$

*A variable $x_i$ is multiplicative for an element $u \in U$, if $i = 1$ and $\deg_1(u) = \max\{\deg_1(v) |\ v \in U\}$, or if $i > 1$, $u \in [d_1, \ldots, d_{i-1}]$, and $\deg_i(u) = \max\{\deg_i(v) |\ v \in [d_1, \ldots, d_{i-1}]\}$.*

The filter axiom can be reformulated as follows:

$$\forall u \in U \cap V \quad M_L(u, U \cup V) \subseteq M_L(u,U) \cap M_L(u,V).$$

Involutive divisions are mainly used to compute involutive bases of polynomial ideals (see [2, 3, 1]). Most algorithms for involutive base computations have a similar structure. The core of the algorithm is a main loop, which finishes the computing for every commutative ideal if the division possesses some special characteristic properties: noetherianity, continuity, constructivity [2, 3], or admissibility [1]. Each iteration of the loop consists of taking the polynomial $g \cdot x$, where $x$ is non-multiplicative for $\mathrm{lm}(g)$ and $g$ belongs to the base $G$ in construction, getting its involutive normal form, and adding it to $G$, if it is non-zero. The logic of any algorithm for involutive base computing supposes that, the greater the size and dimension of the sets of involutive multiples, the less the number of involutive prolongations should be considered. Starting from this idea, it is worth finding optimal divisions with the widest possible involutive cones. According to the filter axiom, these are the divisions with property $\forall u \in U \cap V$

$$M_L(u, U \cup V) = M_L(u,U) \cap M_L(u,V). \tag{1}$$

This class coincides with the class of **pairwise** involutive divisions, introduced in [4].

**Definition 2.** *[4] An involutive division $L$ is pairwise, if* $\forall U$ , $\forall u \in U (U \setminus \{u\} \neq \emptyset)$ *the following condition holds:*

$$M_L(u, U) = \bigcap_{v \neq u, v \in U} M_L(u, \{u, v\}) \tag{2}$$

The proof of equivalence is given in [6]. In [4], it is proved that Thomas and Janet divisions are pairwise. Besides, they are disjoint. Another important example of pairwise disjoint involutive division is $\succ$-division, where $\succ$ is an admissible monomial ordering.

**Definition 3 ($\succ$-division).** *Let $U$ be a finite monomial set of different elements and $L$ be an involutive division. A variable $x_i$ $(1 \leqslant i \leqslant n)$ is non-multiplicative for $u \in U$ if there exists $u_1 \in U$ such that $u_1 \succ u$, $i = \min\{j \mid \deg_j(u) < \deg_j(u_1)\}$.*

Janet division is a $\succ_{\text{lex}}$-division, where lex is a lexicographical ordering for which $x_1 \succ x_2 \succ \ldots \succ x_n$. A $\succ$-division may be defined equivalently using the pair property.

**Definition 4.** *The pairwise involutive $\succ$-division $L$ is defined on sets, consisting of no more that two elements (basic set), as follows. For an $\{u, v\}$ $(u \succ v)$:*

1. $NM_L(u, \{u, v\}) = \{\}$,
2. $NM_L(v, \{u, v\}) = x_p$, *where* $p = \min\{l \mid \deg_l(v) < \deg_l(u)\}$

*If the set $U$ contains one element, namely, $U = \{u\}$, then $NM_L(u, \{u\}) = \{\}$.*

Pairwise $\succ$-divisions can be viewed as candidates to be optimal among all involutive divisions, if the optimality is understood as the maximal width of the involutive cones. Consider the following argument to support this view:

If a set $U$ consists of one element, namely, $U = \{u\}$, then $NM_L(u, \{u\}) = \{\}$ is the best variant for involutive division. If a set $U$ contains two distinct elements, the optimal possibility is when one element has an $n$-dimensional involutive cone and another element has an $(n-1)$-dimensional cone or, in other words, one non-multiplicative variable. The sizes of involutive cones for sets $U$, which consist of more than two elements, are the maximal possible for a pairwise division, if the division on all two-element sets is completely defined.

In papers [2, 3] a concept of *continuity* of involutive divisions is introduced. Let an involutive division $L$ be given. A sequence of monomials $\{u_t\}$ $(t \in \{1, 2 \ldots\})$ in a finite monomial set $U$ with distinct elements, such that there exists $x_{k(t)} \in NM_L(u_t, U)$ and $[u_{t+1}|_L u_t \cdot x_{k(t)}]$ is called a *chain*.

**Definition 5.** *[2, 3] An involutive division $L$ is* **continuous**, *if, in every chain, the inequality $\forall s_1 \neq s_2\ u_{s_1} \neq u_{s_2}$ holds, and, hence, every chain is finite.*

**Definition 6.** *[2, 3] A continuous involutive division $L$ is constructive on a finite set $U \subset \mathbb{M}$ with distinct elements if for any $u \in U$, $x_i \in NM_L(u, U)$ such that $u \cdot x_i$ has no involutive divisors in $U$ and*

$$(\forall v \in U)(\forall x_j \in NM_L(v, U))(v \cdot x_j | u \cdot x_i, v \cdot x_j \neq u \cdot x_i) \Rightarrow$$

$$\Rightarrow [v \cdot x_j \in C_L(U)]$$

*the following condition holds:*

$$\forall w \in C_L(U)[u \cdot x_i \notin C_L(U \cup \{w\})].$$

*A continuous involutive division $L$ is constructive if it is constructive on every finite set $U$ with distinct elements. In [7] it is proven, that all $\succ$-divisions are continuous.*

The constructivity property [2, 3] is a sufficient condition for the existence and uniqueness of a minimal involutive base. Results on classification of constructive $\succ$-divisions are presented in the next section.

# 3 Constructivity. Main Results

In this part we present results concerning constructivity. The most important of them is the description of all constructive divisions in three variables.

**Definition 7.** *Let $L$ be an involutive $\succ$-division. A set of distinct monomials $\{u_1, u, w, \widehat{u}\}$ is a $\gamma$-configuration, if it satisfies the following conditions:*

1. $u \prec \widehat{u}$,
2. $w = u_1 \times v$, $v \in L(u_1, \{u_1, u, \widehat{u}\})$,
3. $u \cdot x_i \in wL(w, \{\widehat{u}, u, u_1, w\})$, *where* $x_i = NM_L(u, \{u, \widehat{u}\})$.

Further, the index $i$ and notation $x_i$ is always used in the previous sense ($x_i = NM_L(u, \{u, \widehat{u}\})$). Also, in case $u_1 \prec u$ the notation $x_j$ and $x_k$ will be used for $x_j = NM_L(u_1, \{u_1, u\})$ and $x_k = NM_L(u_1, \{u_1, \widehat{u}\})$. Classification of $\gamma$-configurations is an important step towards describing constructive involutive divisions.

**Lemma 1.** *Let $L$ be an involutive $\succ$-division. Every $\gamma$-configuration $\{u_1, u, w, \widehat{u}\}$ can be of two possible types ($m = \gcd(u_1, u)$).*

1. *(1st type)* $u_1 = m,\ u = mx_1^{k_1} \dots x_n^{k_n}$,
2. *(2nd type)* $u_1 = mx_i,\ u = mx_1^{k_1} \dots x_{i-1}^{k_{i-1}} x_{i+1}^{k_{i+1}} x_n^{k_n}$.

*Proof.* It can be seen from the definition that $u_1|u \cdot x_i$ in the conventional sense. If $\deg_i(u_1) \leq \deg_i(u)$, then $m = u_1$ and the $\gamma$-configuration is of the first type. Otherwise, it belongs to the second type. □

**Lemma 2.** *[7] (Sufficient conditions of constructivity) Let $L$ be an involutive $\succ$-division. If there exists no $\gamma$-configuration $\{u_1, u, w, \widehat{u}\}$ for which the relations $u_1 \prec u$, and $u_1 <_{\mathrm{lex}(\prec)} u$ are simultaneously valid, then $L$ is constructive.*

**Lemma 3.** *[7] For every $\gamma$-configuration $\{u_1, u, w, \widehat{u}\}$ corresponding to a $\succ$-division, where $u_1 \prec u$, the relation $x_i \succ x_j$ is valid.*

**Lemma 4.** *Let $L$ be an involutive $\succ$-division with $\gamma$-configuration $\{u_1, u, w, \widehat{u}\}$. Then $\deg_i(u_1 v) = \deg_i(u) + 1$.*

*Proof.* It is obvious that $\deg_i(u_1 v) \leq \deg_i(u) + 1$. If $\deg_i(u_1 v) < \deg_i(u) + 1$ then $u \cdot x_i = u_1 v \times w, u = u_1 v \times \frac{w}{x_i}$, which contradicts the properties of disjoint division. □

**Lemma 5.** *Let $L$ be an involutive $\succ$-division. Let $\{u_1, u, w, \widehat{u}\}$ be a $\gamma$-configuration of the first type. Then $j < i$.*

*Proof.* Obviously $\deg_i(u_1 v) = \deg_i(u) + 1$ and $u_1 = m$ is multiplied by $x_i$. According to its definition, $x_j$ is the variable for which the following relation is satisfied:

$$j = \min\{l | \deg_l(u_1) < \deg_l(u)\}.$$

Due to lemma 3 $i \neq j$. If $i < j$ then $\deg_i(u) = \deg_i(u_1)$. As the $\gamma$-configuration is of the first type, $\deg_{j-}(u) = \deg_{j-}(u_1) = \deg_{j-}(m)$. This relation implies $NM_L(u, \{u, \widehat{u}\}) = NM_L(u_1, \{u_1, \widehat{u}\}) = \{x_i\}$. The last equality contradicts the fact that $x_i$ is multiplicative for $u_1$. □

**Lemma 6.** *Consider the second-type of $\gamma$-configuration for $n$ variables, for which the relations $u_1 \prec u$, and $u_1 <_{\mathrm{lex}(\prec)} u$ are simultaneously valid. The variables can be notated as $x_i, x_j$, and $x_{l_1}, \ldots, x_{l_{n-2}}$. Then there exists $l_s$ such that $l_s > j$, $x_{l_s} \succ x_j$.*

*Proof.* Consider the second-type of $\gamma$-configuration for $n$ variables. There are variables $x_i, x_j$, and $x_{l_1}, \ldots, x_{l_{n-2}}$. The relations which define the configuration type have the following form:

$$x_i \prec x_j^{k_j} x_{l_1}^{r_1} \ldots x_{l_{n-2}}^{r_{n-2}} \prec x_i x_{l_1}^{r_1 - p_1} \ldots x_{l_{n-2}}^{r_{n-2} - p_{n-2}},$$

and the same for the ordering $\mathrm{lex}(\prec)$. Due to the definition of $j$, for all $l_q$, such that $r_q > 0$, $l_q > j$ holds.

Also $k_j > 0$. This yields

$$x_j^{k_j} x_{l_1}^{p_1} \ldots x_{l_{n-2}}^{p_{n-2}} \prec x_i \prec x_j^{k_j} x_{l_1}^{r_1} \ldots x_{l_{n-2}}^{r_{n-2}},$$

and

$$x_j^{k_j} x_{l_1}^{p_1} \ldots x_{l_{n-2}}^{p_{n-2}} <_{\mathrm{lex}(\prec)} x_i <_{\mathrm{lex}(\prec)} x_j^{k_j} x_{l_1}^{r_1} \ldots x_{l_{n-2}}^{r_{n-2}}.$$

If $p_t > 0$ then $x_{l_t} \prec x_i$. If all $p_1, \ldots, p_{n-2}$ are greater than zero, then all $x_{l_1}, \ldots, x_{l_{n-2}} <_{\mathrm{lex}(\prec)} x_i$, and $x_i <_{\mathrm{lex}(\prec)} x_j^{k_j} x_{l_1}^{r_1} \ldots x_{l_{n-2}}^{r_{n-2}}$ is impossible. So there exists an $s$ such that $p_s = 0$.

We can choose $s$, for which also $r_s \neq 0$ and $x_{l_s} >_{\mathrm{lex}(\prec)} x_i$. Due to $x_i \succ x_j$, $x_{l_s} >_{\mathrm{lex}(\prec)} x_j$ and $x_{l_s} \succ x_j$ take place. Then for any $q, r_q - p_q > 0$, $l_q > j$. Therefore $l_s > j$. □

**Theorem 1.** *[7] Consider the $\succ$-division $L$. If for the ordering $\succ$ one of four conditions*

1. $\exists i < j < k$ *s.t.* $x_j \prec x_i \prec x_k$,
2. $\exists i < j < k$ *s.t.* $x_i \prec x_j \prec x_k$,
3. $\exists i < j < k, p \in \mathbb{N}, p > 1$ *s.t.* $x_i \prec x_k \prec x_j \prec x_k^p$,
4. $\exists i < j < k < l$ *s.t.* $x_j \succ x_k \succ x_i$, $x_j \succ x_k \succ x_l$

*is satisfied, $L$ is non-constructive.*

*Proof.* The proof is done according to the following scheme. First, the set $U$ is constructed, and $L$ is defined on $U$ according to the pairwise property. Then, for each case, the example of non-constructivity is given by specifying the set $U$, the non-multiplicative prolongation $\overline{u}$ of the element $u \in U$, and the multiplicative prolongation $w$ of the element $v \in U$. All the conditions on these elements can be checked from tables of variable partitioning for $U$ and $U \cup \{w\}$. The tables in all cases can be seen in the proof of this theorem in [7].

In case 1, relation $x_j \prec x_i \prec x_k$ implies $x_i \prec x_i x_j \prec x_j x_k \prec x_i x_k \prec x_i^2 x_j x_k$. Then $U = \{x_i, x_i x_j, x_j x_k, x_i^2 x_j x_k\}$, and $w = x_i \times x_k$. The main relation is $x_i x_j \cdot x_k = x_j x_k \cdot x_i = x_i x_k \times x_j$.

In case 2, relation $x_i \prec x_j \prec x_k$ implies $x_j \prec x_i x_j \prec x_i x_k \prec x_j x_k \prec x_j^2 x_k$. Then $U = \{x_j, x_i x_j, x_i x_k, x_j^2 x_k\}$, and $w = x_j \times x_k$. The main relation is $x_i x_k \cdot x_j = x_i x_j \cdot x_k = x_j x_k \times x_i$.

In case 3, relation $x_i \prec x_k \prec x_j \prec x_k^p$ implies $x_k \prec x_i x_k \prec x_i x_j \prec x_j x_k \prec x_k^{p+1}$. Then $U = \{x_k, x_i x_k, x_i x_j, x_k^{p+1}\}$, $w = x_k \times x_j$. The main relation is $x_i x_k \cdot x_j = x_i x_j \cdot x_k = x_j x_k \times x_i$.

The fourth case is proven as follows: Consider the case $x_j \succ x_k \succ x_l \succ x_i$. If $x_i x_l \succ x_k$ then $x_l^2 \succ x_k$ and the variables $x_l^2 \succ x_k \succ x_l \succ x_i$ satisfy case 3. Then $x_i x_l \prec x_k$, and it is evident that $x_j x_l \succ x_k$, which also yields non-constructivity by Theorem 7.

Consider the case $x_j \succ x_k \succ x_i \succ x_l$. If $x_j x_l \prec x_k^2$ then $x_j \prec x_k^2$ and the relation $x_k^2 \succ x_j \succ x_k \succ x_i$ satisfies case 3. Then $x_j x_l \succ x_k^2$ and $x_k^2 \succ x_i x_l$ which also yields non-constructivity by Theorem 7. Thus, all four cases are considered, thus proving the theorem. □

The next step is to describe all possible $\gamma$-configurations for three variables. If $x_3$ is greater than all other variables, the $\succ$-division is non-constructive. Consider the special case of a second-type $\gamma$-configuration for $n = 3$. There are

three variables $x_i, x_j$, and $x_l$. Lemma 6 yields $x_l \succ x_i \succ x_j$, and $l > j$. So lemmas 5 and 6 give an intuition about which configuration may yield constructive divisions for $n = 3$. The cases of the four remaining variable permutations are listed below in the following table.

| *Ordering* | 1*st type* | 2*nd type* |
|---|---|---|
| $x_1 \succ x_2 \succ x_3$ | – | – |
| $x_1 \succ x_3 \succ x_2$ | $i = 3, j = 2, l = 1$ | – |
| $x_2 \succ x_1 \succ x_3$ | $i = 2, j = 1, l = 3$ | – |
| $x_2 \succ x_3 \succ x_1$ | $i = 3, j = 1, l = 2$; $\quad i = 2, j = 1, l = 3$ | $i = 3, j = 1, l = 2$ |

**Theorem 2.** *[7] Let $\succ$ be an admissible monomial ordering, such that $x_1 \succ x_2 \succ \ldots \succ x_n$. There is no $\gamma$-configurations with $u_1 \prec u$ and $u_1 <_{\mathrm{lex}(\prec)} u$ and involutive $\succ$-division $L$ is constructive.*

*Proof.* As for first-type configurations, we know that $x_i \succ x_j$, and $i > j$ from lemma 5. This is impossible with this ordering and thus there are no $\gamma$-configurations of the first type. As for second-type configurations, there exists $l_s$ such that $l_s > j$, $x_{l_s} \succ x_j$. This is impossible and there are no $\gamma$-configurations of the second type. From lemma 2 we know that if there are no $\gamma$-configurations, division is constructive. □

**Example 3.** *Example of non-constructive $\succ$-division.*

Consider the $\succ$-division $L$. If the ordering $\succ$ satisfies the conditions: $x_2 \succ x_1 \succ x_3$ and $\exists s, t \geq 0$ s.t. $x_2^{t+1} \succ x_3^{s+1} \succ x_1 x_2^t$, then $L$ is non-constructive.

The relation $x_2^{t+1} \succ x_3^{s+1} \succ x_1 x_2^t$ implies $x_1^2 x_3^{s+1} \prec x_1^2 x_2^{t+1}$ and $x_1^2 x_2^{t+1} \prec x_1 x_2 x_3^{s+1}$.

Then consider $U = \{x_1 x_3^{s+1}, x_1^2 x_3^{s+1}, x_1^2 x_2^{t+1}\}$, $w = x_1 x_3^{s+1} \times x_2$. The main relation is $x_1^2 x_3^{s+1} \cdot x_2 = x_1 x_2 x_3^{s+1} \times x_1$.

| $U$ | $NM_L(U)$ | $U \cup \{x_1 x_2 x_3^{s+1}\}$ | $NM_L(U \cup \{x_1 x_2 x_3^{s+1}\})$ |
|---|---|---|---|
| $x_1 x_3^{s+1}$ | $x_1$ | $x_1 x_3^{s+1}$ | $x_1, x_2$ |
| $x_1^2 x_3^{s+1}$ | $x_2$ | $x_1^2 x_3^{s+1}$ | $x_2$ |
| $x_1^2 x_2^{t+1}$ | – | $x_1^2 x_2^{t+1}$ | $x_3$ |
| | | $x_1 x_2 x_3^{s+1}$ | – |

**Theorem 3.** *Let $\succ$ be an ordering, for which $x_2 \succ x_1 \succ x_3$ is valid, and there do not exist $s, t \geq 0$, $x_2^{t+1} \succ x_3^{s+1} \succ x_1 x_2^t$. Then no $\gamma$-configurations exist for $\succ$ and $\succ$-division $L$ is constructive.*

*Proof.* According to lemma 2 it is sufficient to prove that there is no $\gamma$-configurations for $L$. Consider the opposite. Let $\{u_1, u, w, \widehat{u}\}$ be a $\gamma$-configuration for $L$. According to the previous table $\{u_1, u, w, \widehat{u}\}$ is a $\gamma$-configuration of the first type, and $i = 2, j = 1, l = 3$.

Then, necessarily $k = 1$, $\deg_1(u_1) < \deg_1(\widehat{u}) \leq \deg_1(u)$, and so we have

$$\deg_2(u) < \deg_2(u) + 1 = \deg_2(u_1 v) \leq \deg_2(\widehat{u}),$$

$$\deg_1(u_1) = \deg_1(u_1 v) < \deg_1(\widehat{u}) \leq \deg_1(u).$$

Also relation $\widehat{u} \prec u_1 v$ is necessary, since in the opposite case

$$NM_L(u_1 v, \{u_1 v, \widehat{u}\}) = x_1, \quad \deg_1(u) > \deg_1(u_1 v)$$

and we obtain a contradiction. So $u \prec \widehat{u} \prec u_1 v$.

Then $\deg_3(\widehat{u}) < \deg_3(u_1 v) \leq \deg_3(u)$, since in the opposite case $\widehat{u}$ would be greater than $u_1 v$ with respect to $\succ$.

Relations for degrees yield $u = m_1 x_1^{1+r+r_1} x_3^{1+s+s_1}, \widehat{u} = m_1 x_1^{1+r} x_2^{1+t}, u_1 v = m_1 x_2 x_3^{1+s}$, $m_1 = \gcd(u, u_1 v, \widehat{u})$.

So we have relations on $x_1, x_2, x_3$:

$$x_1^{1+r+r_1} x_3^{1+s+s_1} \prec x_1^{1+r} x_2^{1+t} \prec x_2 x_3^{1+s}.$$

The second relation yields $x_1 x_2^t \prec x_3^{1+s}$ and the first yields $x_3^{1+s} \prec x_2^{1+t}$ for some $s, t \geq 0$. □

**Example 4.** *Example of non-constructive $\succ$-division.*

Consider the $\succ$-division $L$. If the ordering $\succ$ satisfies the condition: $x_1 \succ x_3 \succ x_2$ and $\exists s, t > 0$ s.t. $x_3^{t+1} \succ x_1^{s+1} \succ x_2 x_3^t$, then $L$ is non-constructive.

The relation $x_3^{t+1} \succ x_1^{s+1} \succ x_2 x_3^t$ implies $x_2^2 x_3^{t+1} \succ x_2^2 x_1^{s+1}$ and $x_2 x_3 x_1^{s+1} \succ x_2^2 x_3^{t+1}$. Also, $x_2^2 x_3^{t+1} \succ x_2 x_1^{s+1}$ is valid.

Then consider $U = \{x_1^{s+1} x_2, x_1^{s+1} x_2^2, x_2^2 x_3^{t+1}\}$, $w = x_1^{s+1} x_2 \times x_3$. The main relation is $x_1^{s+1} x_2^2 \cdot x_3 = x_1^{s+1} x_2 x_3 \times x_2$.

| $U$ | $NM_L(U)$ | $U \cup \{x_2 x_3 x_1^{s+1}\}$ | $NM_L(U \cup \{x_2 x_3 x_1^{s+1}\})$ |
|---|---|---|---|
| $x_1^{s+1} x_2$ | $x_2$ | $x_1^{s+1} x_2$ | $x_2, x_3$ |
| $x_1^{s+1} x_2^2$ | $x_3$ | $x_1^{s+1} x_2^2$ | $x_3$ |
| $x_2^2 x_3^{t+1}$ | – | $x_2^2 x_3^{t+1}$ | $x_1$ |
| | | $x_1^{s+1} x_2 x_3$ | – |

**Theorem 4.** *Let $\succ$ be an ordering, for which $x_1 \succ x_3 \succ x_2$ is valid. Then $\succ$-division $L$ is constructive and no $\gamma$-configurations exist, if and only if there do not exist $t, s > 0$: $x_3^{t+1} \succ x_1^{s+1} \succ x_2 x_3^t$.*

*Proof.* Obviously, if $t, s > 0$, and $x_3^{t+1} \succ x_1^{s+1} \succ x_2 x_3^t$ exist, the division is non-constructive, according to the previous example.

The next step is to prove that every non-constructive division for an ordering $\succ$, for which $x_1 \succ x_3 \succ x_2$, satisfies the second property. Suppose the opposite. Let $\{u_1, u, w, \widehat{u}\}$ be a $\gamma$-configuration. Due to the classification above, it is of the first type, and $i = 3, j = 2, l = 1$.

Consider the monomial $u_1 v$. As $u_1 v \succ_{\mathrm{lex}(\prec)} u$, then necessarily $\deg_1(u_1 v) = \deg_1(u)$.

$$\deg_1(\widehat{u}) \leqslant \deg_1(u) = \deg_1(u_1 v) = \deg_1(u_1),$$

$$\deg_2(u_1) = \deg_2(u_1 v) < \deg_2(\widehat{u}) \leqslant \deg_2(u),$$

$$\deg_3(u_1) \leqslant \deg_3(u) < \deg_3(u_1 v) \leqslant \deg_3(\widehat{u}).$$

These relations yield $\widehat{u} \prec u_1 v$, since if $u_1 v \prec \widehat{u}$, $NM_L(u_1 v, \{u_1 v, \widehat{u}\}) = \{x_2\}$. Then we have $u \prec \widehat{u} \prec u_1 v$. Relations for degrees lead to $u = m_1 x_1^{s+1} x_2^{1+r+\lambda}$, $\widehat{u} = m_1 x_2^{1+r} x_3^{t+1}$, $u_1 v = m_1 x_1^{s+1} x_3$, $m_1 = \gcd(u, u_1 v, \widehat{u})$.

So we have relations on $x_1, x_2, x_3$:

$$x_1^{s+1} x_2^{1+r+\lambda} \prec x_2^{1+r} x_3^{t+1} \prec x_1^{s+1} x_3.$$

They are equivalent to the following relations:

$$x_1^{s+1} x_2^{\lambda} \prec x_3^{t+1} (\Rightarrow x_1^{s+1} \prec x_3^{t+1}), x_2^{1+r} x_3^{t} \prec x_1^{s+1} (\Rightarrow x_2 x_3^{t} \prec x_1^{s+1}).$$

□

**Theorem 5.** *Consider the case when $x_2 \succ x_3 \succ x_1$. Then $\succ$-division $L$ is constructive if and only if*

$$\forall p, q \in \mathbb{M} \quad \deg_2(p) > \deg_2(q) \Rightarrow p \succ q.$$

*Proof.* If the second condition of the theorem is not satisfied then $\exists p > 1$, $x_3^p \succ x_2$, and the division is non-constructive due to Theorem 1. Then, if division is constructive, the second condition of the theorem is satisfied necessarily. Our aim is to prove that any $\succ$-division, for which $x_2 \succ x_3 \succ x_1$ and the second condition satisfied, is constructive.

Consider the opposite. Then there exists a set $U$ for which constructivity does not hold on elements $u_1, u, u_1 v \in U$ and a corresponding $\gamma$-configuration $\{u_1, u, w, \widehat{u}\}$ for which relations $u_1 \prec u$ and $u_1 <_{\mathrm{lex}(\prec)} u$ hold simultaneously.

In the proof of the theorem we will consider that $\widehat{u}$ is the maximal possible with respect to $\succ$, which satisfies the corresponding properties of a $\gamma$-configuration.

Due to lemma 3, $i \neq 1$ so

$$\deg_1(\widehat{u}) \leq \deg_1(u).$$

According to the definition of the ordering $\succ$, we have

$$\deg_2(u_1) \leq \deg_2(u) \leq \deg_2(\widehat{u}).$$

If $\deg_2(u) < \deg_2(\widehat{u})$ then $i = 2$, and otherwise $i = 3$.

The case $\deg_2(u_1) = \deg_2(u) = \deg_2(\widehat{u})$ is impossible, as it yields $2 \notin \{i, j, k\}$, and the proof is analogous to that of Theorem 6. So,

$$\deg_2(u_1) < \deg_2(\widehat{u}).$$

As $u_1 v >_{\mathrm{lex}(\prec)} u \deg_2(u) \leq \deg_2(u_1 v)$, and, if $i = 2$, $\deg_2(u_1 v) = \deg_2(u)+1$.

The next step is to prove that $\deg_2(u_1) < \deg_2(u_1 v)$. If $i = 2$ it is obvious. If $i = 3$ then $\deg_2(u) = \deg_2(\widehat{u})$, and $\deg_2(u_1) < \deg_2(u) = \deg_2(u_1 v)$.

Hence, $x_2 \in M_L(u_1, \{u_1, u\})$ and $x_2 \in M_L(u_1, \{u_1, \widehat{u}\})$. Due to $\deg_2(u_1) < \deg_2(\widehat{u})$, $k = 1$, which yields

$$\deg_1(u_1) < \deg_1(\widehat{u}) \leq \deg_1(u),$$

and $j = 1$.

The last relation implies that if $u_1 v \prec \widehat{u}$ then $NM_L(u_1 v, \{u_1 v, \widehat{u}\}) = x_1$ which is impossible since $\deg_1(u_1 v) < \deg_1(u)$. So, $\widehat{u} \prec u_1 v$, hence

$$\deg_2(u_1 v) = \deg_2(\widehat{u}).$$

For $i = 3$ we show below that $u_1 v | \widehat{u}$ and $NM_L(u_1 v, \{u_1 v, \widehat{u}\}) = x_j$. This leads to a contradiction and proves the theorem. □

**Corollary 1.** *Let $L$ be a $\succ$-division in three variables. It is constructive in and only in the following cases:*

- $x_1 \succ x_2 \succ x_3$,
- $x_2 \succ x_1 \succ x_3$, $\nexists s > 0, t > 0$ *s.t.* $x_2^{t+1} \succ x_3^{s+1} \succ x_1 x_2^t$
- $x_2 \succ x_3 \succ x_1$, $\forall p, q \in \mathbb{M} \quad \deg_2(p) > \deg_2(q) \Rightarrow p \succ q$,
- $x_1 \succ x_3 \succ x_2$, $\nexists s > 0, t > 0$ *s.t.* $x_3^{t+1} \succ x_1^{s+1} \succ x_2 x_3^t$.

Below are outlined two theorems, which consider the cases of two and four variables.

**Theorem 6.** *[7] Consider the two-variable case, i.e. when $n = 2$. Then every involutive $\succ$-division $L$ is constructive.*

**Theorem 7.** *Consider the $\succ$-division $L$. If for the ordering $\succ$ the condition is satisfied: $\exists s, m, q \in \mathbb{N}$ s.t. $x_2^{m+1} x_4^q \succ x_3^s \succ x_1 x_2^m x_4^q$, then $L$ is non-constructive.*

*Proof.* The relation $x_2^{m+1} x_4^q \succ x_3^s \succ x_1 x_2^m x_4^q$ implies $x_1^2 x_2^{m+1} x_4^q \succ x_1^2 x_3^s$ and $x_1 x_2 x_3^s \succ x_1^2 x_2^{m+1} x_4^q$. Also, $x_1^2 x_2^{m+1} x_4^q \succ x_1 x_3^s$ is valid.

Then consider $U = \{x_1 x_3^s, x_1^2 x_3^s, x_1^2 x_2^{m+1} x_4^q\}$, $w = x_1 x_3^s \times x_2$. The main relation is $x_1^2 x_3^s \cdot x_2 = x_1 x_2 x_3^s \times x_1$.

| $U$ | $NM_L(U)$ | $U \cup \{x_1 x_2 x_3^s\}$ | $NM_L(U \cup \{x_1 x_2 x_3^s\})$ |
|---|---|---|---|
| $x_1 x_3^s$ | $x_1$ | $x_1 x_3^s$ | $x_1, x_2$ |
| $x_1^2 x_3^s$ | $x_2$ | $x_1^2 x_3^s$ | $x_2$ |
| $x_1^2 x_2^{m+1} x_4^q$ | – | $x_1^2 x_2^{m+1} x_4^q$ | $x_3$ |
| | | $x_1 x_2 x_3^s$ | – |

That proves the theorem. □

# References

[1] Apel J. The Theory of Involutive Divisions and an Application to Hilbert Function Computations. *Journal of Symbolic Computations,* (1998) **25**, no 6, 683–704.

[2] Gerdt V.P., Blinkov Yu. A. Involutive Bases of Polynomial Ideals. *Mathematics and Computers in Simulation,* (1998) **45**, 519–542.

[3] Gerdt V.P., Blinkov Yu. A. Minimal Involutive Bases. *Mathematics and Computers in Simulation,* (1998) **45**, 543–560.

[4] Gerdt V.P. Involutive Division Technique: Some Generalizations and Optimizations. *Journal of Mathematical Sciences*, (2002) 108(6), 1034–1051.

[5] Pankrat'ev E.V. The Elements of Computer Algebra (in Russian). *INTUIT-BINOM*, Moscow, 2007.

[6] Semenov A.S. Pair Analysis of Involutive Divisions. *Fundamental and Applied Mathematics*, (2003) 9(3), 199–212.

[7] Semenov A.S. On Connection Between Constructive Involutive Divisions and Monomial Orderings. *CASC-2006 Proceedings*, Springer Verlag, (2006) 261–278.

[8] Zharkov A. Yu. and Blinkov Yu. A. Involutive Approach to Investigating Polynomial Systems. Proceedings of "SC93", International IMACS Symposium on Symbolic Computations: New Trends and Developments (Lille, June 14-17,1993). *Mathematics and Computers in Simulation*, (1996) **42**, 323–332.

[9] Zharkov A. Yu. and Blinkov Yu. A. Involutive Bases of Zero-Dimensional Ideals. Preprint No. E5-94-318, Joint Institute for nuclear Research, Dubna, 1994.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# Advantages and Disadvantages of Differential Standard Bases

M. V. Kondratieva, E. V. Pankratiev,
D. V. Trushin, and A. I. Zobnin

*Faculty of Mechanics and Mathematics,*
*Moscow State University*

**Abstract.**

We give a brief sketch of recent results on finiteness conditions for differential standard bases in ordinary rings of differential polynomials $\mathcal{F}\{y\}$.

## 1 Introduction

Consider a differential polynomial ring $\mathcal{F}\{y\}$ over a field of constants $\mathcal{F}$ of characteristic zero. The elements of this ring correspond to autonomous differential equations in one unknown function $y$. The algebraic study of systems of such differential equations leads to problems concerning differential ideals in $\mathcal{F}\{y\}$. The membership problem is the classical one [13]. It is undecidable for infinitely generated differential ideals in general [3]. On the other hand, for finitely generated ideals it is still open. There is a beautiful algorithmic solution to this problem in the case of radical ideals via decomposition of a radical ideal into finite intersections of characterizable (or regular) components. Further, one can algorithmically check whether a given polynomial belongs to an isobaric ideal [3, 13]. We consider the third case, the case of ideals that have a finite (or parametric) differential standard basis (DSB) w.r.t. some admissible ordering [1, 8, 14]. DSBs are natural generalizations of Gröbner bases to the case of differential ideals. The difference between DSBs and characteristic sets lies in the reduction process. While the characteristic sets technique involves pseudo-reduction, DSBs need full differential reduction (without initials and separants). As a result, DSBs may be infinite even in very simple cases, but they can be applied to non-radical ideals. The interest in DSBs reappeared after new examples of finite and recursive bases w.r.t. more general admissible orderings had been found by the authors [12]. In fact, the existence of a finite DSB of an ideal depends on admissible ordering. A. Zobnin extended the notion of admissible ordering on differential monomials and introduced special classes of orderings that share common properties w.r.t. differential operations [14]. It turned out that finiteness conditions of DSBs are closely related to the existence of linear and quasi-linear polynomials (or their powers) in the ideal [14]. Moreover, the authors found a link between the existence of a finite DSB and the property

of being radical for ideals generated by at most first-order polynomials. One of the key tools in researching DSBs of certain differential ideals is the ideal of separants, introduced by D. Trushin [11].

# 2 Preliminaries

We give necessary information on Gröbner bases: Fix a polynomial ring $K[x_1, \ldots, x_n]$ over a field $K$. Denote by $\mathbb{M}_n$ the set of all monomials in $x_1, \ldots, x_n$, i.e., the expressions of form $\prod_{i=1}^{n} x_i^{\alpha_i}$, $\alpha_i \geqslant 0$. $\mathbb{M}_n$ is a multiplicative monoid. A term is a monomial with a nonzero coefficient from $K$.

**Definition.** An *(admissible) monomial ordering* on $\mathbb{M}_n$ is a linear order that satisfies the following properties:

1. $M \prec N \implies MP \prec NP \quad \forall\, M, N, P \in \mathbb{M}_n$;
2. $1 \preccurlyeq M \quad \forall\, M \in \mathbb{M}_n$.

**Examples.** Fix an order on variables (for example, $x_1 \prec x_2 \prec \ldots \prec x_n$). Let $M = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ and $N = x_1^{\beta_1} \ldots x_n^{\beta_n}$ be arbitrary monomials in $\mathbb{M}_n$. The following binary relations on $\mathbb{M}_n$ are monomial orderings:

- *Lexicographical* ordering (lex):

$$M \prec_{\text{lex}} N \iff (\alpha_1, \ldots, \alpha_n) \prec_{\text{lex}} (\beta_1, \ldots, \beta_n).$$

- *First by degree, then lexicographical* ordering (deglex):

$$M \prec_{\text{deglex}} N \iff (\deg M, \alpha_1, \ldots, \alpha_n) \prec_{\text{lex}} (\deg N, \beta_1, \ldots, \beta_n).$$

- *First by degree, than reverse lexicographical* ordering (degrevlex):

$$M \prec_{\text{degrevlex}} N \iff (\deg M, \beta_n, \ldots, \beta_1) \prec_{\text{lex}} (\deg N, \alpha_n, \ldots, \alpha_1).$$

It is well-known that a monomial ordering well orders the set $\mathbb{M}$. Fix a monomial ordering $\prec$. For any $f \in K[x_1, \ldots, x_n] \setminus \{0\}$ one can define the *leading monomial* $\operatorname{lm}_\prec f$ and the *leading coefficient* $\operatorname{lcf}_\prec f$.

**Definition.** The set $G$ is called a *Gröbner basis* of the ideal $I \lhd K[x_1, \ldots, x_n]$ w.r.t. $\prec$ if $G \subset I$ and $\forall f \in I, f \neq 0, \quad \exists g \in G : \ \operatorname{lm}_\prec g \mid \operatorname{lm}_\prec f$.

Any ideal in $K[x_1, \ldots, x_n]$ has finite Gröbner basis w.r.t. any monomial ordering, because the ring of polynomials over a field is Noetherian. In the differential polynomial case this is not so.

## 2.1 Matrix Representation of Monomial Orderings

It is well known that any admissible ordering on $\mathbb{M}_n$ can be specified by a matrix $\mathcal{M} \in M_{m,n}(\mathbb{R})$ of the size $m \times n$ (for some $m \geqslant 1$) with zero kernel over $\mathbb{Q}$ and lexicographically positive columns:

$$x_1^{\alpha_1} \ldots x_k^{\alpha_k} \prec x_1^{\beta_1} \ldots x_n^{\beta_n} \iff \mathcal{M} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \prec_{\text{lex}} \mathcal{M} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}.$$

Conversely, any matrix $\mathcal{M}$ possessing such properties specifies a monomial ordering. Such matrices are called *monomial matrices*. By definition, the identity matrix specifies a lexicographical ordering. Surely, the same ordering can be specified by different matrices. For example, a lexicographical ordering can be specified by any nonsingular upper-triangular matrix.

**Example.** The orderings deglex and degrevlex with $x_1 \prec x_2 \prec \ldots \prec x_n$ can be specified by $n \times n$ matrices

$$\begin{pmatrix} 1 & 1 & 1 & \ldots & 1 & 1 & 1 \\ & & & & & & 1 \\ & & & & & 1 & \\ & & & \ldots & & & \\ & & 1 & & & & \\ & 1 & & & & & \\ 1 & & & & & & \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 & 1 & 1 \\ -1 & & & & & & \\ & -1 & & & & & \\ & & -1 & & & & \\ & & & \ldots & & & \\ & & & & -1 & & \\ & & & & & -1 & \end{pmatrix}, \quad \text{respectively.}$$

## 2.2 Rings of Differential Polynomials

We give necessary definitions from differential algebra that can be found in books by Ritt and Kolchin [9, 4].

An *ordinary differential ring* $\mathcal{R}$ is a commutative ring with a derivative operator $\delta$, i.e., a linear operator that satisfies the product rule: $\delta(ab) = \delta a\, b + a\, \delta b$. We put $\Theta := \{\delta^k : k \geqslant 0\}$. An ideal $I$ of $\mathcal{R}$ is called *differential* if $\delta I \subset I$. The intersection of differential ideals is a differential ideal. If $F \subset \mathcal{R}$ then $[F]$ denotes the differential ideal generated by $F$, i.e., the smallest differential ideal containing $F$. It is equal to the intersection of all differential ideals containing $F$.

Let $\mathcal{F}$ be a *field of constants*, i.e., a differential field such that $\delta\mathcal{F} = 0$. We assume that the characteristic of $\mathcal{F}$ is zero. Let $y$ be a *differential indeterminate*. The polynomial ring $\mathcal{F}\{y\} := \mathcal{F}[\Theta y] = \mathcal{F}[y, y_1, y_2, \ldots]$ in infinitely many *differential variables* is called the *ring of differential polynomials* over $\mathcal{F}$. We denote a differential variable $\delta^i y$ by $y_i$. The rule $\delta y_k = y_{k+1}$ makes $\mathcal{F}\{y\}$ a differential ring. We put $\mathcal{R}_k = \mathcal{F}[y_0, y_1, \ldots, y_k]$.

Any *differential monomial* $M \in \mathcal{R}\{y\}$ can be written as $M = \prod_{i=0}^{n} y_i^{\alpha_i}$, where $\alpha_i \geqslant 0$. The *weight* of monomial $M$ is $\operatorname{wt} M := \sum_{i=1}^{n} i\alpha_i$. Any differential polynomial is a finite sum of *differential terms*, i.e., expressions of the form $c \cdot M$, where $c \in \mathcal{F}$ and $M$ is a differential monomial.

We shall often omit the word "differential" for short.

# 3 Differential Standard Bases

Standard bases of differential ideals in rings of differential polynomials are similar to Gröbner bases in polynomial rings. The first generalizations of Gröbner bases in differential algebra appeared independently in the late 1980s in works of G. Carrà Ferro [1, 2] ("differential Gröbner bases") and F. Ollivier [8] ("standard bases of differential ideals"). In the early 1990s the work of E. Mansfield [7] appeared, where the notion of differential Gröbner basis was redefined. Mansfield used pseudo-reduction instead of differential reduction, and her bases extend

the concept of the characteristic set of the ideal. In contrast to Mansfield's construction, bases introduced by Ollivier and Carrà Ferro allow testing membership in the ideal. To avoid confusion, we use the term "differential standard basis" for the bases in the sense of Carrà Ferro and Ollivier.

## 3.1 Admissible Orderings on Differential Monomials

Let $\mathbb{M}$ be the set of all differential monomials of $\mathcal{F}\{y\}$. An *admissible ordering* on differential monomials is a linear order $\prec$ on $\mathbb{M}$, satisfying the following properties:

1. $M \prec N \implies MP \prec NP \quad \forall\, M, N, P \in \mathbb{M}$;
2. $1 \preccurlyeq P \quad \forall\, P \in \mathbb{M}$;
3. $y_i \prec y_j \iff i < j$.

A. Zobnin proved [12] that these properties are sufficient to guarantee that $\prec$ well orders $\mathbb{M}$.

We denote by $\mathcal{R}_k$ the polynomial ring $\mathcal{F}[y, y_1, \ldots, y_k]$. Let $\mathbb{M}_k$ now be the set of all monomials in $y = y_0, y_1, \ldots, y_k$, and $\prec_k$ be the restriction of $\prec$ to $\mathbb{M}_k$. The orderings $\prec_k$ are *concordant*: $\prec_k$ is the restriction of $\prec_{k+1}$ to $\mathbb{M}_k$. The authors proved in [14] that one can choose monomial matrices specifying $\prec_k$ also concordantly in some sense.

**Definition.** The set of monomial matrices $\{\mathcal{M}_k\}$ is called *concordant*, if the matrix $\mathcal{M}_{k-1}$ can be obtained from $\mathcal{M}_k$ by removing the rightmost column and then by removing a row of zeroes, if such a row exists.

**Proposition ([14]).** *If monomial matrices $\mathcal{M}_{k-1}$ and $\mathcal{M}_k$ are concordant then the corresponding orderings $\prec_{k-1}$ and $\prec_k$ are concordant too.* □

**Theorem 1 (Zobnin, [14]).** *Any concordant set of admissible orderings such that $y_k \prec_{k+1} y_{k+1}$ can be specified by a concordant set of matrices.* □

**Corollary 1 (Zobnin, [14]).** *Any admissible ordering on differential monomials can be specified by a concordant set of monomial matrices, or, equivalently, by an "infinite" monomial matrix.* □

*Remark.* By an *infinite monomial matrix* we mean an ordered system of infinite rows with real elements such that for any $k \geqslant 0$ the first $k+1$ columns contain only finitely many non-zero rows. Moreover, this finite set of rows must form a monomial matrix $\mathcal{M}_k$ specifying $\prec_k$. Then $\{\mathcal{M}_k\}$ is a concordant set of matrices. We emphasize that the ordinal type of the ordered system of rows in an infinite monomial matrix may differ from the ordinal types of $\mathbb{N}$ and $\mathbb{Z}$. Of course, any concordant set of monomial matrices, as well as the corresponding infinite matrix, specifies an admissible ordering.

The "infinite" matrix from Corollary 1 is not uniquely determined, but the author introduced [14] the notion of *canonical* infinite matrix specifying the ordering, which is unique.

**Example.** The ordering (lex) is specified by the infinite matrix

$$\begin{pmatrix} \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ & & \dots & & & & 1 & & \dots \\ & & \dots & & & 1 & & & \dots \\ & & \dots & & 1 & & & & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ & 1 & \dots & & & & & & \dots \\ 1 & & \dots & & & & & & \dots \end{pmatrix},$$

or by the concordant set of matrices

$$(1),\ \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ 1 & \mathbf{0} \end{pmatrix},\ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{1} \\ 0 & 1 & \mathbf{0} \\ 1 & 0 & \mathbf{0} \end{pmatrix},\ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ 0 & 0 & 1 & \mathbf{0} \\ 0 & 1 & 0 & \mathbf{0} \\ 1 & 0 & 0 & \mathbf{0} \end{pmatrix},\ \dots$$

(new rows and columns of matrices are printed in bold).

**Example.** The ordering degrevlex (first by degree, then reverse lexicographical) can be specified by the matrix

$$\begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 & 1 & \dots \\ -1 & & & \dots & & & & \dots \\ & -1 & & \dots & & & & \dots \\ & & -1 & \dots & & & & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ & & & \dots & -1 & & & \dots \\ & & & \dots & & -1 & & \dots \\ & & & \dots & & & -1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

The first row in this matrix can be replaced by $(1\ 2\ 3\ 4\ \dots)$, and that gives the new admissible ordering (wt+deg)revlex. Considering the rows of the form

$$\begin{pmatrix} 1 & 1 & 1 & 1 & \dots \\ 0 & 1 & 2 & 3 & \dots \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 1 & 2 & 3 & \dots \\ 1 & 1 & 1 & 1 & \dots \end{pmatrix},$$

one can construct matrices specifying orderings degwtrevlex and wtdegrevlex.

## 3.2 Special Classes of Admissible Orderings

**Definition (Levi, [6]).** Let $p > 0$. A monomial $M = \prod_{i=0}^{k} y_i^{\alpha_i}$ is called an *$\alpha_p$-monomial*, if for all $i$ from 1 to $k$ we have $\alpha_{i-1} + \alpha_i < p$. All other monomials are called *$\beta_p$-monomials*.

It is easy to check that for any $k \geqslant 0$ the polynomial $\delta^k y^p$ contains the unique $\beta_p$-monomial. If $k = ap + b$, where $0 \leqslant b < p$, this $\beta_p$-monomial is equal to $y_a^{p-b} y_{a+1}^b$. Certainly, when $b = 0$, one obtains $y_a^p$.

**Theorem 2 (Levi, [6, Theorem 1.1]).** *For a monomial $P$ one has $P \equiv \sum c_i M_i \mod [y^p]$, where $M_i$ are $\alpha_p$-monomials, and $c_i$ are uniquely determined rational coefficients. No nontrivial linear combination of $\alpha_p$-monomials with rational coefficients is in $[y^p]$.* □

**Definition ([14]).** A polynomial $f$ is called *$\prec$-quasi-linear* if $\deg \mathrm{lm}_\prec f = 1$.

An admissible ordering $\prec$ is called:

- *$\delta$-lexicographical*, if $\mathrm{lm}_\prec \delta M = \mathrm{lm}_{\mathrm{lex}} \delta M\ \forall M \neq 1$;
- *$\beta$-ordering*, if $\mathrm{lm}_\prec \delta^k y^n = \mathrm{lm}_{\mathrm{degrevlex}} \delta^k y^n\ \forall k, n \geqslant 1$ (i.e., the leading monomial of $\delta^k y^n$ is $\beta_n$-monomial);
- *concordant with quasi-linearity*, if the derivative of any $\prec$-quasi-linear polynomial is also $\prec$-quasi-linear.

**Example ([12, 14]).**

- the orderings lex, deglex and wtlex are $\delta$-lexicographical;
- degrevlex, (wt+deg)revlex, degwtrevlex and wtdegrevlex are $\beta$-orderings;
- all these orderings are concordant with quasi-linearity.

## 3.3 Differential Standard Bases

For simplicity, we consider differential standard bases in the differential polynomial ring $\mathcal{F}\{y\}$ with one derivative operator $\delta$. In the cases of partial differential rings or several indeterminates all definitions are similar. Let an admissible ordering $\prec$ be fixed[1].

**Definition.** A set $S \subset I$ is *a differential standard basis* of the differential ideal $I \lhd \mathcal{F}\{y\}$ w.r.t. $\prec$, if $\Theta S$ is an (infinite) algebraic Gröbner basis of the ideal $I$ considered as ideal in $\mathcal{F}[y, y_1, y_2, \ldots]$.

Differential standard bases in $\mathcal{F}\{y\}$ can be viewed as a suitable parametrization of infinite algebraic Gröbner bases in $\mathcal{F}[\Theta y]$. This parametrization is provided by differential operators and is compatible with the differential structure of the ring $\mathcal{F}\{y\}$. Thus, we can work with the single element $f$ instead of the family $\Theta f$.

As in the polynomial case, differential standard bases allow several equivalent definitions. They are based on the notion of *differential reduction* (we shall simply write *reduction* if there are no misunderstandings). A polynomial $f$ *elementary (differentially) reduces* w.r.t. a polynomial $s$, if there exists $k \geqslant 0$ such that $M = \operatorname{lm} \delta^k s$ divides a monomial $Q$ occurring in $f$. Every chain of elementary differential reductions is finite, since the set of all differential monomials is well ordered w.r.t. $\prec$ [12].

**Example (Ollivier, [8]).** The ideal $[y^2+y+1]$ has a finite differential standard basis w.r.t. $\prec_{\text{lex}}$ consisting of two elements: $\{y^2 + y + 1,\, y'\}$.

If a finite differential standard basis of an ideal is known, membership in this ideal can be effectively tested. For example, this is the case for ideals generated by linear polynomials. In the general case, differential standard bases may be infinite. Nevertheless, in some cases (for instance, for radical or isobaric ideals) one can still test membership using other methods [3, 13].

Recently, the authors found abundant beautiful examples of finite differential standard bases, as well as finiteness criteria for such bases under some orderings [14], which are more general than that of Carrà Ferro [2]. Examples of finite and parametric DSBs and the proofs can be found in [11, 14, 15] and at the web page http://shade.msu.ru/~difalg/DSB. These examples have stimulated further research. The following theorems (finiteness conditions for DSBs) have been proven:

[1] Note, that Carrà Ferro and Ollivier used strong additional properties of $\prec$ connected with differentiations.

**Theorem 3 ([15]).** *Let $f$ be a quasi-linear polynomial and $n > 1$. The singleton $\{f^n\}$ is a finite differential standard basis of $[f^n]$ w.r.t. any concordant with quasi-linearity $\beta$-ordering. The ideal $[f^n]$ has no finite differential standard basis w.r.t. any $\delta$-lexicographical ordering.* □

**Theorem 4 (Zobnin, [14]).** *Let an ordering $\prec$ be concordant with quasi-linearity, and $[F]$ contain an $\prec$-quasi-linear polynomial. Then $[F]$ has a finite differential standard basis w.r.t. $\prec$.* □

**Theorem 5 (Zobnin, [14]).** *Let $\prec$ be a $\delta$-lexicographical ordering, and $[F]$ be a proper nonzero differential ideal in $\mathcal{F}\{y\}$. $[F]$ has a finite differential standard basis w.r.t. $\prec$ iff $[F]$ contains an $\prec$-quasi-linear polynomial.* □

**Theorem 6 (Zobnin, [14]).** *If an ideal has a finite differential standard basis w.r.t. a $\delta$-lexicographical ordering, then it has a finite differential standard basis w.r.t. pure lexicographical ordering.* □

**Proposition.** *For any $k \geqslant 0$ there exists an admissible ordering $\prec_{(k)}$ such that for any lexicographically quasi-linear polynomial $f$ of order $k$ and for all $n \geqslant 1$ the ideal $[f^n]$ has a finite differential standard basis w.r.t. $\prec_{(k)}$ consisting only of $f^n$.* □

Ollivier [8] suggested a completion process that returns a minimal differential standard basis if by chance it stops. This process uses criteria for avoiding useless $s$-pairs that are similar to Buchberger criteria. But even if there exists a finite basis, Ollivier's process may not stop. In contrast, the authors suggested and implemented a completion process [14] that returns a reduced differential standard basis w.r.t. a $\delta$-lexicographical ordering if it is finite. This process is based on Theorem 5. It is called "Improved Ollivier Process". Unlike the original Ollivier process, the improved one surely stops if the ideal has a finite standard basis.

We summarize the results in the following corollaries.

**Corollary 2.** *The following are equivalent:*
- *the ideal $I \lhd \mathcal{F}\{y\}$ has the finite lex DSB;*
- *$I$ contains a* lex*-quasi-linear polynomial;*
- *the factoralgebra $\mathcal{F}[y, y_1, y_2, \ldots]/I$ is finitely generated.* □

**Corollary 3.** *The following are equivalent:*
- *the ideal $I \lhd \mathcal{F}\{y\}$ has the finite deglex DSB;*
- *$I$ contains a linear polynomial.* □

**Acknowledgments.** A.Z. thanks the organizers of the International Workshop on Computer Algebra and Differential Equations (CADE-2007). This research was supported by the Russian Foundation for Basic Research, project no. 05-01-00671, and by a joint grant of the Russian Foundation for Basic Research and Academy of Finland (project no. 06-01-90628).

# Bibliography

[1] Carrà Ferro G., *Groebner Bases and Differential Algebra,* Lecture Notes in Computer Science, vol. 356, 141–150, 1989.

[2] Carrà Ferro G., *Differential Gröbner Bases in One Variable and in the Partial Case,* Math. Comput. Model., vol. 25, 1–10, 1997.

[3] Gallo G., Mishra B., Ollivier F., *Some Constructions in Rings of Differential Polynomials,* LNCS, vol. 539, 171–182, 1991.

[4] Kolchin E.R., *Differential Algebra and Algebraic Groups,* 1973.

[5] Kondratieva M.V., Levin A.B., Mikhalev A.V., Pankratiev E.V., *Differential and Difference Dimension Polynomials,* 1999.

[6] Levi H., *On the Structure of Differential Polynomials and on Their Theory of Ideals,* Trans. AMS, vol. 51, 532–568, 1942.

[7] E. Mansfield: *Differential Gröbner Bases*, Ph.D. Thesis, 1991.

[8] Ollivier F., *Standard Bases of Differential Ideals,* Lecture Notes in Computer Science, 304–321, 508, 1990.

[9] Ritt J.F., *Differential Algebra*, 1950.

[10] Sit W.Y., *The Ritt-Kolchin Theory for Differential Polynomials,* in *Differential Algebra and Related Topics*, 2000.

[11] Trushin D.V., *Ideal of Separants in a Ring of Differential Polynomials,* to be published in Journal of Mathematical Sciences, 2007.

[12] Zobnin A.: *Essential Properties of Admissible Orderings and Rankings.* In Contributions to General Algebra 14, pp. 205-221 (2004).

[13] Zobnin A.: *On Testing the Membership to Differential Ideals.* In Proceedings of CASC-2004, St-Petersburg, pp. 485–496 (2004).

[14] Zobnin A.: *Admissible Orderings and Finiteness Criteria for Differential Standard Bases.* In Proceedings of ISSAC-2005, Beijing, China, pp. 365–372 (2005).

[15] Zobnin A.: *Differential Standard Bases under Composition.* To be published in Journal of Mathematical Sciences (2007).

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# On Gröbner Bases over $\mathbb{F}_2$

Vladimir P. Gerdt and Mikhail V. Zinin

*Laboratory of Information Technologies*
*Joint Institute for Nuclear Research*
*141980 Dubna, Russia*
gerdt@jinr.ru mzinin@gmail.com

## 1 Introduction

Over last decades Gröbner bases have been widely used as a universal algorithmic tool to deal with polynomial, differential and difference equations [1]. Applying Gröbner bases one can in particular:

- verify consistency of equations;
- detect the dimension of a solution space;
- convert an initial system of equations into another form more suitable for solution by numerical methods;
- determine some special features such as Lie symmetries and/or the integrability of nonlinear evolution equations.

Recently the need arose to compute Gröbner bases for multivariate polynomial systems over the field $F_2$ with all the variables taking values also from $F_2$. Such systems are of interest for simulation of quantum computation, namely for computing the unitary matrix defined by a quantum circuit [2, 3], and in cryptoanalysis of HFE (hidden function equations) systems [4, 5, 6].

There are a number of computer algebra systems that can compute Gröbner bases over $F_2$: *Maple* [7], *Mathematica* [8], *Singular* [9], *CoCoA* [10] and some others. But all of them exploit extension of the system with the binomials $x_i^2 + x_i$ whose number is equal the number of the polynomial variable $x_i$. In the present paper we try to avoid explicit use of these extra binomials.

### 1.1 Preliminaries

Throughout the paper we use the following notations and definitions:

$\mathbb{X} = \{x_1, \ldots, x_n\}$ is the set of polynomial variables.

$R = \mathbb{K}[\mathbb{X}]$ is a polynomial ring over field $\mathbb{K}$ of characteristic 0.

$R' = \mathbb{F}_2[\mathbb{X}]$ is a polynomial ring over field $\mathbb{F}_2$.

$\mathrm{Id}(F)$ is the ideal generated by polynomial set $F$.

$\tilde{R} = \mathbb{F}_2[\{x_1, \ldots, x_n\} \in \mathbb{F}_2^n]$ is a polynomial ring over field $\mathbb{F}_2$ with variables from $\mathbb{F}_2$. Therefore multiplication of monomials in this ring reads:

$$m_1 \cdot m_2 = x_1^{i_1} \cdots x_n^{i_n} \cdot x_1^{j_1} \cdots x_n^{j_n} = x_1^{\max(i_1, j_1)} \cdots x_n^{\max(i_n, j_n)}.$$

**Definition 1.1.** *Admissible monomial order.* A linear monomial order $\succ$ is called *admissible* if the conditions

$$(i) \;\; m \neq 1 \Longleftrightarrow m \succ 1, \quad (ii) \;\; m_1 \succ m_2 \Longleftrightarrow m_1 m \succ m_2 m$$

hold for any monomials $m, m_1, m_2 \in \mathbb{M}$.

**Definition 1.2.** *Monomial quotient in* $\tilde{R}$. A monomial $m$ will be called a quotient for division of monomial $b$ by monomial $a$ in the ring $\tilde{R}$ if $m \cdot a = b$ and $\deg(m) = \min\{\deg(m_i) \mid m_i \otimes a = b\}$.

**Definition 1.3.** *Gröbner basis.* Given an ideal $\mathcal{I} \subset R(R', \tilde{R})$ and an order $\succ$, a finite subset $G \subset R(R', \tilde{R})$ is called *Gröbner basis* of $\mathcal{I}$ if

$$(\forall f \in \mathcal{I})(\exists g \in G) \;\; [\mathrm{lm}(g) \mid \mathrm{lm}(f)],$$

where $u \mid v$ denotes divisibility of monomial $v$ by monomial $u$.

For further notations and definitions we refer to [11, 12].

# 2 Implemented algorithms

## 2.1 Buhberger's algorithm

We implemented the following algorithm as a version of Buchberger's algorithm [13] for the purpose of preliminary testing of data structures and for experimental comparison with the Involutive algorithm [11, 12] whose version for $\tilde{R}$ is considered below. Unlike the Involutive, the Bushberger algorithm deals with all $S$-polynomials (critical pairs). At the initialization step in our version of Buchberger's algorithm an autoreduction of the input basis is done (step 1) and all possible $S-$polynomials for the autoreduced initial basis are collected at step 2 in set $B$. Here $k$ is the number of polynomials in $G$.

Then a **while**-loop is started at step 3. Here a pair from the set $B$ is selected at step 4 and the two Buchberger criteria [13] are verified (step 5). If both criteria fail the normal form of the selected $S$-polynomial is computed (step 6). The non-vanishing normal form is added to the basis at step 8 and the set of $B$ is upgraded to include new $S-$polynomials. The loop is terminated with a Gröbner basis [1] in $G$. Its autoreduction at step 12 gives the reduced Gröbner basis as the algorithm output.

**Buchberger's Algorithm($F$)**

**Input:** $F \in R(R') \setminus \{0\}$ – finite polynomial set, $\succ$ – monomial order
**Output:** $G$ – autoreduced Gröbner bases of Id($F$)

```
 1: G :=AutoReduce(F)
 2: k := card(G); B := {[i, j] : 1 ≤ i < j ≤ k}
 3: while B ≠ ∅ do
 4:    [i, j] :=SelectPair(B, G); B := B \ {[i, j]}
 5:    if criterion1([i, j], G) and criterion2([i, j], B, G) then
 6:       h :=NormalForm(Spoly(G_i, G_j), G)
 7:       if h ≠ 0 then
 8:          G := G ∪ {h}; k := k + 1; B := B ∪ {[i, j] : 1 ≤ i < k}
 9:       fi
10:    fi
11: od
12: return AutoReduce(G)
```

There are some functions (subalgorithms) used in Buhberger' algorithm. The *AutoReduce* function performs autoreduction of its argument. The *NormalForm* function computes a normal form of the given polynomial modulo a given set. The *SelectPair* function implements a selection strategy for $S-$polynomials for further processing in the **while**-loop. Efficiency of the whole algorithm substantially depends on this function. In our implementation we choose the normal selection strategy [13]: the $S-$polynomial is selected with the minimal least common multiple (lcm) w.r.t. the ordering $\succ$. As a monomial order $\succ$ we implemented degree-reverse-lexicographical.

## 2.2 Involutive algorithm

This is our main interest. The Involutive algorithm [11, 12] constructs an involutive basis which is generally a redundant Gröbner basis. In doing so, the reduced Gröbner basis is a well-defined subset of the involutive basis and can be extracted from the last basis as shown in line 15 of the algorithm. Correctness of step 15 is proved in [12]. We implemented the Involutive algorithm for the Janet monomial division [11]. For this reason we call the algorithm *Janet Division Based Involutive Algorithm.*

**Janet Division Based Involutive Algorithm**$(F, \succ)$

**Input:** $F \subset R(R') \setminus \{0\}$ – finite set of polynomials, $\succ$ – monomial order
**Output:** $G$ – reduced Gröbner bases of $\mathrm{Id}(F)$

```
1:  F:=AutoReduce(F)
2:  choose f ∈ F such that lm(f) = min{lm(F)}
3:  T := {f, f, ∅}; Q := {{q, q, ∅} | q ∈ F \ {f}}
4:  while Q ≠ ∅ do
5:     choose p ∈ Q such that lm(pol(p)) = min{lm(pol(Q))}
6:     Q := Q \ {p}; h :=JanetNormalForm(p, T)
7:     if h ≠ 0 then
8:        T := T ∪ {h, anc(p), nmp(p)}
9:        for all q ∈ T and x ∈ NM_J(q, T) \ nmp(q) do
10:          Q := Q ∪ {{pol(q) · x, anc(q), ∅}}
11:          nmp(q) := nmp(q) ∩ NM_J(q, T) ∪ {x}
12:       od
13:    fi
14: od
15: return {pol(f) | f ∈ T : anc(f) = f}
```

At the initial steps 1-3 the conventional autoreduction is done, and a polynomial with the minimal leading monomial is chosen. This polynomial is inserted in set $T$ whereas the remaining polynomials are inserted in set $Q$. Both sets $T$ and $Q$ consist of triples. Each triple includes a polynomial itself (pol), the ancestor [12] of this polynomial (anc), and a (possibly empty) subset of non-multiplicative variables for the polynomial (nmp).

Then the **while-**loop starts. At step 5 a triple with the minimal leading monomial of its polynomial is chosen from the set $Q$. Then this polynomial is Janet-reduced modulo set $T$. If its Janet normal form $h$ is nonzero, then the corresponding new triple is inserted in $T$ at step 8. The insertion causes the upgrade of the set $Q$ done at steps 10 and 11. After termination of the loop, a reduced Gröbner basis is extracted from $T$ as the final step (step 15).

The *AutoReduce* function is just the same as in the Buhberger algorithm. The *JanetNormalForm* function computes the Janet normal form [11, 12] of a polynomial in a given triple modulo polynomials in triples of a given set $T$. In the *JanetNormalFrom* function we verify four involutive criteria (see [12]) which in aggregate are equivalent [14] to the Buchberger criteria.

# 3 Peculiarities of implementation

## 3.1 Polynomials in $\tilde{R}$

By definition, a polynomial in ring $\tilde{R}$ is a sum of monomials whose coefficients and degrees of variables belong to the field $F_2$. All operations over such polynomials are performed in accordance with the field operations in $F_2$. As the following simple example of multiplication by a variable shows, in $\tilde{R}$ the lex-

icographical or degree-compatible monomial orders are not admissible if one follows Definition 1.1.

$$\text{Addition: } (xy+x+1)+(xy+y+1)=x+y$$
$$\text{Multiplication: } (xy+x+1)\cdot y=xy+xy+y=y$$

Furthermore, any monomial order $\succ$ in the ring $\tilde{R}$ is not admissible, as is easily seen in the following: Let $m_1, m_2 \in \tilde{R}$ be two different monomials satisfying $m_1 \succ m_2$, and let $m_3 \in \tilde{R}$ be the third monomial such that $m_3 = \mathrm{lcm}(m_1, m_2)$. Then we obtain $m_1 \cdot m_3 = m_2 \cdot m_3$ that contradicts the second admissibility condition $(ii)$ in Definition 1.1. This fact can be explicitly verified:

$$\begin{aligned}
&m_1 := x_1^{i_1}\cdots x_n^{i_n},\ m_2 := x_1^{j_1}\cdots x_n^{j_n},\ m_1 \succ m_2\,,\\
&m_3 := \mathrm{lcm}(m_1,m_2) = x_1^{\max(i_1,j_1)}\cdots x_n^{\max(i_n,j_n)}\,,\\
&m_1\cdot m_3 = x_1^{\max(i_1,\max(i_1,j_1))}\cdots x_n^{\max(i_n,\max(i_n,j_n))} = m_3\,,\\
&m_2\cdot m_3 = x_1^{\max(j_1,\max(i_1,j_1))}\cdots x_n^{\max(j_n,\max(i_n,j_n))} = m_3\,.
\end{aligned}$$

Therefore, one has to check carefully the applicability of the Involutive algorithm to the ring $\tilde{R}$ .

One more unusual feature of $\tilde{R}$ : unlike the rings $R$ and $R'$, a basis consisting of a single polynomial may not be a Gröbner basis. We give a simple example.

$$\langle xy+x+1\rangle = \langle x+1, y\rangle$$

If one multiplies polynomial $xy+x+1$ by all monomials in the bivariate ring $\tilde{R}$ – there are 15 of them – it becomes obvious that the Gröbner basis consists of two polynomials: $x+1$ and $y$. Therefore, neither Buchberger's nor the Involutive algorithms can be straightforwardly applied to the ring $\tilde{R}$ .

## 3.2 Maximal vectorization

On the other hand, in ring $\tilde{R}$ we ask if it is possible to use the maximal vectorization for the monomial inner data structures. By maximal vectorization we mean an opportunity to store exponents of the monomial variables by using one bit only for each variable. Here we use bit massives of lengths 32, 64 and so on depending on the number of variables under consideration: up to 32, up to 64, etc. Thus, the corresponding variable has the degree one if and only if this variable occurs in the monomial. From the programming point of view exploitation of the maximal vectorization is a useful means to speed up computation.

An example of the monomial representation based on maximal vectorization reads:

$$x_0x_1x_4x_5 = \underbrace{\langle 00000000000000000000000000110011\rangle}_{32(64,96,128)bits}$$

Thus, we are faced with two problems:

1. How to compute a Gröbner basis (Definition 1.3) in the ring $\tilde{R}$ where known algorithms cannot be straightforwardly applied.
2. How to use the maximal vectorization in the computation.

### 3.3 Ideals in rings $\tilde{R}$ and $R'$: a solution to problem 1

It is easy to prove the following theorem.

**Theorem 3.1.** Let $F := \{\tilde{f}_1, \dots \tilde{f}_m\}$ be a finite set of polynomials in $\tilde{R}$. By using the natural homomorphism of $\tilde{R}$ into $R'$ consider the polynomial set $F' := \{f_1, \dots, f_m\} \cup B \subset R'$ where $f_i$ is an image of $\tilde{f}_i$ and $B = \{x_1^2 + x_1, \dots, x_n^2 + x_n\}$. Then if $G' := \{\tilde{g}_1, \dots, \tilde{g}_k\} \cup B_1$ is a Gröbner basis set of $\mathrm{Id}(F')$ with $B_1 \subseteq B$ and with the polynomials $g_i$ $(1 \leq i \leq k)$ having their variables in degree at most one, then $G := \{g_1, \dots, g_k\}$ where each $g_i \in \tilde{R}$ is a preimage of $\tilde{g}_i$ is the Gröbner basis of $\mathrm{Id}(F) \subset \tilde{R}$.

Therefore, one can implicitly work in ring $\tilde{R}$ by applying the above algorithms to the elements in ring $R'$ rather then in $\tilde{R}$ . In doing so, though some operations in $R'$ look identical to the corresponding operations in $\tilde{R}$ , they are still performed in the ring $R'$.

### 3.4 Using maximal vectorization: a solution to problem 2

With respect to both the initial and resulting polynomials belonging to ring $\tilde{R}$ - and, hence having no variables with degree more then 1 - one can implement operations in the algorithms to avoid the presence in intermediate polynomials of variables with degree two and higher. Having all binomials included in the initial basis, one can perform reductions in such a way as to provide every variable with degree 0 or 1. For instance,

$$\text{Elementary reductions: } xy \cdot yz = xy^2z \longmapsto_{y^2+y} xyz \Longrightarrow xy \cdot yz = xyz$$

On the one hand, all binomials $x_i^2 + x_i$ cannot be stored in a computer memory by using the maximal vectorization. On the other hand, after performing elementary reductions, binomials are still needed for construction of new $S$-polynomials in Buhgerger's algorithm and for separation of variables into multiplicative and non-multiplicative in the Involutive algorithm. At present we have managed to surmount this obstacle only for Buhberger's algorithm.

Instead of $S$-polynomial $Spoly(f_i, x_j^2 + x_j)$ of $f_i \in R'$ and $x_j^2 + x_j$ one can consider the product $p_{ij} := f_i x_j$. It is easy to prove that the image $\tilde{p}_{ij} \in \tilde{R}$ of $p_{ij}$ is the image of $Spoly(f_i, x_j^2 + x_j)$ reduced modulo $x_j^2 + x_j$.

This fact allows us to use the maximal vectorization.

## 4 Computer experiments

We compared the running time for implementation of the above-described algorithms with some other computer algebra systems and packages implementing computation of Gröbner bases over $F_2$. In particular, we compared our timings with those for *CoCoA* 4.6, *Singular* 3.0.2 and *FGb* 1.34 [15] library for Maple. All these software packages were running on the 2xOpteron-242 (1.6GHz) computer

with 6Gb of RAM under Gentoo Linux 2005.1. Both of our implementations were built with gcc-4.1.0 compiler.

As benchmarks we took some of the serial examples in the collections [16] widely used for Gröbner bases software. Those of the serial benchmarks having no variables with degrees higher than 1 were used. They are *cyclic, eco, red-cyclic, redeco, noon* and *katsura*. In addition to these famous series, a new one was taken from [17]. We designate these benchmarks by *life*, since they were derived [17] from analysis of the famous Game of Life by J.Conway. Every example of the new series consists of the single polynomial in variables $x_0, \ldots, x_i$ of the form:

$$x_i + x_{i-1}(\sigma_{i-2} + \sigma_{i-3} + \sigma_3 + \sigma_2) + \sigma_{i-2} + \sigma_3$$

where $\sigma_k$ is the $k-$th symmetric polynomial of variables $x_0, \ldots, x_{i-2}$.

Some of the timings obtained are depicted in Figures 1-3.

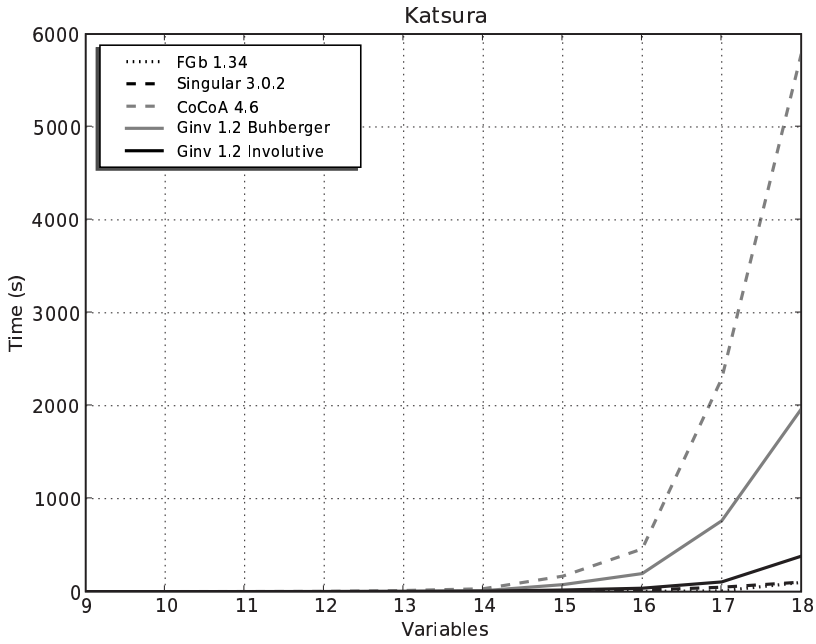



Figure 1: Timings for the *katsura* benchmarks

# 5 Effectivenes of involutive criteria

We also experimentally analyzed the role of criteria [12] in our implementation of the Involutive algorithm. Table 1 shows the timings for some benchmarks

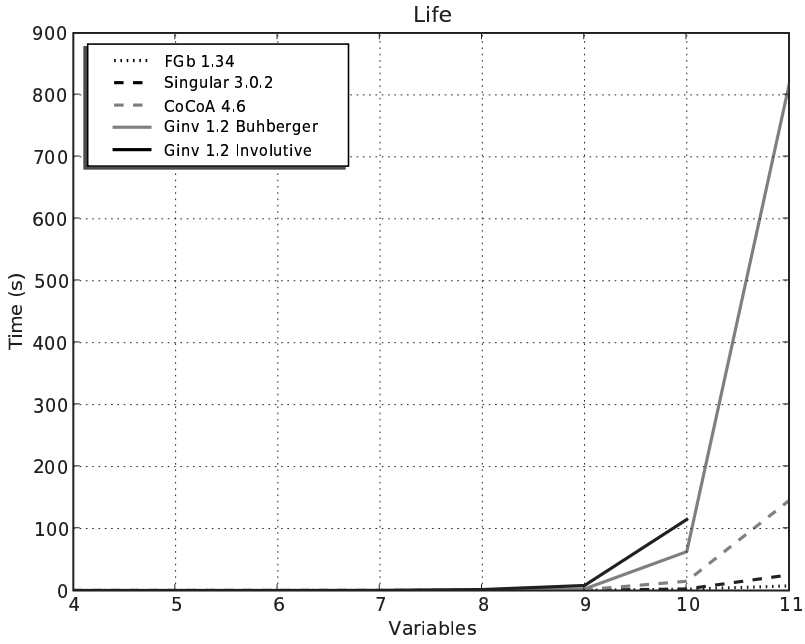



Figure 2: Timings for the *life* benchmarks

under application of different criteria. The row C0 corresponds to the Involutive algorithm without application of any criteria; C1 - with the 1st criterion applied, C1+C2 - with the 1st criterion applied and, if it fails, the 2nd criterion applied; and so on. As one can see, the effect of the criteria application, except example *life10*, is insignificant. We analyzed the action of criteria for many other examples, and only in a few cases did their application lead to a substantial speeding-up of computation. This experimental observation is in agreement with our earlier experimentation with the role of criteria in computation over $\mathbb{Q}$ [12].

Table 1: Avoidance of useless critical pair in involutive completion

| Example | Number of variables | C0 | C1 | C1+C2 | C1+C2+ +C3 | C1+C2+ +C3+C4 |
|---|---|---|---|---|---|---|
| life9 | 10 | 6.77 | 6.92 | 5.06 | 6.58 | 7.85 |
| life10 | 11 | 651.27 | 667.76 | 100.92 | 112.88 | 114.86 |
| kats17 | 18 | 122.47 | 108.90 | 100.27 | 100.92 | 101.21 |

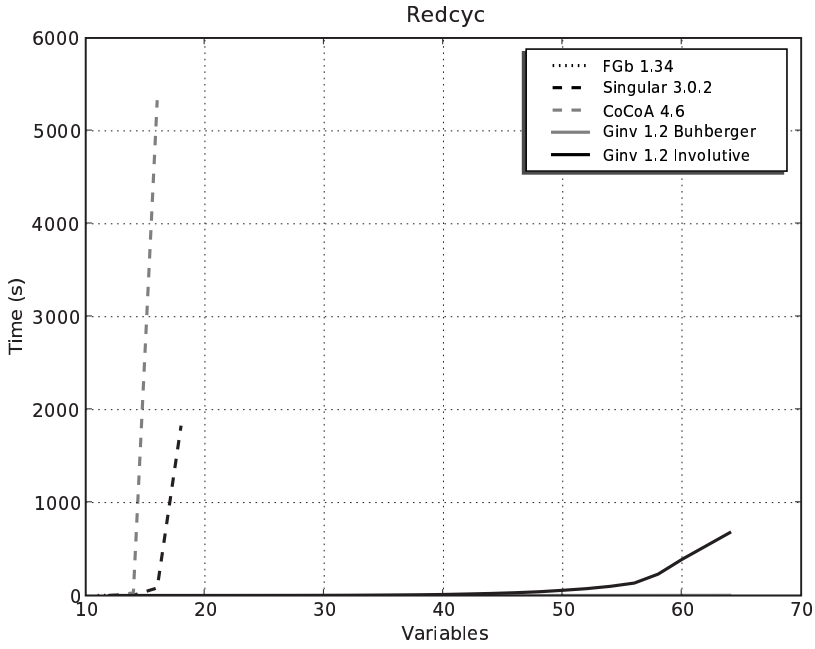



Figure 3: Timings for the *redcyc* benchmarks

# 6 Future work

We are going to work to improve the version of the Involutive algorithm and its implementation presented above for constructing Gröbner bases in the ring $\tilde{R}$. In particular, we plan the following:

- Improvement of the inner data structures. In the present implementation we use bitsets for storing monomials and one-way lists for polynomials, but there may be more suitable structures.
- Optimization of some functions. For example, the very often-used operation of multiplication of a polynomial by a monomial consumes a substantial part of CPU resources and should be optimized.
- Search for heuristically better selection strategies for non-multiplicative prolongations.
- Investigation of applicability to the ring $\tilde{R}$ of involutive divisions different from the Janet one (Pommaret division, etc.).

**Acknowledgements.** The research presented in this paper was partially supported by grant 07-01-00660 from the Russian Foundation for Basic Research and by grant 5362.2006.2 from the Ministry of Education and Science of the Russian Federation.

# References


[1] Buchberger B. and Winkler F. (Eds.): *Gröbner Bases and Applications*, Cambridge University Press, 1998.

[2] Dawson C.M., Nielsen, M.A. et al.: *Quantum computing and polynomial equations over the finite field* $Z_2$. arXiv:quant-ph/0408129.

[3] Gerdt V.P., Kragler R. and Prokopenya A.N.: *A Mathematica Package for Construction of Circuit Matrices in Quantum Computation*, this volume.

[4] Levy-dit-Vehel F. and Ludovic Perret L.: *Polynomial equivalence problems and applications to multivariate cryptosystems*. Report No. 5119, INRIA, 2004. http://www.inria.fr/rrrt/rr-5119.html

[5] Faugère J.-C.: *Algebraic cryptanalysis of HFE using Gröbner bases*. INRIA report: RR- 4738, 2003. http://www.inria.fr/rrrt/rr-4738.html

[6] Faugère J.C. and Joux A.: Algebraic cryptanalysis of Hidden Field Equations (HFE) Using Gröbner Bases. LNCS 2729, Springer-Verlag, 2003, pp. 44–60.

[7] `http://maplesoft.com/`

[8] `http://www.wolfram.com/`

[9] Greuel G.-M., Pfister G. and Schönemann H. *Singular 3.0. A Computer Algebra System for Polynomial Computations*. Centre for Computer Algebra, University of Kaiserslautern, 2005. `http://www.singular.uni-kl.de`

[10] `http://cocoa.dima.unige.it/`

[11] Gerdt V.P. and Blinkov Yu.A.: Involutive Bases of Polynomial Ideals. *Mathematics and Computers in Simulation* 45 (1998) 519–542, arXiv:math.AC/9912027; Minimal Involutive Bases. Ibid., 543–560, arXiv:math.AC/9912029.

[12] Gerdt V.P.: Involutive Algorithms for Computing Gröbner Bases. In: *Computational Commutative and Non-Commutative Algebraic Geometry*, S.Cojocaru, G.Pfister and V.Ufnarovski (Eds.), NATO Science Series, IOS Press, 2005, pp. 199-225. arXiv:math.AC/0501111.

[13] Buchberger B.: Gröbner Bases: an Algorithmic Method in Polynomial Ideal Theory. In: *Recent Trends in Multidimensional System Theory*, N.K. Bose (ed.), Reidel, Dordrecht, 1985, pp. 184–232.

[14] Apel J. and Hemmecke R.: Detecting unnecessary reductions in an involutive basis computation, RISC Linz Report Series 02-22, 2002.

[15] `http://fgbrs.lip6.fr/salsa/Software/`

[16] `http://www-sop.inria.fr/saga/POL`
`http://www.math.uic.edu/~jan/demo.html`

[17] Kornyak V.V.: On Compatibility of Discrete Relations. *LNCS 3718*, Springer-Verlag, 2005, pp.272-284. arXiv:math-ph/0504048.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# POTHMF, a Program to Compute Matrix Elements of the Coupled Radial Equations for a Hydrogen-like Atom in a Homogeneous Magnetic Field

O. Chuluunbaatar[1], A.A. Gusev[1], V.P. Gerdt[1],
V.A. Rostovtsev[1], T.V. Tupikova[1], S.I. Vinitsky[1],
A.G. Abrashkevich[2], M.S. Kaschiev[3], and V.V. Serov[4]

[1] *Joint Institute for Nuclear Research, Dubna, Moscow region, Russia,*
[2] *DB2 RAS/PD Development, IBM Toronto Lab,*
*8200 Warden Avenue, Markham, ON L6G 1C7, Canada*
[3] *Institute of Mathematics and Informatics, Sofia, Bulgaria*
[4] *Saratov State University, Saratov, Russia*

**Abstract.** We describe POTHMF, a program to compute matrix elements of the coupled radial equations for a hydrogen-like atom in a homogeneous magnetic field. POTHMF computes with a prescribed accuracy the oblate angular spheroidal functions, which depend on a parameter and corresponding eigenvalues, and the matrix elements, which are integrals of the eigenfunctions multiplied by their derivatives with respect to the parameter. The program, implemented in Maple-Fortran, consists of a package of symbolic-numerical algorithms that reduce a singular two-dimensional boundary value problem for an elliptic second-order partial differential equation to a regular boundary value problem for a system of second-order ordinary differential equations using the Kantorovich method.

## 1 Introduction

The calculation of the dynamics of electron states of hydrogen-like atoms in a magnetic field in atomic physics is reduced to a boundary value problem for an elliptic second-order partial differential equation in a two-dimensional region for fixed values of the magnetic number and parity [1]. Efficient algorithms for the numerical solution of this problem are based on its reduction to a system of ordinary differential equations by the Kantorovich method, using the oblate angular spheroidal functions [2] as the basis for the expansion of the unknown solution. For an efficient application of the Kantorovich method we elaborate the POTHMF program as a set of symbolic-numerical algorithms for computing the following quantities to a prescribed accuracy [3]:

• oblate angular spheroidal functions on a bounded interval of the parameter values,

• derivatives with respect to the parameter of the angular functions and of the matrix elements (integrals of the eigenfunctions multiplied by their derivatives with respect to the parameter),

• asymptotics of the radial parameter of the eigenfunctions and of the matrix elements that appear as variable coefficients in the system of ordinary differential equations,

• asymptotics of the solutions to the system of ordinary differential equations for small and large values of the radial variable,

• solutions of the boundary value problem for the system of second-order ordinary differential equations.

The program also calculates asymptotic regular and irregular matrix solutions of the system of second-order ordinary differential equations at the end of interval in the radial variable needed for solving the corresponded boundary problem with third-type boundary conditions.

## 2 The problem statement

The Schrödinger equation for the hydrogen atom in an axially symmetric magnetic field $\mathbf{B} = (0, 0, B)$ in spherical coordinates $(r, \eta = \cos\theta, \varphi)$ can be written as the 2D-equation in the region $\Omega = \{0 < r < \infty, -1 < \eta < 1\}$ [1]

$$\left(-\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{\hat{A}^{(0)}(r,\eta)}{r^2} - \frac{2Z}{r} - \epsilon\right)\Psi(r,\eta) = 0. \tag{1}$$

The operator $\hat{A}^{(0)}(r,\eta) = A^{(0)}(r,\eta) + \gamma m r^2$, where $A^{(0)}(r,\eta)$ is given by

$$A^{(0)}(r,\eta) = -\frac{\partial}{\partial\eta}(1-\eta^2)\frac{\partial}{\partial\eta} + \frac{m^2}{1-\eta^2} + \left(\frac{\gamma r^2}{2}\right)^2(1-\eta^2). \tag{2}$$

Here $m = 0, \pm1, \ldots$ is the magnetic quantum number, $\gamma = B/B_0$, $B_0 \cong 2.35 \times 10^5 T$ is a dimensionless parameter which determines the field strength $B$, and the atomic units ($a.u.$) $\hbar = m_e = e = 1$ are used under the assumption of infinite mass of the nucleus having charge $Z$. In these expressions, $\epsilon = 2E$ is the doubled energy (in Rydbergs, 1Ry=(1/2) $a.u.$) of the state $|m\sigma\rangle$ at fixed values of $m$ and $z$-parity; $\sigma = \pm1$; $\Psi \equiv \Psi_{m\sigma}(r,\theta) = (\Psi_m(r,\theta) + \sigma\Psi_m(r,\pi-\theta))/\sqrt{2}$ is the corresponding wave function. The wave functions $\Psi_i^{Em\sigma}(r,\eta)\exp(\imath m\varphi)/\sqrt{2\pi}$ with fixed parity $\sigma$ and azimuthal quantum number $m$ is expanded over the one-dimensional basis, $\Phi_j^{m\sigma}(\eta;r)$,

$$\Psi_i^{Em\sigma}(r,\theta) = \sum\nolimits_{j=1}^{j_{\max}} \Phi_j^{m\sigma}(\eta;r)\chi_j^{(m\sigma i)}(E,r),$$

with unknown radial vector-solutions $\chi_j^{(m\sigma i)}(E,r)$. The basis functions are solutions of the eigenvalue problem for the angular oblate spheroidal functions[2]

$$\hat{A}^{(0)}(r,\eta)\Phi_j(\eta;r) = E_j(r)\Phi_j(\eta;r), \quad I_{ij}(r) = \int_{-1}^{1}\Phi_i(\eta;r)\Phi_j(\eta;r)d\eta = \delta_{ij}. \tag{3}$$

Thus, the Schrödinger equation in the 2D-region, $\Omega = \{R^1_+ \times [-1,1]\}$, is reduced to a set of coupled differential equations

$$\left(-\mathbf{I}\frac{1}{r^2}\frac{d}{dr}r^2\frac{d}{dr} + \frac{\mathbf{U}(r)}{r^2} + \mathbf{Q}(r)\frac{d}{dr} + \frac{1}{r^2}\frac{d\,r^2\mathbf{Q}(r)}{dr}\right)\boldsymbol{\chi}^{(i)}(r) = \epsilon_i\,\mathbf{I}\boldsymbol{\chi}^{(i)}(r). \quad (4)$$

The matrix of effective potentials $\mathbf{U}(r)$ and $\mathbf{Q}(r)$ of $j_{\max} \times j_{\max}$ are given by

$$U_{ij}(r) = \frac{E_i(r)+E_j(r)}{2}\delta_{ij} - 2Zr\delta_{ij} + r^2 H_{ij}(r), \quad I_{ij} = \delta_{ij}, \quad (5)$$

$$H_{ij}(r) = \int_{-1}^{1} \frac{\partial\Phi_i(\eta;r)}{\partial r}\frac{\partial\Phi_j(\eta;r)}{\partial r}d\eta, \quad Q_{ij}(r) = -\int_{-1}^{1}\Phi_i(\eta;r)\frac{\partial\Phi_j(\eta;r)}{\partial r}d\eta.$$

The wave function $\boldsymbol{\chi}(r) = \{\chi_j^{(m\sigma i)}(E,r)\}_{j=1}$ satisfies the following boundary conditions at $r \to 0$

$$\lim_{r\to 0} r^2\left(\mathbf{I}\frac{d}{dr} - \mathbf{Q}(r)\right)\boldsymbol{\chi}(r) = 0, \quad (6)$$

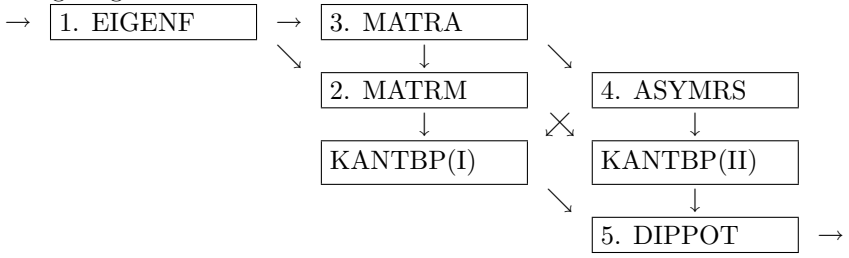
and at large $r = r_{\max} \gg 1$

$$\boldsymbol{\chi}(r) = 0, \quad \text{for the discrete spectrum,} \quad (7)$$

$$\left(\mathbf{I}\frac{d}{dr} - \mathbf{Q}(r)\right)\boldsymbol{\chi}(r) = \mu(r)\boldsymbol{\chi}(r), \quad \text{for the continuous spectrum.} \quad (8)$$

Note, the energy $\epsilon \equiv \epsilon(r_{\max})$ plays the role of eigenvalues of the boundary problem (4)–(7) on a finite interval $0 \le r \le r_{\max}$, while the unknown parameters $\mu \equiv \mu(r_{\max}, \epsilon)$ at fixed value of $\epsilon$ play the role of eigenvalues of the logarithmic normal derivative matrix of the solution of the boundary problem (4)–(6), (8).

To reduce the system of radial equations to the finite interval $r \in (0, r_{\max})$ with homogeneous boundary conditions of the third type, symbolic algorithms for evaluating asymptotics of the effective potentials and the solutions of radial equations at small and large values of $r$ are elaborated. The resulting system of radial equations, which contains the first-derivative coupling terms, is solved using the finite element method by means of the KANTBP program, implemented in FORTRAN. POTHMF calculates energy values, the reaction matrix and unknown radial vector-solutions, and the photoionization cross-sections. POTHMF prepares input files for KANTBP and has the structure given by the following diagram:



In procedure EIGENF, the eigenvalue problem for one-dimension differential equation is reduced to the algebraic eigenvalue problem, which is solved for a finite set of values of parameter $r$.
In procedure MATRM, using solutions of the algebraic eigenvalue problem above, the parametric derivatives of the basis functions, $\partial\Phi_j(\eta;r)/\partial r$ and matrix elements $Q_{ij}(r)$, $H_{ij}(r)$, $\partial Q_{ij}(r)/\partial r$, etc, are calculated.
In procedure MATRA, the asymptotic solutions of the algebraic eigenvalue problem generated in EIGENF as well as its matrix elements are calculated as a power series in the parameter $r^2$ and its inverse.
In procedure ASYMRS, using asymptotics of matrix elements, the asymptotics of the fundamental radial solutions at small and large values $r$ are calculated and the needed boundary conditions for a reduced interval $[0, r_{max}]$ are generated.
In procedure DIPPOT, the transition matrix elements are evaluated using the results of program KANTBP.

# 3 The procedure EIGENF

In procedure EIGENF, the eigenvalue problem for a one-dimensional differential equation is reduced to the algebraic eigenvalue problem, which is solved for a finite set of values of parameter $r$.

We obtain eigenfunctions $\Phi_j(\eta;\hat{r})$ in the form of a series expansion at fixed values $\sigma = \pm 1$ and $m$,

$$\Phi_j(\eta;\hat{r}) = \sum_{s=(1-\sigma)/2}^{s_{\max}} c_{sj}^{m\sigma}(\hat{r}) P_{|m|+s}^{|m|}(\eta). \tag{9}$$

Here, $s$ is an even (odd) integer at $\sigma = (-1)^s = \pm 1$ until $s_{\max} = 2(N_{\max} - 1) + (1-\sigma)/2$, where $N_{\max}$ is the number of even or odd terms of expansion, and $P_{|m|+s}^{|m|}(\eta)$ are the normalized associated Legendre polynomials defined by the relation [2]:

$$-\frac{d}{d\eta}(1-\eta^2)\frac{d}{d\eta}P_{|m|+s}^{|m|}(\eta) + \frac{m^2}{1-\eta^2}P_{|m|+s}^{|m|}(\eta) = \lambda_s^{m\sigma}(0)P_{|m|+s}^{|m|}(\eta), \tag{10}$$

$$\lambda_s^{m\sigma}(0) = (|m|+s)(|m|+s+1), \quad s = 2(j-1) + (1-\sigma)/2,$$

$$\int_{-1}^{1} P_{|m|+s}^{|m|}(\eta) P_{|m|+s'}^{|m|}(\eta) d\eta = \delta_{ss'}. \tag{11}$$

The coefficients $c_{sj}^{m\sigma}(\hat{r})$ satisfy the relation

$$\sum_{s=(1-\sigma)/2}^{s_{\max}} c_{sj}^{m\sigma}(\hat{r}) c_{sj'}^{m\sigma}(\hat{r}) = \delta_{jj'}. \tag{12}$$

The eigenvalue problem for eigenvectors $\mathbf{c}_j = \{c_{sj}^{m\sigma}(\hat{r})\}_{(1-\sigma)/2}^{s_{\max}}$, and eigenvalues $\lambda_j(\hat{r})$ take the form

$$\mathbf{A}^{(0)}\mathbf{c}_j = \lambda_j(\hat{r})\mathbf{c}_j, \tag{13}$$

where matrix $\mathbf{A}^{(0)}$ is the symmetric tridiagonal $N_{\max} \times N_{\max}$ matrix:

$$A^{(0)}_{ss-2} = A^{(0)}_{s-2s} = \frac{-p^2}{(2s+2|m|-1)}\sqrt{\frac{(s-1)s(s+2|m|-1)(s+2|m|)}{(2s+2|m|-3)(2s+2|m|+1)}},$$
$$A^{(0)}_{ss} = (s+|m|)(s+|m|+1) + 2p^2\frac{(s^2+s+2s|m|+2m^2+|m|-1)}{(2s+2|m|-1)(2s+2|m|+3)}. \quad (14)$$

The expansion (9), which provides stability of the numerical calculation with double precision arithmetic (the relative machine precision is $eps \approx 2 \cdot 10^{-16}$), was implemented using the subroutine DSTEVR from the LAPACK Fortran Library [4]. The orthogonality relations (12) were computed with an accuracy of the order of *eps*.

### 3.1 Finding the optimal value of $s_{\max}$ and the matching point $R_{\text{match}}$ of numerical and asymptotic solutions

At large $s$ elements of matrix $\mathbf{A}^{(0)}$ (14) take the form

$$A^{(0)}_{ss} = \frac{(2s+2|m|+1)^2-1}{4} + \frac{p^2}{2} + O(s^{-2}), \quad A^{(0)}_{ss\pm2} = -\frac{p^2}{4} + O(s^{-2}). \quad (15)$$

On intervals $s \in (s_b, s_e)$ at $s_b, s_e \gg 1$, we suppose that the elements of matrix $\mathbf{A}^{(0)}$ have slow dependence on $s$. Therefore, for a given value of $\lambda$, the solution of the algebraic problem (13), (15) will be represented in the form

$$c_s = xc_{s+2}, \quad c_{s-2} = xc_s. \quad (16)$$

From (16), (13), and (15) we have the following algebraic equation with respect to factor $x$

$$x + x^{-1} = d \equiv p^{-2}\left((2s+2|m|+1)^2 - 1 - 4\lambda + 2p^2\right). \quad (17)$$

For $s > s_2$, where $s_2 = (\sqrt{4\lambda+1} - 2|m| - 1)/2$ is determined from equation (17) at $d = 2$, equation (17) has two real solutions. One of them,

$$x_s = p^{-2}\left(\sqrt{(s-s_2)(s+s_2+2|m|+1)} + \sqrt{p^2+(s-s_2)(s+s_2+2|m|+1)}\right)^2, \quad (18)$$

exceeds unity by its absolute value and the other, $x^{-1}$, is smaller than one. Thus, the solution of (16) with decreased coefficients $c_s$ at increased $s$ exists. For $s < s_2$ we have two solutions with oscillating coefficients $c_s$. Then, the solution of Eq. (17) allows us to determine an algorithm for evaluating $s_{\max}$:

$$\prod_{s=s_2}^{s_{\max}-1} x_s < 1/eps, \qquad \prod_{s=s_2}^{s_{\max}} x_s > 1/eps. \quad (19)$$

We need an approximate value of the eigenvalue $\lambda$ for the above calculation. If we use the fact all diagonal elements $A^{(0)}_{ss}$ of the tridiagonal matrix $\mathbf{A}^{(0)}$ and

eigenvalues $\varepsilon_j(p)$ or $\lambda_j(p)$ increased by the number $j$, then we can obtain the upper bound of the eigenvalue $\lambda_N$ with the help of Wilkinson's shift [5]:

$$shift = G + A^{(0)}_{s_N s_N} + \sqrt{G^2 + (A^{(0)}_{s_N s_N - 2})^2}, \quad G = \frac{A^{(0)}_{s_N - 2 s_N - 2} - A^{(0)}_{s_N s_N}}{2}, \quad (20)$$

where $s_N = 2(N-1) + (1-\sigma)/2$. But $shift \gg \lambda_N$ at $p \gg 1$. In this case we use an asymptotic expression of the eigenvalue (29) at $p \geq 2s_N$, since the asymptotic expression gives an upper bound of the eigenvalue.

The matching point $R_{\rm match}$ of the numerical and asymptotic solution is calculated by the MATRA algorithm as follows:

$$r_{\rm match} = \max(r_\varepsilon, r_h, r_q),\ r_\varepsilon = \sqrt[18]{\frac{|\varepsilon^{(18)}_N|}{eps}},\ r_h = \sqrt[18]{\frac{|H^{(18)}_{NN}|}{eps}},\ r_q = \sqrt[17]{\frac{|Q^{(17)}_{NN-1}|}{eps}}, \quad (21)$$

since $|\varepsilon^{(2k)}_j| < \gamma|\varepsilon^{(2k+2)}_j|$, $|Q^{(2k-1)}_{jj'}| < \gamma|Q^{(2k+1)}_{jj'}|$, $|H^{(2k)}_{jj'}| < \gamma|H^{(2k+2)}_{jj'}|$ and $|Q^{(17)}_{jj'}| \leq |Q^{(17)}_{NN-1}|$, $|H^{(18)}_{jj'}| \leq |H^{(18)}_{NN}|$.

# 4 The procedure MATRM

In the procedure MATRM, based on solutions of the above algebraic eigenvalue problem, the parametric derivatives of the basis functions $\partial\Phi_j(\eta;r)/\partial r$ and matrix elements $Q_{ij}(r)$, $H_{ij}(r)$, $\partial Q_{ij}(r)/\partial r$, etc., are calculated.

The derivatives of functions $\Phi_j(\theta;r)$ at fixed values of $\sigma = \pm 1$ and $m$ can be represented as the following expansion in terms of the normalized Legendre polynomials (9):

$$\Phi^{(n)}_j(\theta;r) = \sum_{s=(1-\sigma)/2}^{s_{\max}} c^{(n)}_{sj} P^{|m|}_{|m|+s}(\eta), \quad c^{(n)}_{sj} \equiv \frac{\partial^n c_{sj}(r)}{\partial r^n}, \quad (22)$$

where $\mathbf{c}^{(0)} \equiv \mathbf{c}_j$ and $\lambda^{(0)} \equiv \lambda_j(r)$.

Following (12)–(13), we solve the following linear recurrence system of algebraic equations:

$$(\mathbf{A}^{(0)}\mathbf{c}^{(k)} - \mathbf{c}^{(k)}\lambda^{(0)}) + (\mathbf{A}^{(k)}\mathbf{c}^{(0)} - \mathbf{c}^{(0)}\lambda^{(k)}) = \mathbf{b}_{(k)}, \quad \mathbf{A}^{(k)} \equiv \frac{\partial^k \mathbf{A}^{(0)}}{\partial r^k},$$
$$\mathbf{b}_{(k)} \equiv \sum_{n=1}^{k-1} \frac{k!}{n!(k-n)!}(\mathbf{c}^{(k-n)}\lambda^{(n)} - \mathbf{A}^{(n)}\mathbf{c}^{(k-n)}), \qquad \mathbf{b}_{(1)} \equiv 0. \quad (23)$$

From the normalization condition (12) we obtain the required additional equality

$$\sum_{n=0}^{k} \frac{k!}{n!(k-n)!} \mathbf{c}^{(k-n)^T}\mathbf{c}^{(n)} = 0, \quad (24)$$

providing the uniqueness of the solution (23). The details of the algebraic realization of the algorithm are given in [3].

# 5 The procedure MATRA

In procedure MATRA, the asymptotic solutions of the algebraic eigenvalue problem generated in EIGENF and the matrix elements are calculated as a power series of the parameter $r^2$ and its inverse at small and large values $r$.

At step 1 we go from coordinate $\eta \in [0,1]$ (or $\eta \in [-1,0]$) to the new coordinate $y$ using the formula $y = 2p(1-\eta)$ (or $y = 2p(1+\eta)$).
At steps 2 and 3 we go from the set of functions $\Phi_j(y)$ to the set of functions $F_n(y)$

$$\Phi_j(y) = \exp\left(-\frac{y}{2}\right)\left(\frac{y}{4p}\right)^{|m|/2}\left(1-\frac{y}{4p}\right)^{|m|/2} F_n(y), \tag{25}$$

that are found as a sum of Laguerre polynomials $L_{n+s}^{|m|}(y)$ [2] with unknowns $C_n(s,r)$

$$F_n(y) = 2^{|m|+1/2} p^{(|m|+1)/2} \sum\nolimits_s C_n(s,r) L_{n+s}^{|m|}(y). \tag{26}$$

In step 4, where we evaluate integrals, we change the domain from $[0,2p]$ to $[0,\infty)$, and then drop exponentially small terms. Step 5 finds $C_n(s,r)$ and $\lambda_n$ as a series expansion

$$C_n(s,r) = c_{s,n}^{(0)} + \sum_{k=1}^{k_{\max}} \frac{c_{s,n}^{(k)}}{(4p)^k}, \quad \lambda_n = 4p\left[\frac{|m|+1}{2} + \beta_n^{(0)} + \sum_{k=1}^{k_{\max}} \frac{\beta_n^{(k)}}{(4p)^k}\right]. \tag{27}$$

Substituting (27) in the result of step 3 and equating coefficients at the same powers of $p$, we arrive at a system of recurrence relations for evaluating coefficients $\beta_n^{(k)}$ and $c_{s,n}^{(k)}$ (except $c_{0,n}^{(k)}$):

$$\begin{aligned} s c_{s,n}^{(k)} &= ((n_s+|m|+1)(2n_s+|m|+1) - (n_s+|m|)(|m|+1)) c_{s,n}^{(k-1)} \\ &- n_s(n_s+|m|) c_{s-1,n}^{(k-1)} - (n_s+|m|+1)(n_s+1) c_{s+1,n}^{(k-1)} + \sum_{k'=1}^{k-|s|} \beta_n^{(k')} c_{s,n}^{(k-k')}, \end{aligned} \tag{28}$$

with initial conditions $\beta_n^{(0)} = n$, $c_{s,n}^{(0)} = \delta_{s0}\sqrt{n!/(n+|m|)!}$.
In step 6, substituting (27) the coefficients $c_{s,j}^{(k)}$ evaluated in step 5 for expressions of the matrix elements evaluated in step 4, we easily find the matrix elements as a series expansion of inverse powers of $r$

$$r^{-2}\varepsilon_j(r) = \sum_{k=0}^{k_{\max}} \frac{\varepsilon_j^{(2k)}}{r^{2k}}, \quad H_{jj'}(r) = \sum_{k=1}^{k_{\max}} \frac{H_{jj'}^{(2k)}}{r^{2k}}, \quad Q_{jj'}(r) = \sum_{k=1}^{k_{\max}} \frac{Q_{jj'}^{(2k-1)}}{r^{2k-1}}. \tag{29}$$

# 6 Procedure ASYMRS

In procedure ASYMRS, using asymptotics of matrix elements, the asymptotics of the fundamental radial solutions at small and large values $r$ are calculated and

the needed boundary conditions for a reduced interval $[0, r_{max}]$ are generated. Now let us consider the asymptotic solution

$$\chi_{ji_o}(r) = R(p_{i_o}, r)\phi_{ji_o}(r) + \frac{dR(p_{i_o}, r)}{dr}\psi_{ji_o}(r), \tag{30}$$

where $R(p_{i_o}, r) = p_{i_o}^{-1/2} r^{-1}(\imath\, F_0(p_{i_o}, r) + G_0(p_{i_o}, r))/2$, $F_0(p_{i_o}, r)$ and $G_0(p_{i_o}, r)$ are the Coulomb regular and irregular functions, respectively [2]. The function $R(p_{i_o}, r)$ satisfies the differential equation

$$\frac{d^2R(p_{i_o}, r)}{dr^2} + \frac{2}{r}\frac{dR(p_{i_o}, r)}{dr} + \left(p_{i_o}^2 + \frac{2Z}{r}\right) R(p_{i_o}, r) = 0. \tag{31}$$

Substituting the function (30) into Eq. (4) using (31) and extracting the coefficients for the Coulomb function and its derivative, we arrive at two coupled differential equations with respect to the unknown functions $\phi_{ji_o}(r)$ and $\psi_{ji_o}(r)$. Then we expand the functions $\phi_{ji_o}(r)$ and $\psi_{ji_o}(r)$ in the inverse power series of $r$:

$$\phi_{ji_o}(r) = \sum\nolimits_{k=0}^{k_{\max}} \phi_{ji_o}^{(k)} r^{-k}, \quad \psi_{ji_o}(r) = \sum\nolimits_{k=0}^{k_{\max}} \psi_{ji_o}^{(k)} r^{-k}. \tag{32}$$

After substituting the expansions (29), (32) to the given coupled differential equations and equating the coefficients of the same powers of $r$, we compute a set of recurrence relations with respect to the unknown coefficients $\phi_{ji_o}^{(k)}$ and $\psi_{ji_o}^{(k)}$

$$\begin{aligned}
&\left(p_{i_o}^2 - 2E + \varepsilon_j^{(0)}\right)\phi_{ji_o}^{(k)} - 2p_{i_o}^2(k-1)\psi_{ji_o}^{(k-1)} - (k-2)(k-3)\phi_{ji_o}^{(k-2)} \\
&\qquad -2Z(2k-3)\psi_{ji_o}^{(k-2)} + \sum_{k'=1}^{k}\left(\varepsilon_j^{(k')} + H_{jj}^{(k')}\right)\phi_{ji_o}^{(k-k')} \qquad (33) \\
&= \sum_{j'=1, j'\neq j}^{N}\sum_{k'=1}^{k}\Big[\left((2k-k'-3)Q_{jj'}^{(k'-1)} - H_{jj'}^{(k')}\right)\phi_{j'i_o}^{(k-k')} \\
&\qquad + \left(2p_{i_o}^2 Q_{jj'}^{(k')} + 4ZQ_{jj'}^{(k'-1)}\right)\psi_{j'i_o}^{(k-k')}\Big], \\
&\left(p_{i_o}^2 - 2E + \varepsilon_j^{(0)}\right)\psi_{ji_o}^{(k)} + 2(k-1)\phi_{ji_o}^{(k-1)} - k(k-1)\psi_{ji_o}^{(k-2)} \\
&+ \sum_{k'=1}^{k}\left(\varepsilon_j^{(k')} + H_{jj}^{(k')}\right)\psi_{ji_o}^{(k-k')} \qquad (34) \\
&= \sum_{j'=1, j'\neq j}^{N}\sum_{k'=1}^{k}\left[\left((2k-k'+1)Q_{jj'}^{(k'-1)} - H_{jj'}^{(k')}\right)\psi_{j'i_o}^{(k-k')} - 2Q_{jj'}^{(k')}\phi_{j'i_o}^{(k-k')}\right].
\end{aligned}$$

From the first four equations of the set (33), (34) for $\phi_{i_oi_o}^{(0)}$, $\phi_{j_0i_o}^{(0)}$, $\psi_{i_oi_o}^{(0)}$, $\psi_{j_0i_o}^{(0)}$, we obtain the leading terms of the eigenfunction, the eigenvalue $p_{i_o}^2$, i.e., the initial

data for solving the recurrence equations (33), (34),

$$\phi^{(0)}_{j_0 i_o} = \delta_{j_0 i_o}, \quad \psi^{(0)}_{j_0 i_o} = 0, \quad p^2_{i_o} = 2E - \varepsilon^{(0)}_{i_o}. \tag{35}$$

Open channels have $p^2_{i_o} \geq 0$, and close channels have $p^2_{i_o} < 0$. Suppose that there are $N_o \leq N$ open channels, i.e., $p^2_{i_o} \geq 0$ for $i_o = 1, \ldots N_o$ and $p^2_{i_o} < 0$ for $i_o = N_o + 1, \ldots N$.

In addition, it should be noted that at large $r$ the linearly independent functions (30) satisfy the Wronskian-type relation

$$\mathbf{W}(\mathbf{Q}(r); \boldsymbol{\chi}^*(r), \boldsymbol{\chi}(r)) = \frac{\imath}{2}\mathbf{I}_{oo}, \tag{36}$$

where $\mathbf{W}(\bullet; \boldsymbol{\chi}^*(r), \boldsymbol{\chi}(r))$ is a generalized Wronskian with a long derivative defined as

$$\mathbf{W}(\bullet; \boldsymbol{\chi}^*(r), \boldsymbol{\chi}(r)) = r^2 \left[ (\boldsymbol{\chi}^*)^T \left( \frac{d\boldsymbol{\chi}}{dr} - \bullet\, \boldsymbol{\chi} \right) - \left( \frac{d\boldsymbol{\chi}^*}{dr} - \bullet\, \boldsymbol{\chi}^* \right)^T \boldsymbol{\chi} \right].$$

These relations will be used to examine the accuracy of the above expansion. The calculations of the above asymptotics were performed using MATRA and ASYMRS implemented in MAPLE 8.

# 7 Procedure DIPPOT

In procedure DIPPOT, the transition matrix elements are evaluated using the results of program KANTBP.

Let us construct the longitudinal and transversal dipole matrix elements $D^{(m\sigma\sigma')}_{jj'}(r)$ and $P^{(mm'\sigma)}_{jj'}(r)$ with the photon polarized along the $z$ axis and along the $XOY$ plane, respectively. Using the expression (9), the above matrix elements can be written in the form

$$\begin{aligned}
D^{(m\sigma\sigma')}_{jj'}(r) &= \delta_{|\sigma+\sigma'|0}\, r \sum_{s=s(\sigma)}^{s_{\max}} \sum_{s'=s(\sigma')}^{s_{\max}} c^{m\sigma}_{sj}(r) c^{m\sigma'}_{s'j'}(r) \delta_{|s-s'|1} \frac{\sqrt{s_>}\sqrt{s_>+2|m|}}{\sqrt{4(s_>+|m|)^2-1}}, \\
P^{(mm'\sigma)}_{jj'}(r) &= \delta_{|m-m'|1} \frac{r}{\sqrt{2}} \sum_{s=s(\sigma)}^{s_{\max}} \sum_{s'=s(\sigma')}^{s_{\max}} c^{m\sigma}_{sj}(r) c^{m'\sigma}_{s'j'}(r) \\
&\quad \times \left\{ \delta_{ss'+2} \sqrt{\frac{s(s-1)}{4(s+m_<)^2-1}} - \delta_{ss'} \sqrt{\frac{(s+2m_<+1)(s+2m_<+2)}{(2s+2m_<+1)(2s+2m_<+3)}} \right\},
\end{aligned} \tag{37}$$

where $s(\sigma) = (1-\sigma)/2$, $s_> = \max(s, s')$ and $m_< = \min(|m|, |m'|)$.

Using the coefficients $c^{(k)}_{s,j}$ obtained in sections 3 and 4, one can easily find longitudinal and transversal dipole matrix elements as the series expansion by the inverse power of $r$ without the exponential terms

$$D^{(m\sigma\sigma')}_{jj'}(r) = r \sum_{k=0}^{k_{\max}} r^{-2k} D^{(2k)}_{jj'}, \quad P^{(mm'\sigma)}_{jj'}(r) = -\sum_{k=0}^{k_{\max}} r^{-2k} P^{(2k)}_{jj'}. \tag{38}$$

# 8 Conclusion

Using elaborated algorithms, for large $r$ one could build asymptotic expansions in the inverse powers of $r$ needed for calculation with a given accuracy of the required set of basis functions for all values of parameter $r$ [3]. As a consequence, at large values of the radial variable $r$, the potential curves, radial matrix elements, and dipole transition matrix elements are calculated using asymptotic formulae and matching points $r_{\text{match}} < r_{\max}$ that are found automatically from the interval of integration $0 \leq r \leq r_{\max}$. Thus, we can build a more efficient algorithm for solving the partial algebraic eigenvalue problem depending on parameter $r$ with an automatic choice of Wilkinson's shift [5]. Thus, we give a constructive *solution of the key problem* to build up a nonsymmetric matrix logarithmic derivative, i.e., **R**-matrix in an adiabatic (parametrically dependent) basis in terms of the recalculation matrix for solution of a boundary problem with the boundary conditions of the third type from the inner region to the outer region.
The LONG WRITE UP of POTHMF and KANTBP will be published in Computer Physics Communications.

**Acknowledgments.** The authors (AG and VG) thank the Academy of Finland and RFBR (projects 06-01-90627 and 07-01-00660) for support.

# References

[1] Dimova, M.G., Kaschiev, M.S., Vinitsky, S.I. - The Kantorovich method for high-accuracy calculations of a hydrogen atom in a strong magnetic field: low-lying excited states *J. Phys. B: At., Mol. Phys.* Vol. 38, 2005, 2337-2352.

[2] Abramovits M. and Stegun I.A. - *Handbook of Mathematical Functions.*, Dover, New York, 1965.

[3] Vinitsky, S.I., et. al, - A Symbolic-Numerical Algorithm for the Computation of Matrix Elements in the Parametric Eigenvalue Problem *Programming and Computer Software.* Vol. 33, 2007, 105-116.

[4] http://www.netlib.org/lapack/

[5] Wilkinson, J.H. - Global convergene of tridiagonal QR algorithm with origin shifts, *Linear Algebra and its Applications* Vol. 1, 1968, 409-420.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# Symbolic-numeric Solution of the Two-dimensional Schrödinger Equation with Double-well Potential

I.N. Belyaeva[1], N.A. Chekanov[1], A.A. Gusev[2], A.N. Lukianenko[1], V.A. Rostovtsev[2], S.I. Vinitsky[2]

[1] *Belgorod State University,*
*Studentcheskaja St.14, Belgorod, 308007, Russia*
chekanov@bsu.edu.ru,

[2] *Joint Institute for Nuclear Research,*
*Dubna, Moscow Region 141980, Russia*
vinitsky@thsun1.jinr.ru

**Abstract.** In this work the self-consistent basis method for solving the two-dimensional stationary Schrödinger equation is presented. The solution shows a potential energy surface with two local minima and a unique saddle point. We have developed a symbolic-numeric algorithm and Maple program (SELFA) that realizes the computation. The low part of energy spectrum and corresponding wave functions for the $C_{2v}$ invariant Hamiltonian were also calculated by means of this program.

## 1 Introduction

For solving an eigenvalue problem, in particular the Schrödinger equation, there are many different methods, for example: diagonalization method [1, 2, 3], quasi-classical approaches [4, 5, 6], different variants of perturbation theory [7, 8], finite element method [9], generalized continuous analog of the Newton method [10], normal form methods [11, 12, 13, 14], the so-called $1/N$ expansion [15], oscillator representation method [16], variation and operation methods [17, 18, 19, 20], symplectic method [21].

As is known, when a dimension of the considered system is increased, the complexity of the differential Schrödinger operator for which an eigenvalue problem is solved results in a concomitant increase in numerical difficulties. Besides this, the accuracy of energy spectra and wave functions calculated decreases if the quantum system allows the existence of dynamical chaos in its classical limit [22].

In Ref. [23] the invariant of the two-dimensional polynomial $C_{3v}$ Schrödinger's equation shows that the potential energy surface (PES) has the only minimum. The self-consistent basis method [24, 25] was used to obtain the solution.

In this paper we consider the $C_{2v}$ symmetric Hamiltonian with four parameters. The parameters were chosen in such a way that PES has two local minima and a unique saddle point (see Fig. 1). On the one hand this choice of the PPE results in simplifying the solution of the Schrödinger equation in comparison with that PES that have more than two local minima. On the other hand, this choice results in the possibility that tunneling effects and classical chaos may exist and affect the properties of the energy spectrum and wave functions.

## 2 Main equations

In this report the self-consistent basis method is used to solve the eigenvalue problem for the $C_{2v}$ invariant Schrödinger operator

$$\hat{H}(x, y, \hat{p}_x, \hat{p}_y) = \frac{1}{2}(\hat{p}_x^2 + \hat{p}_y^2) + V(x, y), \tag{1}$$

where the potential energy surface (PES)

$$V(x, y) = \frac{a}{2}(x^2 + y^2) - \frac{a^{'}}{2}x^2 + bx^2y^2 + c(x^2 + y^2)^2, \tag{2}$$

has two local minima for $a = 1.8490$, $a^{'} = 8.257825$, $b = -0.287070$, $c = 0.375509$ (see Fig. 1).

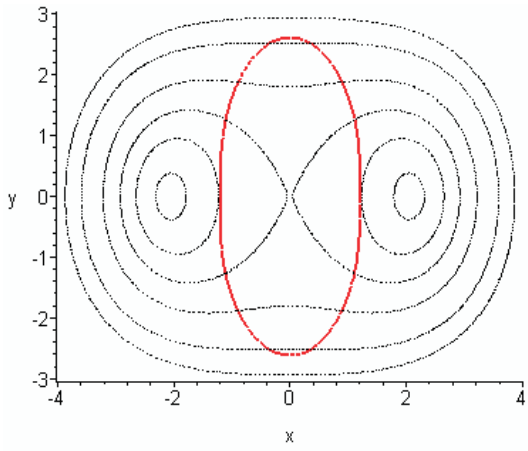



Figure 1: Isolines of the PES (dotted) and the Gaussian curvature zero line (continuous) for Eq. (2).

Expressed as polar coordinates, $x = r\cos\varphi$, $y = r\sin\varphi$, Eqs. (1)-(2) we have

$$\hat{H}(r, \varphi)\psi(r, \varphi) = E\psi(r, \varphi), \tag{3}$$

$$\hat{H}(r, \varphi) = -\frac{1}{2}\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}\right) + \frac{r^2}{2} + \frac{br^3}{2}\sin 3\varphi + cr^4. \tag{4}$$

A regular and bounded solution of the partial eigenvalue problem for Eqs.(3)-(4) can be found in the form of the following Fourier series:

$$u(r,\varphi) = \sqrt{r}\psi(r,\varphi) = \frac{A_0(r)}{2} + \sum_{l=1}^{n}[A_l(r)\cos l\varphi + B_l(r)\sin l\varphi]. \tag{5}$$

Projecting the unknown solution $u(r,\varphi)$ onto basis functions, $\sin l'\varphi$ and $\cos l'\varphi$ ($l' = 1, ..., n$), we have four linear systems of second-order ordinary differential equations (ODE), the consequence of a discrete symmetry, $C_{2V}$, of the Hamiltonian (4), which correspond to four irreducible representations:

$$\begin{aligned}
&\mathrm{A}_1: \quad u(r,\varphi) = \frac{A_0(r)}{2} + \sum_{l=1} A_{2l}(r)\cos 2l\varphi,\\
&\mathrm{A}_2: \quad u(r,\varphi) = \sum_{l=1} B_{2l}(r)\sin 2l\varphi,\\
&\mathrm{B}_1: \quad u(r,\varphi) = \sum_{l=1} A_{2l+1}(r)\cos(2l+1)\varphi,\\
&\mathrm{B}_2: \quad u(r,\varphi) = \sum_{l=1} B_{2l+1}(r)\sin(2l+1)\varphi.
\end{aligned}$$

As result we obtained the following infinite system for the second-order differential equations:

$$\begin{aligned}
&\mathrm{A}_1:\\
&A_0'' + \alpha_0 A_0 + 2\beta A_2 + 2\gamma A_4 = 0,\\
&A_2'' + \alpha_2 A_2 + \beta(A_0 + A_4) + \gamma(A_6 + A_2) = 0,\\
&A_l'' + \alpha_l A_l + \beta(A_{l-2} + A_{l+2}) + \gamma(A_{l-4} + A_{l+4}) = 0,\\
&\qquad\qquad l = 4, 6, 8, \ldots,
\end{aligned} \tag{6}$$

$$\begin{aligned}
&\mathrm{A}_2:\\
&B_2'' + (\alpha_2 - \gamma)B_2 + \beta B_4 + \gamma B_6 = 0,\\
&B_4'' + \alpha_4 B_4 + \beta(B_2 + B_6) + \gamma B_8 = 0,\\
&B_l'' + \alpha_l B_l + \beta(B_{l-2} + B_{l+2}) + \gamma(B_{l-4} + B_{l+4}) = 0,\\
&\qquad\qquad l = 6, 8, 10, \ldots,
\end{aligned} \tag{7}$$

$$\begin{aligned}
&\mathrm{B}_1:\\
&A_1'' + A_1\alpha_1 + A_3(\beta + \gamma) + \gamma A_5 = 0,\\
&A_l'' + A_l\alpha_l + \beta(A_{l-2} + A_{l+2}) + \gamma(A_{l-4} + A_{l+4}) = 0,\\
&\qquad\qquad l = 3, 5, 7, \ldots,
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\mathrm{B}_2:\\
&B_1'' + \alpha_1 B_1 + \beta B_3 + \gamma(B_5 - B_3) = 0,\\
&B_3'' + \alpha_3 B_3 + \beta(B_1 + B_5) + \gamma(B_7 - B_1) = 0,\\
&B_l'' + \alpha_l B_l + \beta(B_{l-2} + B_{l+2}) + \gamma(B_{l+4} - B_{l-4}) = 0,\\
&\qquad\qquad l = 5, 7, 9, \ldots,
\end{aligned} \tag{9}$$

where $\alpha_l = 2E - ((4l^2 - 1)/4)r^2 - ar^2 + (a'/2)r^2 - (b/4)r^4 - 2cr^4$, $\beta = (a'/4)r^2$, $\gamma = (b/8)r^4$.

We rewrite linear systems (6)–(9) in the form of a first-order ODE using a transformation of the above functions $A_i(r)$, $B_j(r)$ $(i,j=1,2)$ to new functions $z_k(r)$:

$$\begin{aligned}
&\mathrm{A}_1 : (A_l \to z_{l+1}, A_l' \to z_{l+2}, l = 0,2,4,...),\\
&\mathrm{A}_2 : (B_l \to z_{l-1}, B_l' \to z_l \quad , l = 2,4,6,...),\\
&\mathrm{B}_1 : (A_l \to z_l \quad , A_l' \to z_{l+1}, l = 1,3,5,...),\\
&\mathrm{B}_2 : (B_l \to z_l \quad , B_l' \to z_{l+1}, l = 1,3,5,...),
\end{aligned}$$

and we obtain the following system:

$$\begin{aligned}
&\mathrm{A}_1 :\\
&z_{l+1}' - z_{l+2} = 0, \qquad l = 0,2,4,...,\\
&z_2' + \alpha_0 z_1 + 2\beta z_3 + 2\gamma z_5 = 0,\\
&z_4' + \alpha_2 z_3 + \beta(z_1 + z_5) + \gamma(z_7 + z_3) = 0,\\
&z_l' + \alpha_l z_{l+1} + \beta(z_{l-1} + z_{l+3}) + \gamma(z_{l-3} + z_{l+5}) = 0,\\
&\qquad\qquad l = 4,6,8,\ldots,
\end{aligned} \tag{10}$$

$$\begin{aligned}
&\mathrm{A}_2 :\\
&z_{l-1}' - z_l = 0, \qquad l = 2,4,6,...,\\
&z_2' + \alpha_2 z_1 + \beta z_3 + \gamma(z_5 - z_1) = 0,\\
&z_4' + \alpha_4 z_3 + \beta(z_1 + z_5) + \gamma z_7 = 0,\\
&z_l' + \alpha_l z_{l-1} + \beta(z_{l-3} + z_{l+1}) + \gamma(z_{l-5} - z_{l+3}) = 0,\\
&\qquad\qquad l = 6,8,10,\ldots,
\end{aligned} \tag{11}$$

$$\begin{aligned}
&\mathrm{B}_1 :\\
&z_l' - z_{l+1} = 0, \qquad l = 1,3,5,...,\\
&z_2' + \alpha_1 z_1 + \beta z_3 + \gamma(z_3 + z_5) = 0,\\
&z_{l+1}' + \alpha_l z_l + \beta(z_{l-2} + z_{l+2}) + \gamma(z_{l-4} + z_{l+4}) = 0,\\
&\qquad\qquad l = 3,5,7,\ldots,
\end{aligned} \tag{12}$$

$$\begin{aligned}
&\mathrm{B}_2 :\\
&z_l' - z_{l+1} = 0, \qquad l = 1,3,5,...,\\
&z_2' + \alpha_1 z_1 + \beta z_3 + \gamma(z_5 - z_3) = 0,\\
&z_4' + \alpha_3 z_3 + \beta(z_1 + z_5) + \gamma(z_7 - z_1) = 0,\\
&z_{l+1}' + \alpha_l z_l + \beta(z_{l-2} + z_{l+2}) + \gamma(z_{l+4} + z_{l-4}) = 0,\\
&\qquad\qquad l = 5,7,9,\ldots.
\end{aligned} \tag{13}$$

Truncating the system obtained (10)–(13) up to the $2Neq$ equations, we cast it as a finite homogeneous system of linear first-order ODE in the unknown functions $z_k(r)$, where the function $\alpha_l$ consists of an arbitrary eigenvalue $E$.

To solve the obtained eigenvalue problem numerically, one needs to define $2Neq$ initial condition data. From the general theory of linear ODE it is known that its general solution has the form of a linear combination:

$$Z_j = \sum_{k=1}^{2Neq} C_k z_j^{(k)}, \tag{14}$$

spanned the $2Neq$ linear-independent solutions

$$z^j = \{z_1^j, z_2^j, ..., z_{2Neq}^j\}, (j = 1, 2, ..., 2Neq), \tag{15}$$

for any system (10)–(13).

To calculate unknown coefficients $C_k$ in Eq. (14), one needs to take into account appropriate boundary conditions placed on functions $A_i(r)$ and $B_j(r)$ that correspond to the general solution with odd index, i.e. $Z_{2j-1}(r)$ $(j = 1, 2, ..., Neq)$ . Then conditions $Z_{2j-1}(0) = 0$ and $Z_{2j-1}(\infty) = 0$ lead to a system of algebraic equations with respect to unknown coefficients $C_k$. A nontrivial solution of this system is calculated by setting to zero the corresponding determinant, $D(E) = |z_j^k(E)| = 0$. Roots of this determinant give us the low part of the energy spectrum $E = E_j, (j = 1, 2, 3, \ldots, Neq)$ for the two-dimensional Schrödinger equation (3)-(4) under consideration. For given $E_j$ we also construct the corresponding eigenfunction.

# 3 Results

Grounded in the self-consistent basis method [23, 24, 25, 26, 27], a symbolic-numeric algorithm and Maple program have been developed to solve the Schrödinger equation (3) with Hamiltonian (4).

For all four state types $A_1$, $B_1$, $A_2$, $B_2$, the energy spectrum $E_n$ and wave functions were calculated. In the Table below, values of the lowest energy levels are presented, while in Figure 2 the structure of its localization in double-well potential is shown. In Figures 3 and 4 some profiles and isolines of the wave functions are plotted.

Table. The energy spectrum of the Hamiltonian (1)-(2).

| $n$ | $E_n$ | Type | $n$ | $E_n$ | Type | $n$ | $E_n$ | Type |
|---|---|---|---|---|---|---|---|---|
| 0 | -3.898 809 | $A_1$ | 9 | 1.346 141 | $B_2$ | 18 | 4.365 241 | $A_2$ |
| 1 | -3.897 242 | $B_1$ | 10 | 1.529 112 | $A_2$ | 19 | 5.300 808 | $B_1$ |
| 2 | -1.423 213 | $B_2$ | 11 | 2.056 203 | $B_1$ | 20 | 5.667 120 | $A_1$ |
| 3 | -1.419 615 | $A_2$ | 12 | 3.242 163 | $B_2$ | 21 | 5.940 909 | $B_2$ |
| 4 | -0.903 803 | $A_1$ | 13 | 3.684 687 | $A_1$ | 22 | 6.293 753 | $B_2$ |
| 5 | -0.807 265 | $B_1$ | 14 | 3.773 432 | $A_1$ | 23 | 6.683 205 | $A_2$ |
| 6 | 1.160 673 | $A_1$ | 15 | 3.918 503 | $B_2$ | 24 | 6.735 355 | $B_1$ |
| 7 | 1.196 685 | $B_1$ | 16 | 3.923 146 | $A_2$ | 25 | 6.740 105 | $A_1$ |
| 8 | 1.217 782 | $A_1$ | 17 | 4.039 048 | $B_1$ | 26 | 6.907 529 | $B_1$ |

# 4 Conclusion and Acknowledgments

A program, SELFA, grounded in the self-consistent basis method, was implemented in MAPLE to solve the two-dimensional Schrodinger equation. The

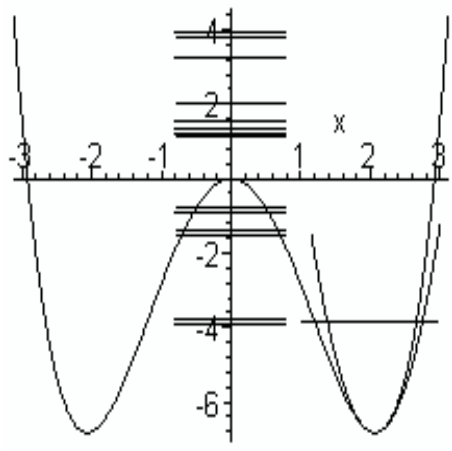



Figure 2: Structure of the energy level localization in double-well potential $V(x, y = 0)$.

efficiency of this program is shown on the C2V symmetric Hamiltonian with two local minimuma, for which the lowest energy levels and wave functions are calculated. Further applications of SELFA include solving the eigenvalue problem for different Hamilton operators, studying tunneling effects, and avoiding crossing phenomena of eigenenergies. An appropriate development of this approach can also be done within the framework of the Kantorovich method using a self-consistent basis with $r$ as a parameter, taking into account the discrete symmetry of Hamiltonian under consideration.

The authors are deeply indebted to Prof. I.V. Puzynin and to the participants of his seminar for a useful discussion. This work is partially supported by BelSU (grant VKG-003-04).

# References

[1] Wilkinson J.H., Reinsch C. - *Handbook for Automatic Computation, Vol.2, Linear Algebra.* Heidelberg New York Springer-Verlag Berlin, 1971.

[2] Golub G.H., Van Loan Ch.F. - *Matrix Computations.* The John Hopkins University Press Baltimore and London, 1989.

[3] Banerjee K. - *General anharmonic oscillators.* Proc. R. Soc. Lond., Vol.A364, 1978, 265-275.

[4] Fröman N., Fröman O. - *JWKB Approximation.* North-Holland Publishing Company, Amsterdam, 1965.

[5] Maslov V.P., Fedoryuk M.V. - *Kvaziklassicheskie pribligheniya dlya uravnenii kvantovoi mekhaniki.* Nauka, Moskva, 1976.

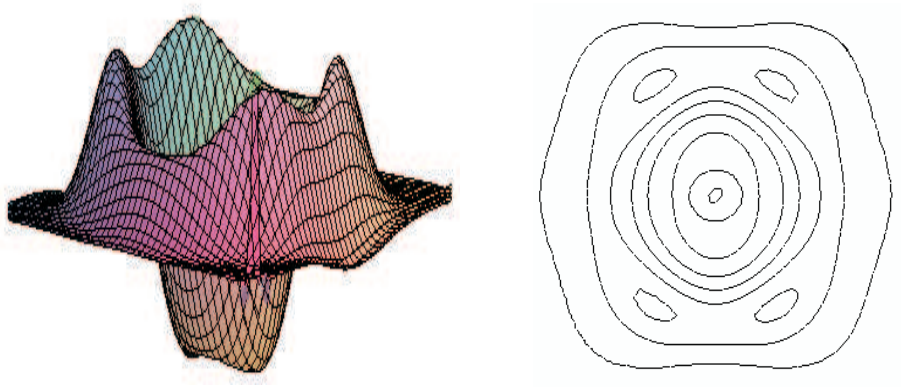


Figure 3: Surface and isolines the wave function of $A_1$-type with energy $E = 1.160673$.

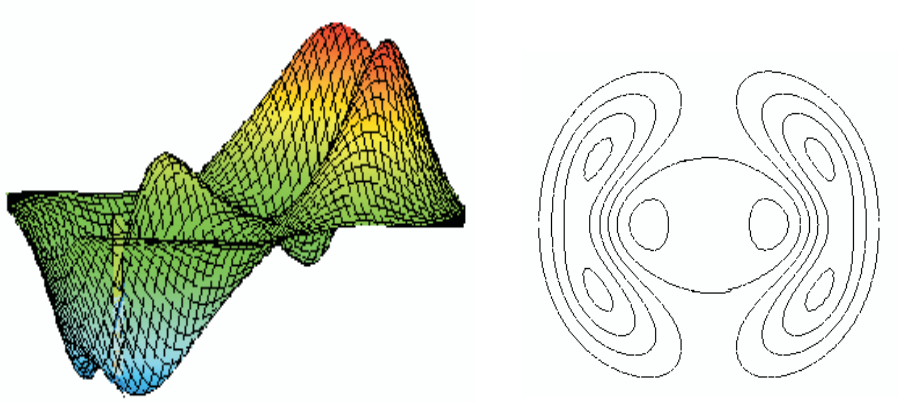


Figure 4: Surface and isolines the wave function of $B_1$-type with energy $E = -0.807265$.

[6] Born M. - *Vorlesungen über atommechanik.* Vol.1, Berlin, 1925.

[7] Bethe H.A. - *Intermediate quantum mechanics.* W. A. Benjamin, inc. New York, Amsterdam, 1964.

[8] Flügge S. - *Practical quantum mechanics.* Vol.1,2, Springer-Verlag, Berlin - Heidelberg - New York, 1971.

[9] Abrashkevich A.G., Abrashkevich D.G., Kaschiev M.S., Puzynin I.V. - *FESSDE, a program for the finite-element solution of the coupled-channel Schroedinger equation using high-order accuracy approximations.* Comp. Phys. Commun., Vol.85, 1995, 65-74.

[10] Puzynin I.V. et.al. - *Obobshennyi nepreryvnyi analog metoda Newtona dlya chislennogo issledovaniya nekotoryh nelineinyh kvantovo-polevyh modelei.* Physics of elementary particles and atomic nuclei. Vol.30(1), 1999, 210-265.

[11] Swimm R.T., Delos J.B. - *Semiclassical calculation of vibrational energy levels for non-separable systems using Birkhoff-Gustavson normal form.* J. Chem. Phys. Vol.71, 1979, 1706-1716.

[12] Ali M.K. - *The quantum normal form and its equivalents.* J. Math. Phys., Vol.26(10), 2565-2572.

[13] Robnik M. - *Algebraic quantization of the Birkhoff-Gustavson normal form.* J. Phys.A: Math.Gen., Vol.17, 1984, 109-130.

[14] Chekanov N.A. - *Kvantovanie normalnoi formy Bikgofa-Gustavsona.* Jadernaya fizika, Vol.50(8), 1989, 344-346.

[15] Jaffe L. - *Large N limits as classical mechanics.* Rev. Mod. Phys. Vol.54, 1982, 407.

[16] Dineykhan M. and Efimov G.V. - *The Schrödinger equation for bound state systems in the oscillator representation.* Reports of Math. Phys., Vol.6(2/3), 1995, 287-308.

[17] Collatz L. - - *Eigenwertaufgaben mit technischen anwendungen.* Akademische verlagsgesellschaft geest & portig K.-G., Leipzig, 1963.

[18] Jafarpour M., Afshar D. - *Calculation of energy eigenvalues for the quantum anharmonic oscillator with a polynomial potential.* J. Phys. A: Math. Gen., Vol.35, 2002, 87-92.

[19] Ivanov I.A. - *Sextic and octic anharmonic oscillator: connection between strong-coupling and weak-coupling expansions.* J. Phys. A: Math. Gen. Vol.31, 1998, 5697-5704.

[20] Ivanov I.A. - *Link between the strong-coupling and weak-coupling asymptotic perturbation expansions for the quartic anharmonic oscillator.* J. Phys. A: Math. Gen. Vol.31, 1998, 6995-7003.

[21] Liu X.S., Su L.W., Ding P.Z. Intern. J. - *Quantum Chem.* 2002, 1-11.

[22] Bolotin Yu.L., Gonchar V.Yu., Tarasov V.N., Chekanov N.A. - *The transition regularity-chaos-regularity and statistical properties of wave function.* Phys. Lett., Vol.A144(8,9), 1990, 459-461.

[23] Belyaeva I.N., Ukolov Yu.A., Chekanov I.N., Ukolov Yu.A. - *Vycheslenie energeticheskogo spectra i volnovyh funkziiobobschennogo gamiltoniana Henon-Heiles metodom samosoglasovanno bazisa.* Vestnik Hersonskogo gosudarstvennogo tekhnicheskogo universiteta, Vol.2(22), 2005, 43-47.

[24] Vinitsky S.I., Inopin E.V., Chekanov N.A. - *Reshenie dvumernogo uravneniya Schredingera v samosoglasovannom basise.* Preprint JINR-P4-93-150, Dubna, 1993, 12.

[25] Chekanov N.A., Ukolov Yu.A. - *Chislennoe reshenie statsionarnogo uravneniya Schredingera v priblizhenii samosoglasovannogo basisa.* International Seminar "Super-computations and computer simulation", Sarov, Russian federal nuclear center – All Russian research institute of experimental physics, 2004, 100-102.

[26] Belyaeva I.N., Ukolov Yu.A., Chekanov N.A. - *Vestnik Kherson. Gosudarst. Techn. Univer.*, Vol.2(22), 2005, 43-47.

[27] Belyaeva I.N., Chekanov N.A, Gusev A.A., Rostovtsev V.A., Ukolov Yu.A., Uwano Y and Vinitsky S.I. - *A MAPLE symbolic-numeric program for solving the 2D-eigenvalue problem by a self-consistent basis method.* Lecture Notes in Computer Science, Vol.3718, 2005, 32-39.

[28] Vinitsky S.I., Gerdt V.P., Gusev A.A., Kaschiev M.S., Rostovtsev V.A., Samoilov V.N., Tyupikova T.V., and Chuluunbaatar O. - *A symbolic-numerical algorithm for the computation of matrix elements in the parametric eigenvalue problem.* Programming and Computer Software, Vol. 33, 2007, 105-116.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# On Compatibility of Systems of PDEs

Valentin Lychagin

*Department of Mathematics,*
*University Tromsø, Norway*

## 1 Introduction

In this work, we review methods for investigating the compatibility (or formal integrability) of systems of partial differential equations (PDEs) based on the jet space approach. They were originated by D. Spencer. To apply them one should, first of all, realize our system, say $E$, as a tower $E^{(k)} \subset J^k(\pi)$, $k = 0, 1, ...$ of subsets (*prolongations*) in the jet-spaces $J^k(\pi)$. Here we meet the first condition for regularity: are all $E^{(k)}$ smooth submanifolds or not? If not, then the simplest case of singularities of vector fields, or 1-st order PDEs, shows how different should be methods for regular or singular cases. So, we'll assume that $E$ is a *regular system of PDEs,* that is, all prolongations $E^{(k)}$ are smooth submanifolds. The second regularity condition arises from the study of compatibility, namely, we investigate the surjectivity of the maps $\pi_{k,k-1} : E^{(k)} \to E^{(k-1)}$. If they are surjective then our system is compatible, that is, we could not get new, "hidden" PDEs from the given ones by algebraic manipulations and derivations. The obstructions to surjectivity are realized in terms of Weyl tensors $W_k$ () which are closely related to metasymplectic structures on jet spaces and which take their values in the 2-nd Spencer cohomology of the PDE system. This cohomology forms a family of vector spaces over $E^{(k)}$, and we arrive at a second, more subtle, condition for regularity: the 2-nd Spencer cohomology should form vector bundles over prolongations $E^{(k)}$. Then the Weyl tensors are sections of these bundles, and the third, the more delicate regularity condition, requires that either $W_k = 0$ identically, or sections $W_k$ are transversal to zero. Then one should reconsider the PDE system $E$ by adding new PDEs $W_k = 0$, and repeat all previous steps. This is a geometrical jet-space interpretation of the *Cartan's method of projections and prolongations.*

The straightforward computations of Weyl tensors are extremely costly (see, for example, [1]). So we propose a different method based on syzygies of the symbolic modules. The method uses Buchsbaum-Rim complexes instead of Spencer $\delta$-complexes and provides us with a constructive form of Weyl tensors. Moreover, for PDEs which we call the *Cohen-Macaulay type,* and especially for *(generalized) complete intersections,* one gets explicit formulae for obstructions in terms of generalized Mayer brackets and multi-brackets.

This paper is based on joint research with Boris Kruglikov.

## 2 Geometry of Compatibility

A *regular system* of PDEs for sections of a smooth bundle $\pi : E(\pi) \to M$ is a tower of smooth bundles

$$\mathcal{E}_{k_1} \to \mathcal{E}_{k_1-1} \to \cdots \to \mathcal{E}_{k_0},$$

where $\mathcal{E}_k$ are subbundles of the jet bundles $\pi_k : J^k(\pi) \to M$,

$$\mathcal{E}_{k+1} \subset (\mathcal{E}_k)^{(1)}$$

and

$$\begin{aligned} \pi_{k+1,k} &: \mathcal{E}_{k+1} \to \mathcal{E}_k, \\ k &= k_0, ...., k_1 - 1, \end{aligned}$$

are smooth bundles too.

The PDEs system is said to be *formally integrable* (or, more concisely, *compatible*) if

$$\pi_{k+1,k} : \mathcal{E}_{k+1} \to \mathcal{E}_k,$$

are smooth bundles for all values of $k$, where

$$\mathcal{E}_{k_1+i} = (\mathcal{E}_{k_1})^{(i)}$$

for all $i \geq 0$.

To check that maps

$$\pi_{k+1,k} : \mathcal{E}_k^{(1)} \to \mathcal{E}_k,$$

are surjective, we introduce *Cartan forms*:

$$\begin{aligned} f \in\ & C^\infty\left(J^k\pi\right) \longmapsto \omega_f \in \Omega^1\left(J^{k+1}\pi\right), \\ & \omega_f = df - \widehat{d}f. \end{aligned}$$

The *Cartan distribution* $\mathcal{C}_k$ on $J^k\pi$ is defined as an annihilator of all Cartan forms $\omega_f$, where $f \in C^\infty\left(J^{k-1}\pi\right)$.

The restriction of the de Rham differential $d\omega_f$ on the Cartan distribution $\mathcal{C}_k$ leads to the *metasymplectic structure*

$$\Omega : f \longmapsto d\omega_f|_{\mathcal{C}_k}$$

on the Cartan distribution:

$$\Omega : S^{k-1}T \otimes \pi^* \to \Lambda^2\mathcal{C}_k^*.$$

For $k = 1, \dim \pi = 1$, $\Omega$ is the canonical symplectic structure on the contact distribution $\mathcal{C}_1$.

In canonical local coordinates in the jet spaces one has the following expressions for the Cartan forms:

$$\begin{aligned} \omega_f &= \sum_{\sigma,j} \frac{\partial f}{\partial u^j_\sigma} \omega^j_\sigma, \\ \omega^j_\sigma &= du^j_\sigma - \sum_i u^j_{\sigma+1_i} dx_i. \end{aligned}$$

The restriction of the Cartan distribution and the metasymplectic structure on a PDEs $\mathcal{E}_k \subset J^k\pi$ leads to the metasymplectic structure

$$\Omega : g^*_{k-1} \to \Lambda^2 \mathcal{C}\left(\mathcal{E}_k\right)^*$$

where $g_{k-1} \subset S^{k-1}T^* \otimes \pi$ is the *symbol* of $\mathcal{E}_{k-1}$ :

$$0 \to g_{k-1} \to T\mathcal{E}_{k-1} \to T\mathcal{E}_{k-2} \to 0.$$

Restriction of the metasymplectic structure on a horizontal subspace, $H$, in the Cartan distribution $\mathcal{C}\left(\mathcal{E}_k\right)$ gives tensors

$$\Omega_H \in g_{k-1} \otimes \Lambda^2 T^*.$$

They are 2-cocycles for the Spencer $\delta$-complex:

$$0 \to g_{k+1} \to g_k \otimes T^* \to g_{k-1} \otimes \Lambda^2 T^* \to g_{k-2} \otimes \Lambda^3 T^* \to \cdots$$

The cohomology classes

$$W_{k-1} \in H^{k-1,2}\left(g\right)$$

do not depend on the choice of the horizontal subspaces $H$ and are called *Weyl tensors (see, [5]).*

**Proposition 1** *A point $x_k \in \mathcal{E}_k$ belongs to the image of the map $\pi_{k+1,k} : \mathcal{E}^{(1)}_k \to \mathcal{E}_k$ if and only if $W_{k-1} = 0$ at this point.*

# 3 Syzygies of Symbolic Modules

The $ST = \oplus_k S^k T$-module, graded by dual symbols $g^*_k$,

$$g^* = \oplus_k g^*_k$$

is called a *symbolic module* (of the system of PDEs at the point of the equation $x_k \in \mathcal{E}_k,\ k \geq k_1$).

The annihilator $I \subset ST$ of the symbolic module is called *a characteristic ideal,* and the corresponding projective variety of $I$-zeros is called *a characteristic variety:*

$$\operatorname{Char} \mathcal{E} \subset \mathbb{P}T^{*\mathbb{C}}.$$

We say that the PDE system is of the *Cohen-Macaulay type* if the symbolic module is a Cohen-Macaulay $ST$-module ( at the point of the equation, $x_k \in \mathcal{E}_k$).

Let the PDE system has form

$$F_1\left(x_1,,,x_n,u^1,..,u^m,...,u^j_\alpha,...\right)=0,\ \cdots,F_N\left(x_1,,,x_n,u^1,..,u^m,...,u^j_\alpha,...\right)=0,$$

in canonical jet coordinates, and let

$$\sigma\left(F_i\right)=\left(\sigma^1\left(F_i\right),...,\sigma^m\left(F_i\right)\right),$$

be the symbol of $i$-th equation, where

$$\begin{aligned}\sigma^j\left(F_i\right) &= \sum_{|\alpha|=k_i}\frac{\partial F_i}{\partial u^j_\alpha}\,\partial^a,\\ \partial^a &= \partial_1^{\alpha_1}\circ\cdots\circ\partial_n^{\alpha_n},\\ \partial_s &= \frac{\partial}{\partial x_s},\ \alpha=(\alpha_1,...,\alpha_n)\end{aligned}$$

and $k_i$ is a order of $i$-th PDE.

Then the symbolic module $g^*$ is the factor of $ST\otimes\mathbb{R}^m$ by submodule $\langle\sigma\left(F_1\right),..,\sigma\left(F_N\right)\rangle$ generated by symbols $\sigma\left(F_1\right),..,\sigma\left(F_N\right)$.

Denote by $\Sigma$ the syzygy $ST$-module for $g^*$. Note that this module as well as $g^*$ depends on a point of the PDE system $E$. We shall require that this modules form a vector bundles over the PDE system (this is the second regularity condition discussed in the introduction).

Under the regularity condition the following theorem holds.

**Theorem 2** *1. Any syzygy* $\lambda=\left(\sum_{|\alpha|=R-k_1}\lambda^1_\alpha\partial^a,....,\sum_{|\alpha|=R-k_N}\lambda^N_\alpha\partial^a\right)$ *produces a compatibility condition:*

$$\begin{aligned}\overline{\lambda} &= \sum_{|\alpha|=R-k_1}\lambda^1_\alpha D^a\left(F_1\right)+\cdots\\ &\quad+\sum_{|\alpha|=R-k_N}\lambda^1_\alpha D^a\left(F_N\right)=0\ \mod\left(D^{\beta^1}\left(F_1\right),...,D^{\beta^N}\left(F_N\right)\right),\end{aligned}$$

*where* $\left|\beta^1\right|<R-k_1,....,\left|\beta^N\right|<R-k_N$, *and* $D^a$ *are the total differential operators corresponding to* $\partial^a$.

*2. The system of PDEs,* $\mathcal{E}$, *is formally integrable (compatible) if and only if* $\overline{\lambda}=0$ *for all syzygies* $\lambda$.

Remark that the obstructions $\overline{\lambda}$ are representatives of the Weyl tensors in the 2-nd Spencer $\delta$-cohomology.

# 4 Compatibility of Scalar PDEs and Mayer Brackets

Applying the description of obstructions for a system $\mathcal{E}$ of scalar PDEs

$$\begin{aligned} &F_1\left(x, u, \frac{\partial u}{\partial x}, ..., \frac{\partial^\sigma u}{\partial x^\sigma}\right) = 0, \\ &..... \\ &F_r\left(x, u, \frac{\partial u}{\partial x}, ..., \frac{\partial^\sigma u}{\partial x^\sigma}\right) = 0 \end{aligned}$$

which satisfy the following condition:

- The equations $F_1 = 0, ...., F_r = 0$ have transversal complex characteristic varieties (or, equivalently: The characteristic ideal is a complete intersection),

we get

**Theorem 3 ([1],[3])** *The PDEs system $\mathcal{E}$ is formally integrable if and only if all* Mayer brackets $[F_i, F_j]$ *vanish due to the system $\mathcal{E}$.*

Here, the (higher) *Mayer bracket* of two functions $F \in C^\infty(\mathbf{J}^k\mathbb{R}^n)$ and $G \in C^\infty(\mathbf{J}^l\mathbb{R}^n)$ is defined as follows

$$[F, G] = \sum_{|\alpha|=k} \frac{\partial F}{\partial u_\alpha} D^\alpha G - \sum_{|\beta|=l} \frac{\partial G}{\partial u_\beta} D^\beta F$$

where $D^\alpha = D_1^{\alpha_1} \cdots D_n^{a_n}$, $\alpha = (\alpha_1, ..., \alpha_n)$, and $D_i$ are the operators of total derivatives.

Remark that

- $[F, G] \in C^\infty(\mathbf{J}^{k+l-1}\mathbb{R}^n)$, if $F \in C^\infty(\mathbf{J}^k\mathbb{R}^n)$ and $G \in C^\infty(\mathbf{J}^l\mathbb{R}^n)$,
- The Mayer bracket is bilinear, skew-symmetric, but
- The Jacobi identity for Mayer brackets is violated.
- If $r = n$ and the above conditions hold then the PDE system $\mathcal{E}$ is solvable in smooth functions.
- If $H$ is a higher symmetry of the PDE system $\mathcal{E}$ such that complex characteristics of $H$ are transversal to complex characteristics of $\mathcal{E}$, then the PDE system $\mathcal{E} \cap H^{-1}(0)$ for $H$-invariant solutions of $\mathcal{E}$ is formally integrable.
- We call a function $G \in C^\infty(\mathbf{J}^l\mathbb{R}^n)$ an *auxiliary integral* for the system $\mathcal{E}$ if all Mayer brackets $[F_i, G]$ vanish due to the system $\mathcal{E} \cap G^{-1}(0)$ and complex characteristics of $H$ are transversal to complex characteristics of $\mathcal{E}$ . In this case the system $\mathcal{E} \cap G^{-1}(0)$ is formally integrable.

**Example 1** *1. The PDE system*

$$\begin{aligned} u_{xx}u_{yy} - u_{xy}^2 &= ae^{bu}, \\ u_{xx} \pm u_{yy} &= 0 \end{aligned}$$

*is formally integrable for any constants* $a, b$.

*2. The PDE system*

$$\begin{aligned} u_{xx} + u_{yy} &= u^{-1}, \\ u_{xx}u_{yy} - u_{xy}^2 &= 0. \end{aligned}$$

*is formally integrable.*

*3. The associativity equation:*

$$u_{xxx}u_{xyy} - u_{xxy}^2 + u_{yyy} = 0$$

*has*

*(a) The 1st order auxiliary integrals*

$$\begin{aligned} &\lambda x u_x + \mu y u_y + (\mu - 4\lambda)u, \\ &xu_x + yu_y - 2u,\ xu_x - u, xu_x + c, u \\ &yu_y - ru, r = 1, 2 \end{aligned}$$

*(b) The 2nd order auxiliary integrals*

$$\begin{aligned} &au_{xx} + bu_{xy} + cu_{yy}, \\ &2u_{xx}u_{yy} - 3u_{xy}^2 \end{aligned}$$

*and*

*(c) The 3rd order auxiliary integrals*

$$\begin{aligned} &au_{xxx} + bu_{xxy} + cu_{xyy} + du_{yyy}, \\ &2u_{xxy} - u_{xxx}^2 \end{aligned}$$

# 5 Compatibility of PDE Systems and Multi-brackets

Let $\mathbf{A}$ be an (non commutative) algebra, and let denote by $\langle \mathbf{a}_1, ...., \mathbf{a}_m \rangle = \mathrm{Ndet} \, \|a_{ij}\| \in \mathbf{A}$ any generalization of the determinant for non commutative algebras for

$$\begin{aligned} \mathbf{a}_1 &= (a_{11}, ..., a_{1m}) \in \mathbf{A}^m, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \mathbf{a}_m &= (a_{m1}, ..., a_{mm}) \in \mathbf{A}^m. \end{aligned}$$

Define a multi-bracket as follows

$$\{\mathbf{a}_1, ...., \mathbf{a}_{m+1}\} = \sum_{k=1}^{m+1} (-1)^k \langle \mathbf{a}_1, ..., \widehat{\mathbf{a}_k}, ..., \mathbf{a}_{m+1} \rangle \mathbf{a}_k.$$

**Example 2** *1. Let $m = 1$. Then*

$$\{a, b\} = ab - ba.$$

*2. Let $m = 2$. Then*

$$\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\} = (x_1, x_2)$$

*where*

$$\begin{aligned} x_1 &= -a_{21}a_{32}a_{11} + a_{31}a_{22}a_{11} + a_{11}a_{32}a_{21} \\ &\quad -a_{31}a_{12}a_{21} - a_{11}a_{22}a_{31} + a_{21}a_{12}a_{31}, \\ x_2 &= -a_{21}a_{32}a_{12} + a_{31}a_{22}a_{12} + a_{11}a_{32}a_{22} \\ &\quad -a_{31}a_{12}a_{22} - a_{11}a_{22}a_{32} + a_{21}a_{12}a_{32}. \end{aligned}$$

Remark that:

- If $\mathbf{A}$ is a commutative algebra, then

  $$\{\mathbf{a}_1, ...., \mathbf{a}_{m+1}\} \equiv 0$$

  for all $\mathbf{a}_1, ...., \mathbf{a}_{m+1}$.

- If $\mathbf{A}$ is an algebra of differential operators, and order of $\mathbf{a}_i \leq k_i$, then order of $\{\mathbf{a}_1, ...., \mathbf{a}_{m+1}\}$ does not exceed

  $$k_1 + \cdots + k_{m+1} - 1.$$

Let $\pi : \mathbb{R}^m \times M \to M$ be the trivial vector bundle and let $\mathrm{dif}_k(\pi, \mathbf{1}) = C^\infty\left(\mathbf{J}^k\pi\right)$ be the module of non-linear scalar differential operators on $\pi$ of order $\leq k$.

Denote by

$$l(F) = (l_1(F), ..., l_m(F))$$

the linearization of differential operator $F \in \mathrm{dif}_k(\pi, \mathbf{1})$.

Define a multi-bracket $\{F_1, ..., F_{m+1}\}$ as follows

$$\frac{1}{m!} \sum_{\alpha \in \mathbf{S}_m, \beta \in \mathbf{S}_{m+1}} (-1)^\alpha (-1)^\beta l_{\alpha_1}(F_{\beta_1}) \cdots l_{\alpha_m}(F_{\beta_m})(F_{\beta_{m+1}}).$$

Note that

$$\{F_1, ..., F_{m+1}\} \in \mathrm{dif}_{k_1 + \cdots k_{m+1} - 1}(\pi, \mathbf{1})$$

if

$$F_i \in \mathrm{dif}_{k_i}(\pi, \mathbf{1}).$$

Let a PDE system $\mathcal{E}$ given by equations

$$F_1 = 0, ..., F_r = 0$$

where $F_i \in \mathrm{dif}_{k_i}(\pi, \mathbf{1})$, and let

$$\mathcal{J}_s = \langle D^\sigma F_i \ \text{ with } \ k_i + |\sigma| \leq s \rangle \subset \mathrm{dif}_s(\pi, \mathbf{1}).$$

Then by the Mayer multi-bracket we mean

$$[F_1, ..., F_{m+1}]_{\mathcal{E}} = \{F_1, ..., F_{m+1}\} \bmod \mathcal{J}_s,$$

where

$$s = k_1 + \cdots + k_{m+1} - 1.$$

We say that the PDE system $\mathcal{E}$ of $r$ differential equations on $m$ unknowns is of *generalized complete intersection type* if

1. $$m \leq r \leq n + m - 1.$$

2. The complex projective characteristic variety has dimension

   $$m + n - r - 2$$

   at each point.

3. The kernel bundle is a 1-dimensional bundle over the characteristic variety.

Remark that:

- this class of systems is included into the class of Cohen-Macaulay type systems introduced in [2].
- The number $r$ of equation, called *codimension of PDE* system $\mathcal{E}$, is defined [2] via the Spencer $\delta$-cohomology as

  $$\dim H^{*,1}(\mathcal{E}).$$

**Theorem 4** *Let a PDE system $\mathcal{E}$ be given by equations*

$$F_1 = 0, ..., F_r = 0$$

*where $F_i \in \mathrm{dif}_{k_i}(\pi, \mathbf{1})$. Then*

1. *If the system is formally integrable, then all Mayer multi-brackets*

   $$[F_{i_1}, ..., F_{i_{m+1}}]_{\mathcal{E}}$$

   *vanish.*

2. *If the system is of generalized complete intersection type, then $\mathcal{E}$ is formally integrable if and only if the all Mayer multi-brackets vanish.*

**Example 3** *The Cauchy-Riemann system*

$$u_x = v_y, \quad u_y = -v_x$$

*has an auxiliary integral of the form*

$$\det \begin{Vmatrix} u_x & v_x \\ u_y & v_y \end{Vmatrix} = G(u, v)$$

*if and only if*

$$\Delta G = \frac{|\nabla G|^2}{G}.$$

# References

[1] B.,S. Kruglikov, V.,V. Lychagin, Mayer brackets and solvability of PDEs – I, Diff. Geom. and its Appl.,17, 251–272 (2002)

[2] B.,S. Kruglikov, V.,V. Lychagin, Mayer brackets and solvability of PDEs – II, Trans. Amer. Math. Soc.,358, no.3, 1077–1103 (2005)

[3] B.,S. Kruglikov, V.,V. Lychagin, A compatibility criterion for systems of PDEs and generalized Lagrange-Charpit method, A.I.P. Conference Proceedings, "Global Analysis and Applied Mathematics: International Workshop on Global Analysis", 729, no. 1, 39–53 (2004)

[4] B.,S. Kruglikov, V.,V. Lychagin, Multi-brackets of differential operators and compatibility of PDE systems, Comptes Rendus Math. 342, no. 8, 557–561 (2006)

[5] V.,V. Lychagin, Homogeneous geometric structures and homogeneous differential equations, AMS Transl., "The interplay between differential geometry and differential equations", ser.2, 167, (1995), 143–164.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# Computation of Positive Roots of Polynomials with Applications to Orthogonal Polynomials

Doru Ştefănescu

*University of Bucharest*
*P. O. Box 39–D5, Bucharest 39*
*Romania*
stef@rms.unibuc.ro

**Abstract.** We consider univariate, nonconstant polynomials $P$ with real coefficients. For computing the real roots of such polynomials it is convenient to isolate the roots. A key step in real isolation of roots is the computation of lower bounds for the real roots. This can be realized as soon as we obtain accurate upper bounds for positive roots. We present some bounds for positive roots and compare them with other methods. We also compute bounds for the roots of orthogonal polynomials.

## 1 Introduction

Orthogonal polynomials have real coefficients and all their zeros are real, distinct, simple, and located in the interval of orthogonality. Therefore, the methods of real root isolation using continued fractions or those based on Descartes' rule of signs can be applied. It is sufficient to estimate the smallest positive root. And this can be done if we are able to compute the largest positive root (LPR).

We obtain new estimates for LPR of orthogonal polynomials using two approaches. We use results of Lagrange, Kioustelidis and Ştefănescu on real roots of polynomials with real coefficients. And, we study the ordinary differential equations satisfied by orthogonal polynomials through the Hessian of Laguerre. The results are compared with other bounds.

## 2 On the Bound of Lagrange

We first recall some results on upper bounds for positive roots of univariate polynomials with real coefficients. The first such bounds were obtained by Lagrange [6] and Cauchy [3]. Other limits were obtained by J. B. Kioustelidis [5] and D. Ştefănescu [10]. These bounds are expressed as functions of the degree and of the coefficients. We compare the bounds of Lagrange [6], Koustelidis [5]

and a bound from our paper [10]. Other comparisons can be found in Akritas–Strzeboński–Vigklas [2] and Ştefanescu [11].

In his treatise on the numerical solution of algebraic equations (1769) Lagrange [6] obtained several bounds for the real roots of univariate polynomials with real coefficients. We remind the reader of two of them.

**Theorem 1 (Lagrange)** *Let $F$ be a nonconstant monic polynomial of degree $n$ over $\mathbb{R}$ and let $\{a_j\,;\,j\in J\}$ be the set of its negative coefficients. Then an upper bound for the positive real roots of $F$ is given by the sum of the largest and the second largest numbers in the set*

$$\left\{\sqrt[j]{|a_j|}\,;\,j\in J\right\}\,.$$

**Theorem 2 (Lagrange)** *Let $P(X)=a_0X^d+\cdots+a_mX^{d-m}-a_{m+1}X^{d-m-1}\pm\cdots\pm a_d\in\mathbb{R}[X]$, with all $a_i\geq 0$, $a_0,a_{m+1}>0$. Let*

$$A\;=\;\max\left\{a_i\,;\,\mathrm{coeff}\,(X^{d-i})\;<\;0\right\}\,.$$

*The number*

$$1\;+\;\left(\frac{A}{a_0}\right)^{1/(m+1)}$$

*is an upper bound for the positive roots of $P$.*

This Theorem 1 result has apparently been completely forgotten. We note that in some particular cases it can be successfully used. On the other hand Theorem 2 is very popular. We give here two results that extend it.

**Theorem 3** *Let $P(X)=a_0X^d+\cdots+a_mX^{d-m}-a_{m+1}X^{d-m-1}\pm\cdots\pm a_d\in\mathbb{R}[X]$, with all $a_i\geq 0$, $a_0,a_{m+1}>0$. We put*

$$A\;=\;\max\left\{a_i\,;\,\mathrm{coeff}\,(X^{d-i})\;<\;0\right\}\,.$$

*The number*

$$1\;+\;\max\left\{\left(\frac{pA}{a_0+a_1+\cdots+a_s}\right)^{1/(m-s+1)}\,,\right.$$
$$\left.\left(\frac{qA}{sa_0+\cdots++2a_{s-2}+a_{s-1}}\right)^{1/(m-s+2)}\right\}$$

*is an upper bound for the positive roots of $P$ for any $s\in\{1,2,\ldots,m\}$ and $p\geq 0\,,q\geq 0$ such that $p+q\;=\;1$.*

**Proof:**

We consider $x \in \mathbb{R}$, $x > 1$. We have

$$\begin{aligned} |P(x)| &\geq |a_0x^d + \cdots + a_mx^{d-m}| - |a_{m+1}x^{d-m-1} \mp + \ldots \mp a_d| \\ &\geq a_0x^d + \cdots + a_sx^{d-s} - A(x^{d-m-1} + \cdots + 1) \\ &\geq (a_0\,x^s + \cdots + a_s)x^{d-s} - A\,\frac{x^{d-m}-1}{x-1} \qquad (1) \\ &= \frac{(a_0\,x^s + \cdots + a_s)(x-1)x^{d-s} - A}{x-1}\cdot\, x^{d-m} + \frac{A}{x-1}\,. \end{aligned}$$

The last right hand side of (1) is strictly positive provided

$$(a_0\,x^s + \cdots + a_s)(x-1)x^{m-s} \;\geq\; A\,. \qquad (2)$$

Now we put $x = 1 + y$ and note that $x^j \geq 1 + jy$ for all $j \in \mathbb{N}$. We observe that

$$\begin{aligned} &(a_0\,x^s + \cdots + a_s)(x-1)x^{m-s} \\ &\geq (a_0(1+sy) + \cdots + a_{s-1}(1+y) + a_0)\; y^{m-s+1} \\ &= (a_0 + \cdots + a_s)\,y^{m-s+1} + (sa_0 + \cdots + 2a_{s-2} + a_{s-1})\,y^{m-s+2}\,. \end{aligned}$$

It follows that (2) is satisfied if

$$\begin{aligned} (a_0 + \cdots + a_{s-1} + a_s)\,y^{m-s+1} &\geq pA\,, \\ (sa_0 + \cdots + 2a_{s-2} + a_{s-1})\,y^{m-s+2} &\geq qA\,, \end{aligned}$$

with $p, q \geq 0,\ p + q = 1$.

These inequalities are satisfied if

$$y \;\geq\; \max\left\{ \left(\frac{pA}{a_0 + a_1 + \cdots + a_s}\right)^{1/(m-s+1)}, \left(\frac{qA}{sa_0 + \cdots + +2a_{s-2} + a_{s-1}}\right)^{1/(m-s+2)} \right\}.$$

□

We also obtain

**Theorem 4** *Let* $P(X) = a_0X^d + \cdots + a_mX^{d-m} - a_{m+1}X^{d-m-1} \pm \cdots \pm a_d \in \mathbb{R}[X]$, *with all* $a_i \geq 0$, $a_0, a_{m+1} > 0$. *We put*

$$A \;=\; \max\left\{a_i\,;\, \text{coeff}\,(X^{d-i}) \;<\; 0\right\}\,.$$

*The number*

$$1 \quad + \quad \max\left\{ \left( \frac{pA}{a_0 + \cdots + a_s} \right)^{1/(m-s+1)} , \right.$$

$$\left( \frac{qA}{sa_0 + \cdots + 2a_{s-2} + a_{s-1}} \right)^{1/(m-s+2)} ,$$

$$\left. \left( \frac{2rA}{s(s-1)a_0 + (s-1)(s-2)a_1 + \cdots + 2a_{s-2}} \right)^{1/(m-s+3)} \right\} .$$

*is an upper bound for the positive roots of* $P$ *for any* $s \in \{2, 3, \ldots, m\}$ *and* $p \geq 0\,, q \geq 0\,, r \geq 0$ *such that* $p + q + r = 1\,.$

**Proof:**
We consider $x \in \mathbb{R}$, $x > 1$. As in the proof of Theorem 3 $P(x) \neq 0$ if

$$(a_0\, x^s + \cdots + a_s)(x-1)x^{m-s} \geq A\,. \tag{3}$$

We consider again $x = 1 + y$ and note that

$$x^j \geq 1 + jy + \frac{j(j-1)}{2} y^2 \qquad \text{for all} \qquad j \geq 2\,.$$

It follows that the inequality (3) is satisfied if the following three conditions are fulfilled:

$$\begin{aligned} (a_0 + \cdots + a_s) y^{m-s+1} &\geq pA\,, \\ (sa_0 + \cdots + 2a_{s-2} + a_{s-1}) y^{m-s+2} &\geq qA\,, \\ (s(s-1)a_0 + (s-1)(s-2)a_1 + \cdots + 2 \cdot 1 \cdot a_{s-2}) y^{m-s+3} &\geq rA \end{aligned}$$

for $p, q, r$ positive with $p + q + r = 1\,.$ □

**Example**
Let $P(X) = X^{15} + X^{13} + X^{12} + 2X^6 - 5X^5 - 3X^2 + 2X - 6$
and denote by $B = B(m, s, p, q, r)$ the bound given by Theorem 4. We have $A = 6$, $m = 9\,.$ We obtain the following bounds

| $s$ | $p$ | $q$ | $r$ | $B$ | LPR |
|---|---|---|---|---|---|
| 2 | 0.5 | 0.5 | 0 | 2.06 | 1.11 |
| 2 | 0.25 | 0.25 | 0.5 | 2.07 | 1.11 |
| 3 | 0.5 | 0.5 | 0 | 2.00 | 1.11 |
| 3 | 0.25 | 0.25 | 0.25 | 2.00 | 1.11 |
| 9 | 0.5 | 0.5 | 0 | 1.84 | 1.11 |
| 9 | 0.25 | 0.25 | 0.5 | 1.67 | 1.11 |

# 3 Another Bound

J. B. Kioustelidis [5] gives the following upper bound for the positive real roots:

**Theorem 5 (Kioustelidis)** *Let $P(X) = X^d - b_1 X^{m_1} - \cdots - b_k X^{m-k} + g(X)$, with $g(X)$ having positive coefficients and $b_1 > 0, \ldots, b_k > 0$. The number*

$$K(P) \;=\; 2 \cdot \max\{b_1^{1/m_1}, \ldots, b_k^{1/m_k}\}\,.$$

*is an upper bound for the positive roots of $P$.*

For polynomials with an even number of variations of sign, we have the following bound:

**Theorem 6 (Ştefănescu)** *Let $P(X) \in \mathbb{R}[X]$ be such that the number of variations of signs of its coefficients is even. If*

$$P(X) \;=\; c_1 X^{d_1} - b_1 X^{m_1} + c_2 X^{d_2} - b_2 X^{m_2} + \cdots + c_k X^{d_k} - b_k X^{m_k} + g(X)\,,$$

*with $g(X) \in \mathbb{R}_+[X]$, $c_i > 0$, $b_i > 0$, $d_i > m_i > d_{i+1}$ for all $i$, the number*

$$S(P) = \max\left\{ \left(\frac{b_1}{c_1}\right)^{1/(d_1-m_1)}, \ldots, \left(\frac{b_k}{c_k}\right)^{1/(d_k-m_k)} \right\}$$

*is an upper bound for the positive roots of the polynomial $P$.*

**Remark**: Note that the bound of Lagrange returns only bounds surpassing unity, so it cannot be used for some classes of orthogonal polynomials. For example the roots of Legendre polynomials are subunitary.

# 4 Applications to Orthogonal Polynomials

Let us consider the polynomials of Laguerre and Chebyshev of first and second kind. Using Theorem 6 we obtain:

**Proposition 7** *Let $L_n$, $T_n$ and $U_n$ be the orthogonal polynomials of degree $n$ of Laguerre, respectively Chebyshev of first and second kind. We have*

*i. The number $K(L_n) = S(L_n) = n^2$ is an upper bound for the roots of $L_n$.*

*ii. The numbers $K(T_n) = \sqrt{n}$ and $S(T_n) = \dfrac{\sqrt{n}}{2}$ are upper bounds for the roots of $T_n$.*

*iii. The numbers $K(U_n) = \sqrt{n-1}$ and $S(U_n) = \dfrac{\sqrt{n-1}}{2}$ are upper bounds for the roots of $U_n$.*

**Proof:**
We use the representations

$$L_n(X) = \sum_{k=0}^{n} \binom{n}{n-k} \frac{(-1)^k}{k!} X^k,$$

$$T_n(X) = \frac{n}{2} \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k\, 2^{n-2k}}{n-k} \binom{n-k}{k} X^{n-2k},$$

$$U_n(X) = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k\, 2^{n-2k} \binom{n-k}{k} X^{n-2k},$$

and Theorems 5 and 6. □

**Examples**:
We denote by LPR the largest positive root of a polynomial and by $L$ the bound of Lagrange (Theorem 2).

1. Laguerre Polynomials

| $n$ | L(P) | K(P) | S(P) | LPR |
|---|---|---|---|---|
| 8 | 376321.0 | 64 | 64 | 22.86 |
| 120 | $1.94 \times 10^{206}$ | 14400 | 14400 | 487.696 |

2. Chebyshev Polynomials of First Kind

| $n$ | L(P) | K(P) | S(P) | LPR |
|---|---|---|---|---|
| 8 | 2.41 | 2.83 | 1.41 | 0.994 |
| 120 | 27917.33 | 10.00 | 5.00 | 0.99991 |

3. Chebyshev Polynomials of Second Kind

| $n$ | L(P) | K(P) | S(P) | LPR |
|---|---|---|---|---|
| 8 | 2.322 | 2.83 | 1.41 | 0.994 |
| 120 | 25864.44 | 9.96 | 4.98 | 0.9996 |

Note that for Chebyshev polynomials we have $K(P) = 2\,S(P)$.

# 5 Applications of the Hessian of Laguerre

A second approach for estimating the largest positive root of an orthogonal polynomial is the consideration of properties of its associated differential equation. This can be realized by considering the positivity of the Hessian of Laguerre

$$\mathrm{H}(f) = (n-1)^2 f'^2 - n(n-1) f f'.$$

We discuss applications to Legendre and Hermite polynomials.

## Definition of the Hessian of Laguerre

Let us suppose that

$$f(X) \,=\, \sum_{j=1}^{n} a_j \, X^j$$

is an univariate polynomial with real coefficients.

We associate to $f$ the corresponding bivariate polynomial

$$F(X,Y) \,=\, \sum_{j=1}^{n} a_j \, X^j \, Y^{n-j}$$

and consider its Hessian

$$\mathrm{H}\,(F) \,=\, \det \begin{pmatrix} F_{XX} & F_{XY} \\ \\ F_{XY} & F_{YY} \end{pmatrix} .$$

We have

$$F_{XX} \,=\, \sum_{j=0}^{n} j(j-1) a_j \, X^{j-2} Y^{n-j},$$

$$F_{XY} \,=\, \sum_{j=0}^{n} j(n-j) a_j \, X^{j-1} Y^{n-j-1},$$

$$F_{YY} \,=\, \sum_{j=0}^{n} (n-j)(n-j-1) a_j \, X^{j} Y^{n-j-2}.$$

We denote

$$f''_{xx} = F_{XX}(x,1),$$

$$f''_{xy} = F_{XY}(x,1),$$

$$f''_{yy} = F_{YY}(x,1),$$

and compute

$$\mathrm{H}\,(F)(x,1) \,:=\, \det \begin{pmatrix} f_{xx} & f_{xy} \\ \\ f_{xy} & f_{yy} \end{pmatrix} .$$

This determinant can be expressed as a function of $f$, $f'$ and $f''$. In fact, we have

$$f''_{xx} = \sum_{j=0}^{n} j(j-1)a_j\, x^{j-2} = f'',$$

$$f''_{xy} = \sum_{j=0}^{n} j(n-j)a_j\, x^{j-1}$$

$$= n\sum_{j=0}^{n} j a_j\, x^{j-1} - \sum_{j=0}^{n} j^2\, a_j\, x^{j-1}$$

$$= nf' - x\sum_{j=0}^{n} j(j-1)a_j\, x^{j-2} - \sum_{j=0}^{n} j a_j\, x^{j-1}$$

$$= (n-1)f' - xf'',$$

$$f''_{yy} = \sum_{j=0}^{n}(n-j)(n-j-1)a_j\, x^{j}$$

$$= (n^2-n)\sum_{j=0}^{n} a_j\, x^j - 2n\sum_{j=0}^{n} j\, a_j\, x^j + \sum_{0}^{n} j(j+1)a_j\, x^j$$

$$= (n^2-n)\sum_{j=0}^{n} a_j\, x^j - 2(n-1)\sum_{j=0}^{n} j\, a_j\, x^j + \sum_{0}^{n} j(j-1)a_j\, x^j$$

$$= (n^2-n)\sum_{j=0}^{n} a_j\, x^j - 2(n-1)x\sum_{j=0}^{n} j\, a_j\, x^{j-1} + x^2\sum_{0}^{n} j(j-1)a_j\, x^{j-2}$$

$$= (n^2-n)f - 2(n-1)xf' + x^2 f''.$$

It follows that

$$\mathrm{H}(F)(x,1) = \begin{vmatrix} f'' & (n-1)f' - xf'' \\ (n-1)f' - xf'' & (n^2-n)f'^2 - 2(n-1)xf' + x^2 f'' \end{vmatrix}$$

$$= n(n-1)ff'' - (n-1)^2 f'^2\,.$$

Note that this differs only by a sign from the *Hessian* considered by Laguerre [7]:

$$\mathrm{H}(f) \,=\, (n-1)^2\, f'^2 - n(n-1)\, f f' \,\geq\, 0\,.$$

We remind the reader of the following:

**Theorem 8 (Laguerre)** *If the polynomial $f$ has real simple roots, then the Hessian $H(f)$ is positive.*

## Applications to Roots of Orthogonal Polynomials

Let $f \in \mathbb{R}[X]$ be a polynomial of degree $n \geq 2$ that satisfies the second–order differential equation

$$p(x)\, y'' + q(x)\, y' + r(x)\, y \,=\, 0\,, \tag{4}$$

with $p$, $q$ and $r$ univariate polynomials with real coefficients, $p(x) \neq 0$.

Let us assume that all the roots of $f$ are simple and real and let $\alpha$ be a root of $f$. Laguerre established

$$4(n-1)\Big(p(\alpha)r(\alpha) + p(\alpha)q'(\alpha) - p'(\alpha)q(\alpha)\Big) - (n+2)q(\alpha)^2 \,\geq\, 0\,. \tag{5}$$

The inequality (4) can be applied succesfully for finding upper bounds for the roots of orthogonal polynomials.

**Example.** Consider the Legendre polynomial $P_n$, which satisfies the differential equation

$$(1-x^2)y'' - 2xy' + n(n+1)y \,=\, 0\,.$$

From (4) it follows that $La(n) = (n-1)\sqrt{\dfrac{n+2}{n(n^2+2)}}$ is a bound for the roots of $P_n$. We have:

Bounds for Zeros of Legendre Polynomials

| $n$ | La(P) | LPR |
|---|---|---|
| 5 | 0.91084 | 0.90617 |
| 8 | 0.96334 | 0.96028 |
| 11 | 0.98021 | 0.97822 |
| 55 | 0.99917 | 0.99906 |
| 100 | 0.99975 | 0.99971 |

**Remark**: The Hessian can give accurate bounds also for other orthogonal polynomials if convenient differential equations are examined.

**Acknowledgment.** This research was supported by the Romanian grant CNCSIS 1068/2006.

# References

[1] A. G. Akritas, A. W. Strzeboński: A comparative study of two real root isolation methods, Nonlin. Anal: Modell. Control, **10**, 297–304 (2005).

[2] A. Akritas, A. Strzeboński, P. Vigklas: Implementations of a new theorem for Computing bounds for positive roots of polynomials, Computing, **78**, 355-367 (2006).

[3] A.–L. Cauchy: Exercises de mathématiques, Paris (1829).

[4] I. Z. Emiris, E. P. Tsigaridas: Univariate polynomial real root isolation: Continued fractions revisited, http://arxiv.org/pdf/cs.SC/0604066 (2006).

[5] J. B. Kioustelidis: Bounds for positive roots of polynomials, J. Comput. Appl. Math., **16**, 241–244 (1986).

[6] J.–L. Lagrange: Traité de la résolution des équations numériques, Paris (1798).

(Reprinted in Œuvres, t. VIII, Gauthier–Villars, Paris (1879).)

[7] E. Laguerre: Mémoire pour obtenir par approximation les racines d'une équation algébrique qui a toutes les racines réelles, Nouv. Ann. Math., 2ème série, **19**, 161-172, 193-202 (1880).

[8] M. Mignotte, D. Ştefănescu: Polynomials – An algorithmic approach, Springer Verlag (1999).

[9] F. Rouillier, P. Zimmermann: Efficient isolation of polynomial's real roots, J. Comput. Appl. Math., **162**, 33–50 (2004).

[10] D. Ştefănescu: New bounds for the positive roots of polynomials, J. Univ. Comp. Sc., **11**, 2125–2131 (2005).

[11] D. Ştefănescu: Inequalities on Upper Bounds for Real Polynomial Roots, in Computer Algebra in Scientific Computing, 284–294, LNCS 4194 (2006).

[12] C. K. Yap: Fundamental problems of algorithmic algebra, Oxford University Press (2000).

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# Analytic Computation of Normally Hyperbolic Invariant Manifolds and their Surrounding Structures

Jesús F. Palacián, Patricia Yanguas

*Departamento de Ingeniería Matemática e Informática,*
*Universidad Pública de Navarra, Campus de Arrosadía,*
*31006 Pamplona (Navarra) Spain*

**Abstract.** A normal form approach is used to get an analytical handle on geometric objects, such as the normally hyperbolic invariant manifolds (NHIMs), and to achieve as a consequence a breakthrough in astrodynamics and reaction dynamics. The technique is illustrated by applying it to the three collinear equilibria ($L_1$, $L_2$ and $L_3$) of the spatial circular restricted three-body problem. The expressions involved in the procedure need to be known very accurately; thus, one has to carry out the computations to a high degree, necessitating the handling of huge expressions through symbolic computations. The advantages of this approach compared to other techniques are emphasized. Finally we give some hints on its application to time-dependent Hamiltonians.

## 1 Introduction

Transition State (TS) has been a subject of continuous interest in chemistry since the early nineties. Classical Transition State Theory (TST) is the cornerstone of reaction rate theory. It postulates a partition of phase space into reactant and product regions, which are separated by a dividing surface that reactive trajectories must cross. Historically, it has been analyzed from various points of view and a great development in the understanding of TS has been made. For a good introduction on the subject, see [2] and references therein. Several techniques for studying two-degrees-of-freedom problems have been proposed and successfully used. The generalization to more than two degrees of freedom is a non-trivial result. It was made possible through a phase space formulation of TS and the computation of geometric structures such as NHIMs, as well as the computation of other related structures [7].

Transition phenomena do not only occur in chemical reactions [2], but also in other kinds of problems, such as those appearing in astrodynamics. For

instance, we can mention capture and escape phenomena in the solar system [4], the design of space missions using low-energy orbits, and the determination of homoclinic and heteroclinic trajectories. All share a common mathematical framework. Thus, TS techniques applied in chemistry can also be applied to astrodynamic problems and vice-versa: celestial mechanics tools can be used to deal with chemical reaction dynamics. As a result, both fields are advanced.

Examples of mathematical models used to study problems in astrodynamics are the spatial (or planar) elliptic (or circular) restricted three-body problem, the quasi-bicircular or tricircular problems, and the restricted four-body problem with Hill approximation [6], among others.

Here we will choose the spatial circular restricted three-body problem as an example to calculate the geometric structures defining the TS around the collinear equilibria ($L_1$, $L_2$ and $L_3$). The calculation of normal forms is a key to computing the TS and their associated structures with great precision. In case of no resonances among the fundamental frequencies of the linearised system around the collinear points, we compute the normal form as a Hamiltonian with zero degrees of freedom. We show how to determine analytically the TS, as well as the NHIM, together with its stable and unstable invariant manifolds. We compute trajectories that start on the NHIM in the energy surface. And, we determine trajectories in either the forward or backward stable and unstable manifolds associated with the NHIM. These trajectories are simply chosen and computed from the normal form vector field. Then, the normal form transformation allows us to visualize these structures in the original coordinates if the number of degrees of freedom is two or three. Thus, we have complete control and knowledge of the exact dynamical trajectories near the TS in an $n$-DOF system.

Our approach is algorithmic in nature in the sense that we provide a series of steps that can be carried out to locate the NHIM, its stable and unstable manifolds, and the TS, as well as describing all possible trajectories near the TS [7].

The advantages of our approach compared to the technique called *reduction to the central manifold* are emphasized. For instance, we can determine halo and Lissajous quasiperiodic orbits only from the normal form without calculating any Lindstedt series. We also mention how to detect homoclinic and heteroclinic orbits without resorting to Montecarlo simulations. The theoretical background of our exposition appears in [7, 8, 9, 2].

We start by recalling the definition of the geometric structures we will compute [7]. In Section 3 we make a brief statement on the technique to compute the geometric structures defined in Section 2: normal forms. Section 4 treats the spatial circular restricted three-body problem. Finally, in Section 5 we consider the extension of the method to time-dependent systems.

## 2 Definition of the geometric structures

We start by stating the definition of the structures we want to compute. For more information, see [7, 2]. Let us consider a Hamiltonian system of the type:

$$\mathcal{H} = \sum_{i=1}^{n-1} \frac{\omega_i}{2}\left(p_i^2 + q_i^2\right) + \lambda q_n p_n + f_1(q_1, \ldots, q_{n-1}, p_1, \ldots, p_{n-1}, \mathcal{I}) + f_2(q_1, \ldots, q_{n-1}, p_1, \ldots, p_{n-1}), \quad (1)$$

where $(q_1, \ldots, q_n, p_1, \ldots, p_n) \in \mathbf{R}^{2n}$ denote canonical coordinates, $\mathcal{I} \equiv p_n q_n$ and $f_1$, $f_2$ are at least of third order, i.e., they are responsible for the non-linear terms in the Hamiltonian vector field, and $f_1(q_1, \ldots, q_{n-1}, p_1, \ldots, p_{n-1}, 0) = 0$. Hamiltonian $\mathcal{H}$ has an equilibrium at the origin.

The dynamics associated with $\mathcal{H}$ occurs in the $(2n-1)$-dimensional energy surface given by:

$$\sum_{i=1}^{n-1} \frac{\omega_i}{2}\left(p_i^2 + q_i^2\right) + \lambda q_n p_n + f_1(q_1, \ldots, q_{n-1}, p_1, \ldots, p_{n-1}, \mathcal{I}) + f_2(q_1, \ldots, q_{n-1}, p_1, \ldots, p_{n-1}) = h = constant > 0. \quad (2)$$

Taking $q_n = p_n = 0$ in (1) we get a $(2n-2)$-dimensional invariant manifold. Its intersection with the $(2n-1)$-dimensional energy surface (2) is given by

$$\mathcal{M}_h^{2\,n-3} = \Big\{(q_1, \ldots, q_n, p_1, \ldots, p_n)\,|\,q_n = p_n = 0, \sum_{i=1}^{n-1} \frac{\omega_i}{2}\left(p_i^2 + q_i^2\right) + f_2(q_1, \ldots, q_{n-1}, p_1, \ldots, p_{n-1}) = h = \text{ constant} > 0\Big\}. \quad (3)$$

This is the NHIM, a "multidimensional version of a saddle" ("saddle sphere") whose dimension is $2\,n-3$ and which possesses a codimension of one stable and one unstable manifold. It is approximated by using high-order normal forms, as we will see in next section. A NHIM roughly means that the stretching and contraction rates under linearised dynamics transverse to the $(2n-3)$-sphere dominate those tangent to the $(2n-3)$-sphere. The dynamics normal to the $(2n-3)$-sphere is described by the exponential contraction and expansion of the saddle point dynamics.

A key advantage of the normal form is that the stable and unstable manifolds of $\mathcal{M}_h^{2n-3}$ can be known explicitly. They are given by the following expressions:

$$W^s\left(\mathcal{M}_h^{2n-3}\right) = \Big\{(q_1, \ldots, q_n, p_1, \ldots, p_n)\,|\,\sum_{i=1}^{n-1} \frac{\omega_i}{2}\left(p_i^2 + q_i^2\right) + f_2(q_1, \ldots, q_{n-1}, p_1, \ldots, p_{n-1}) = h = constant > 0,\, q_n = 0\Big\},$$

$$W^u\left(\mathcal{M}_h^{2n-3}\right) = \left\{(q_1,\ldots,q_n,p_1,\ldots,p_n)\,|\,\sum_{i=1}^{n-1}\frac{\omega_i}{2}\left(p_i^2+q_i^2\right)\right.$$
$$\left.+f_2(q_1,\ldots,q_{n-1},p_1,\ldots,p_{n-1})=h=constant>0,\ p_n=0\right\}. \quad (4)$$

Hence, the stable and unstable manifolds of the sphere have the structure of $\mathcal{M}_h^{2n-3}\times\mathbf{R}$. They are $(2\,n-2)$-dimensional objects which act as multidimensional separatrices. The central manifold (with dimension $2\,n-2$) associated with the equilibrium (the origin) is given by $p_n=q_n=0$. The stable manifold of dimension 1 is given by $q_i=p_i=0$ for $i=1,\ldots,n-1$, $q_n=0$. The unstable manifold of dimension 1 is given by $q_i=p_i=0$ for $i=1,\ldots,n-1$, $p_n=0$. Precisely the intersection of the central manifold of the origin with the energy surface (2) is the NHIM. Taking into account that the dynamics takes place on the energy surface, the NHIM, rather than the central manifold, is the right structure to compute in order to study the dynamics.

The TS for (1) is obtained by taking $q_n=p_n$:

$$\mathcal{TS}_h^{2\,n-2} = \Big\{(q_1,\ldots,q_n,p_1,\ldots,p_n)\,|\,q_n=p_n,$$
$$\sum_{i=1}^{n-1}\frac{\omega_i}{2}\left(p_i^2+q_i^2\right)+f_1(q_1,\ldots,q_{n-1},p_1,\ldots,p_{n-1},p_n^2)$$
$$+\,f_2(q_1,\ldots,q_{n-1},p_1,\ldots,p_{n-1})=\,h=\,constant>\,0\Big\}.$$

This structure has all the properties that define the so-called TS: it is locally a "surface of no return" where the trajectories crossing $\mathcal{TS}_h^{2\,n-2}$ correspond to "reactive" trajectories and all reactive particles must pass through it.

# 3 Normal forms

All the structures defined in the previous section can be obtained analytically by means of normal forms [1]. We outline here the main ideas of the procedure.

Consider a Hamiltonian vector field $\mathcal{H}$ such that

$$\mathcal{H}(\mathbf{q},\mathbf{p};\varepsilon)=\sum_{i\geq 0}\frac{\varepsilon^i}{i!}\,\mathcal{H}_i(\mathbf{q},\mathbf{p}),\qquad |\mathcal{H}_0|\gg\varepsilon\,|\mathcal{H}_1|\gg\varepsilon^2\,|\mathcal{H}_2|/2\gg\ldots,$$

where $(\mathbf{q},\mathbf{p})=(q_1,\ldots,q_n,p_1,\ldots,p_n)$, $\varepsilon$ is a small parameter and:

$$\frac{d\,\mathbf{q}}{d\,t}=\frac{\partial\mathcal{H}}{\partial\mathbf{p}},\qquad \frac{d\,\mathbf{p}}{d\,t}=-\frac{\partial\mathcal{H}}{\partial\mathbf{q}}.$$

Denoting $\mathbf{x}=(\mathbf{q},\mathbf{p})$, calculating a normal form associated with $\mathcal{H}$ means constructing an application $\mathbf{X}:(\mathbf{y};\varepsilon)\to\mathbf{x}$, as a solution of the initial value problem

$$\frac{d\,\mathbf{x}}{d\,\varepsilon}=\frac{\partial\,\mathcal{W}}{\partial\,\mathbf{x}}(\mathbf{x};\varepsilon),\quad \mathbf{x}(\mathbf{0})=\mathbf{y}$$

through a generating function

$$\mathcal{W}(\mathbf{x};\varepsilon) = \sum_{i\geq 0} \frac{\varepsilon^i}{i!}\,\mathcal{W}_{i+1}(\mathbf{x})$$

such that the vector field $\mathcal{H}$ is transformed into

$$\mathcal{K}(\mathbf{x};\varepsilon) = \sum_{i\geq 0} \frac{\varepsilon^i}{i!}\,\mathcal{K}_i(\mathbf{y}).$$

The procedure to construct $\mathcal{K}$ is recurrent and, at each step, one has to solve the homology equation, which is the partial differential equation:

$$\{\,\mathcal{H}_0^{(0)}\,,\,\mathcal{W}_i\,\} + \mathcal{K}_i \;=\; \tilde{\mathcal{H}}_0^{(i)}$$

where terms $\tilde{\mathcal{H}}_0^{(i)}$ are the ones depending on the previous orders and are already known and $\mathcal{K}_i$ is chosen to be simple" according to the goal of the transformation. The application $\mathbf{x} = \mathbf{X}(\mathbf{y};\varepsilon)$ relates the old coordinates $\mathbf{x}$ with the new ones $\mathbf{y}$ and is a near-identity transformation. This is a direct change of variables, given by

$$\mathbf{x} = \mathbf{X}(\mathbf{y};\varepsilon) \;=\; \mathbf{y} + \sum_{i\geq 1} \frac{\varepsilon^i}{i!}\,\mathbf{y}_0^{(i)}$$

with $\mathbf{y}_0^{(0)} \equiv \mathbf{y}$, $\mathbf{y}_i^{(0)} \equiv 0$ if $i \geq 1$. The inverse change of variables is also given explicitly through the expression:

$$\mathbf{y} = \mathbf{Y}(\mathbf{x};\varepsilon) \;=\; \mathbf{x} + \sum_{i\geq 1} \frac{\varepsilon^i}{i!}\,\mathbf{x}_0^{(i)} \tag{5}$$

with $\mathbf{x}_0^{(0)} \equiv \mathbf{x}$, $\mathbf{x}_i^{(0)} \equiv 0$ if $i \geq 1$. The way of computing the terms $\mathbf{y}_0^{(i)}$ and $\mathbf{x}_0^{(i)}$ is recurrent and uses the generators given by $\mathcal{W}_i$.

The advantage of the normal form is that once it is obtained, one can compute analytically the NHIM, its stable and unstable manifolds, and the TS in the original coordinates in terms of so-called "normal form" coordinates. To accomplish that, we make use of direct and inverse changes of the variables given above.

The NHIM is obtained by setting $q_n = p_n = 0$ in the normal form; its stable manifold is got by setting $q_n = 0$; the unstable manifold is obtained by setting $p_n = 0$; the TS results after making $q_n = p_n$.

# 4 The circular restricted three-body problem

In this section we focus on the circular restricted three-body problem and calculate the TS, the NHIM, and its stable and unstable manifolds. An extended study of this system will be made in a subsequent paper.

The Hamiltonian corresponding to the circular restricted three body problem centered on the center of mass of the primaries is:

$$\begin{aligned}\mathcal{H} &= \frac{1}{2}\left(P_1^2+P_2^2+P_3^2\right)-(x_1\,P_2-x_2\,P_1)-\frac{1-\mu}{\sqrt{(\mu+x_1)^2+x_2^2+x_3^2}}\\ &\quad-\frac{\mu}{\sqrt{(1-\mu-x_1)^2+x_2^2+x_3^2}},\end{aligned}$$

where $(x_1, x_2, x_3, P_1, P_2, P_3)$ represent Cartesian coordinates and momenta and $\mu = m_1/(m_1+m_2) \in (0, 1/2]$ is the reduced mass.

We know from classical analysis that this system has five equilibria: the so-called Euler equilibria $L_1, L_2, L_3$, which are unstable, and the Lagrange equilibria, $L_4$ and $L_5$, which are linearly stable up to a certain value of $\mu$ (the Routh limit), see for instance [5]. The three Lagrange equilibria are of interest to us because after expanding (4) around each of them and performing appropriate changes of variables, one gets a Hamiltonian of the type (1).

First, we translate the equilibrium (either $L_1$, $L_2$ or $L_3$) to the origin and then Taylor-expand the Hamiltonian (4) around the origin up to order $n$. We get $\mathcal{H} = \sum_{i=0}^{n} \mathcal{H}_i$, where each $\mathcal{H}_i$ denotes a homogeneous polynomial of degree $i+2$. The eigenvalues of the matrix associated with $\mathcal{H}_0$ are:

$$\begin{aligned}\gamma_{1,2} &= \pm\sqrt{\alpha}\,\imath,\\ \gamma_{3,4} &= \pm\sqrt{\frac{\alpha-2-\sqrt{\alpha\,(9\,\alpha-8)}}{2}},\quad \gamma_{5,6} = \pm\sqrt{\frac{\alpha-2+\sqrt{\alpha\,(9\,\alpha-8)}}{2}},\end{aligned}$$

where $\alpha = \mu/d_2^3 + (1-\mu)/d_1^3$ and $d_i$ is the distance from the equilibrium to the body $i$. As $\alpha > 1$, $\gamma_{5,6}$ are real and $\gamma_{1,2}$ and $\gamma_{3,4}$ are purely imaginary. Using the eigenvectors associated to the linear part of the system we perform a linear canonical change of coordinates and arrive at a Hamiltonian whose quadratic part is

$$\mathcal{H}_0 = \frac{\omega_1}{2}\,(q_1^2+p_1^2) + \frac{\omega_2}{2}\,(q_2^2+p_2^2) + \mu\,q_3\,p_3,$$

or

$$\mathcal{H}_0 = \imath\,\omega_1\,\bar{q}_1\,\bar{p}_1 + \imath\,\omega_2\,\bar{q}_2\,\bar{p}_2 + \mu\,\bar{q}_3\,\bar{p}_3,$$

where $\omega_1 = \gamma_1$, $\omega_2 = \gamma_3$ and $\mu = \gamma_5$. Once having the linear part in an adequate form we apply the change to the non-linear terms.

The following step transforms $\mathcal{H}$ into $\mathcal{K}$ through a normal form calculation in such a way that $\mathcal{K}$ admits three independent formal integrals in involution. For that, we identify $\mathcal{K}_0 = \mathcal{H}_0$ and, for each $i \in [1, n]$, choose $\mathcal{K}_i$ so that the Poisson bracket $\{\,\mathcal{K}_i\,,\,\mathcal{K}_0\,\} = 0$. Picking a monomial $m_i$ of degree $i+2$:

$$m_i = c\,\bar{q}_1^{j_1}\,\bar{q}_2^{j_2}\,\bar{q}_3^{j_3}\,\bar{p}_1^{k_1}\,\bar{p}_2^{k_2}\,\bar{p}_3^{k_3} \quad\text{with}\quad \sum_{l=1}^{3}(j_l+k_l) = i+2,$$

notice that $\{m_i; \mathcal{H}_0\} = 0$ if and only if $j_1 = k_1$, $j_2 = k_2$ and $j_3 = k_3$. In this case $m_i$ is part of $\mathcal{K}_i$; otherwise, the monomial $m_i/(\mu\,(k_3-j_3) + \imath\,\omega_1\,(k_1-j_1) +$

$\imath\,\omega_2\,(k_2 - j_2))$ goes to $\mathcal{W}_i$. This, then, is a step-by-step method for constructing the normal form $\mathcal{K}$ and the generating function $\mathcal{W}$.

In order to check the validity of the transformation, we estimate its global error. Here we only present the calculations for $L_1$ but the results are similar for the remaining equilibria. We fix a neighborhood of the origin, then we take a sample of points such that

$$\|(x_1, x_2, x_3, P_1, P_2, P_3)\| \leq 10^{-2}$$

and compose the transformed Hamiltonian $\mathcal{K}$ with the inverse change (5):

$$|\mathcal{K}(\mathbf{Y}(\mathbf{x})) - \mathcal{H}(\mathbf{x})| = \mathbf{E}(\mathbf{x}) \,=\, \mathbf{E_{n+1}}(\mathbf{x}) + \mathcal{O}(\|\mathbf{x}\|^{\mathbf{n+2}}).$$

For Jupiter and the Sun, the relative error $E_{n+1}(\mathbf{x})/|\mathcal{H}(\mathbf{x})|$ yields

| | |
|---|---|
| degree 4 | 0.0000292729... |
| degree 6 | $1.2966485295... \times 10^{-7}$ |
| degree 8 | $3.5986352013... \times 10^{-9}$ |
| degree 10 | $1.3577798902... \times 10^{-10}$ |
| degree 12 | $5.8177310016... \times 10^{-12}$ |
| degree 14 | $1.2314332549... \times 10^{-14}$ |

Once the normal form is obtained, we calculate the geometric structures defined in Section 2. The dynamics occurs in a 5-dimensional (5D) energy surface. Setting $q_3 = p_3 = 0$ in the vector field associated with $\mathcal{H}$, then $\dot{q}_3 = \dot{p}_3 = 0$ and we get the central manifold. Its intersection with the 5D energy surface is the NHIM. Given an energy value $h > 0$, the NHIM of $L_1$ associated with the Hamiltonian in the normal form coordinates has the form:

$$\mathcal{M}_h^3(q_1, q_2, p_1, p_2) \,= \\ \left\{(q_1, q_2, p_1, p_2) \mid \frac{\omega_1}{2}\,(q_1^2 + p_1^2) + \frac{\omega_2}{2}\,(q_2^2 + p_2^2) + f_2(q_1\,p_1, q_2\,p_2) = h\right\},$$

where $f_2$ represents polynomials of degree at least three. This is topologically a deformed sphere, the generalization of the Lyapunov orbit of the planar problem.

The NHIM has 4D stable and unstable manifolds in the 5D energy surface since normal hyperbolicity is preserved under perturbations. In the normal form coordinates they are:

$$W^s\left(\mathcal{M}_h^3\right) = \{(q_1, q_2, q_3, p_1, p_2, p_3) \mid q_3 = 0,$$
$$\frac{\omega_1}{2}\left(q_1^2 + p_1^2\right) + \frac{\omega_2}{2}\left(q_2^2 + p_2^2\right) + f_2(q_1\, p_1, q_2\, p_2) = h\Big\},$$

$$W^u\left(\mathcal{M}_h^3\right) = \{(q_1, q_2, q_3, p_1, p_2, p_3) \mid p_3 = 0,$$
$$\frac{\omega_1}{2}\left(q_1^2 + p_1^2\right) + \frac{\omega_2}{2}\left(q_2^2 + p_2^2\right) + f_2(q_1\, p_1, q_2\, p_2) = h\Big\},$$

$$\mathcal{TS} = \{(q_1, q_2, q_3, p_1, p_2, p_3) \mid p_3 = 0,$$
$$\frac{\omega_1}{2}\left(q_1^2 + p_1^2\right) + \frac{\omega_2}{2}\left(q_2^2 + p_2^2\right) + \mu\, p_3^2 + f_1(q_1\, p_1, q_2\, p_2, p_3^2)$$
$$+ f_2(q_1\, p_1, q_2\, p_2) = h\}\,.$$

The expressions of the manifolds of $L_1$ in the original variables are:

- $\mathbf{x}_{\text{NHIM}} = \psi\, \mathbf{X}(q_1, q_2, 0, p_1, p_2, 0)$ and we add $\mathcal{K}(q_1, q_2, 0, p_1, p_2, 0) = h$ (3D NHIM),
- $\mathbf{x}_{S\text{NHIM}} = \psi\, \mathbf{X}(q_1, q_2, 0, p_1, p_2, p_3)$ with $\mathcal{K}(q_1, q_2, 0, p_1, p_2, p_3) = h$ (4D stable manifold of the NHIM),
- $\mathbf{x}_{U\text{NHIM}} = \psi\, \mathbf{X}(q_1, q_2, q_3, p_1, p_2, 0)$ and $\mathcal{K}(q_1, q_2, q_3, p_1, p_2, 0) = h$ (4D unstable manifold of the NHIM).

The existence of true invariant sets of $\mathcal{H}$ close to the ones we have computed can be guaranteed, provided the global error of the process is kept bounded. The manifolds $\mathbf{x}_{S\text{NHIM}}$ and $\mathbf{x}_{U\text{NHIM}}$ can be globalized numerically. The first integrals for the original Hamiltonian can be calculated from the integrals of motion for $\mathcal{K}$, which are: $\bar{q}_1\, \bar{p}_1$, $\bar{q}_2\, \bar{p}_2$ and $\bar{q}_3\, \bar{p}_3$. Their inverses via the Lie transformation and linear changes, i.e. the polynomials in $(x_1, x_2, x_3, P_1, P_2, P_3)$ which are approximate integrals of motion for $\mathcal{H}$, are:

$$J_1 = \psi\, \mathbf{Y}(\bar{q}_1\, \bar{p}_1), \quad J_2 = \psi\, \mathbf{Y}(\bar{q}_2\, \bar{p}_2), \quad J_3 = \psi\, \mathbf{Y}(\bar{q}_3\, \bar{p}_3).$$

Checking that the Poisson brackets $\{\, \mathcal{H}\,,\, J_i\,\} \approx 0$, we conclude that the integrals are approximated up to 14 digits within balls of radii $10^{-2}$ and $n = 15$.

It is also possible to calculate halo and Lissajous orbits, invariant 2-tori and other quasi-periodic trajectories of $\mathcal{H}$. For that we make use of action and angle coordinates:

$$q_1 = \sqrt{2\, I_1}\, \cos\, \theta_1,\, p_1 = \sqrt{2\, I_1}\, \sin\, \theta_1, q_2 = \sqrt{2\, I_2}\, \cos\, \theta_2,\, p_2 = \sqrt{2\, I_2}\, \sin\, \theta_2.$$

The expression

$$\mathcal{T}_h = \psi\, \mathbf{X}\left(\sqrt{2\, I_1}\, \cos\, \theta_1, \sqrt{2\, I_2}\, \cos\, \theta_2, 0, \sqrt{2\, I_1}\, \sin\, \theta_1, \sqrt{2\, I_2}\, \sin\, \theta_2, 0\right)$$

with $\theta_1, \theta_2 \in [0, 2\,\pi)$ and $0 \leq I_2 \leq I_{2MAX}$ and $I_1$ obtained from $\mathcal{K}(I_1, I_2, 0) = h$ represents a one-parameter family of 2-tori. For an example, see Figure 1. By doing in $\mathcal{T}_h$, $\theta_1 = \nu_1\, t + \phi_1$, $\theta_2 = \nu_2\, t + \phi_2$ with $\nu_1, \nu_2, \phi_1$ and $\phi_2 \in \mathbf{R}$ we get Lissajous trajectories. By doing in $\mathcal{T}_h$, $\theta_1 \equiv \theta_2 = \nu\, t + \phi$, we get halo orbits.

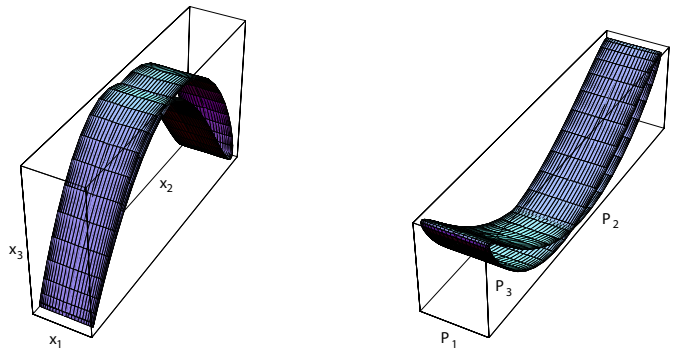



Figure 1: Projection of the NHIM foliated by invariant 2-tori

# 5 Generalizations and Further Considerations

In this section we point out some generalizations of TST to time-dependent Hamiltonians. We see through an example how to handle, for instance, non-autonomous Hamiltonians appearing in some extensions of the circular restricted three-body problem. Problems of molecular dynamics where the reactions are driven by laser pulses [3] are modeled through non-autonomous Hamiltonians. Let us consider the following example borrowed from astrodynamics [6]:

$$\mathcal{H} = \frac{1}{2}\left(p_x^2 + p_y^2 + p_z^2\right) - (m+1)\left(x\,p_y - y\,p_x\right) + \frac{1}{2}(m+1)^2\left(x^2+y^2\right) + \mathcal{V}(x,y,z),$$

where the potential is given through

$$\begin{aligned}\mathcal{V}(x,y,z) &= \frac{1}{2}\left(1+2m+\frac{3}{2}m^2\right)(x^2+y^2) - \frac{1}{2}m^2\,z^2 \\ &\quad + \frac{3}{4}m^2\Big[(x^2-y^2)\,\cos\,2\tau - 2x\,y\,\sin\,2\tau\Big] + \frac{m^2}{a_0^3}\left(\frac{1-\nu}{R_{1-\nu}} + \frac{\nu}{R_\nu}\right),\end{aligned}$$

and

$$\begin{aligned}R_{1-\nu} &= \sqrt{(x+\nu\,\mathbf{R}_x)^2 + (y+\nu\,\mathbf{R}_y)^2 + z^2}, \\ R_\nu &= \sqrt{[x-(1-\nu)\,\mathbf{R}_x]^2 + [y-(1-\nu)\,\mathbf{R}_y]^2 + z^2}, \\ a_0 &= m^{2/3}\left(1 - \frac{2}{3}\,m + \frac{7}{18}\,m^2 + \cdots\right).\end{aligned}$$

The model is the so-called restricted four-body problem with Hill approximation. It enlarges the restricted three-body problem (recovered after setting $m=0$), as well as the Hill problem (recovered after setting $\nu=0$). When both parameters $\nu$ and $m$ do not vanish, $\mathcal{H}$ represents a time-dependent Hamiltonian system with three degrees of freedom. First we expand it around a small periodic orbit surrounding the collinear points. Then we eliminate the time variable by applying canonical Floquet theory. The quadratic part of the system is

associated with the linear ordinary differential equation (ODE)

$$\mathbf{x}(t) = A(t; \mathbf{c})\, \mathbf{x}(t).$$

We need to solve the equation above accurately in order to get the monodromy matrix. The numerical resolution can lead to poor results since the ODE is stiff, in general. An alternative to Floquet theory in the periodic case could be an approximation of the fundamental matrix of the ODE using the Picard iterations:

$$X_0 = I_m, \qquad X_{n+1}(t) = X_0 + \int A(t)\, X_n(t)\, dt.$$

In this case the monodromy matrix is obtained through the expression $M = X_p(T)$. Then one follows the same steps as in canonical Floquet theory. The canonical lineal change of coordinates is calculated either following a recurrent scheme or using Fourier series adjusting their coefficients by least squares. The system is converted into an autonomous one of three degrees of freedom and similar techniques to those explained here are applied to get the invariant structures.

The theory described throughout the paper and applied in Section 4 works also for resonant systems (i.e. systems where resonances among the central directions appear), although in these situations the normal form Hamiltonian has more than zero degrees of freedom.

As we have seen in Section 4, we determine the invariant 2-tori, Lissajous and halo orbits without using Lindstedt-Poincaré techniques. We can look for heteroclinic trajectories without using Monte Carlo simulations, but using the 2-tori. Except for estimating of the error, the approach is purely analytical.

The calculations have been made with *Mathematica.*

**Acknowledgments.** This research was partially supported by Project # MTM2005-08595 of Ministerio de Educación y Ciencia (Spain).

# References

[1] Arnol'd, V.I., Kozlov, V.V. and Neishtadt, A.I. - *Mathematical Aspects of Classical and Celestial Mechanics.* Springer Verlag, New York, 1988.

[2] Jaffé, C., Kawai, S., Palacián, J.F., Yanguas, P. and Uzer, T. - A new look at the transition state: Wigners dynamical perspective revisited. *Adv. Chem. Phys.* Vol. 130 A, 2005, 171-216.

[3] Kawai, S., Bandrauk, A.D., Jaffé, C., Bartsch, T., Palacián, J.F. and Uzer, T. - Transition state theory for laser-driven reactions. *J. Chem. Phys.* Vol. 126, 2007, 164306.

[4] Koon, W.S., Lo, M.W., Marsden, J.E. and Ross, S.D. - Heteroclinic connections between periodic orbits and resonance transitions in celestial mechanics. *Chaos* Vol. 10(2), 2000, 427-469.

[5] Meyer, K.R. and Hall, G.R. - *Introduction to Hamiltonian dynamical systems and the N-body problem.* Springer Verlag, New York, 1992.

[6] Scheeres, D.J. and Bellerose, J. - The restricted Hill full 4-body problem: application to spacecraft motion about binary asteroids. *Dynamical Systems: An International Journal* Vol. 20(1), 2005, 23-44.

[7] Uzer, T., Jaffé, C., Palacián, J.F., Yanguas, P. and Wiggins, S. - The geometry of reaction dynamics. *Nonlinearity* Vol. 15, 2002, 957-992.

[8] Waalkens, H., Burbanks, A. and Wiggins, S. - A computational procedure to detect a new type of high-dimensional chaotic saddle and its application to the 3D Hill's problem. *J. Phys. A* Vol. 37, 2004, L257-L265.

[9] Waalkens, H., Burbanks, A. and Wiggins, S. - Phase space conduits for reaction in multi-dimensional systems: HCN isomerization in three dimensions. *J. Chem. Phys.* Vol. 121, 2004, 6207-6225.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# Algebraic Combinatoric Aspects of Asymptotic Analysis of Nonlinear Dynamical System with Singular Inputs

Hoang Ngoc Minh

*Université Lille II*
*1, Place Déliot, 59024 Lille, France*

## 1 Introduction

We are studying the convergence criterium (see Section 2.3) and the algebraic combinatoric aspects of the asymptotic behaviour of the coefficients $\{y_n\}_{n\geq 0}$ of the following formal serial expansion (see Section 2.4)

$$y(z) = \sum_{n\geq 0} y_n z^n, \tag{1}$$

of the output $y$ of the nonlinear dynamical system with singular inputs

$$\left\{ \begin{array}{lcl} y(z) & = & f(q(z)), \\ \dot{q}(z) & = & A_0(q)u_0(z) + A_1(q)u_1(z), \\ q(z_0) & = & q_0, \end{array} \right. \tag{2}$$

where, $u_0(z)$ and $u_1(z)$ are respectively the singular inputs $z^{-1}$ and $(1-z)^{-1}$, the state $q = (q_1, \ldots, q_n)$ belongs to the complex analytic manifold of dimension $n$ denoted by $Q$ and $q_0$ denotes the initial state, the observation $f$ belongs to the ring of holomorphic functions over $Q$ denoted by $\mathcal{O}$, and for $i = 0..1, A_i$ is the analytic vector field over $Q$ defined by

$$A_i(q) = \sum_{j=1}^{n} A_i^j(q) \frac{\partial}{\partial q_j}, \quad \text{with } A_i^j(q) \in \mathcal{O}, \quad \text{for } j = 1, \ldots, n. \tag{3}$$

Let us consider the smallest algebra containing $\mathcal{C} = \mathbb{C}[z, u_0(z), u_1(z)]$ and being stable by derivation and by integration with respect to the following differential forms $\omega_0(z) = u_0(z)\, dz$ and $\omega_1(z) = u_1(z)\, dz$. This is called *polylogarithm algebra* and it is denoted by $\mathrm{LI}_{\mathcal{C}}$ (see Section 2.1). Let us consider also the algebra generated by the special value at the singularity $z = 1$, when it exists, of the elements belonging $\mathrm{LI}_{\mathcal{C}}$. This is called *polyzêtas algebra* and it is denoted by $\mathcal{Z}$ (see Section 2.1).

In this work, to manipulate these algebras, let us introduce the set of words (resp. Lyndon words) on the finite encoding alphabet $X = \{x_0, x_1\}$, denoted by $X^*$ (resp. $\mathcal{L}ynX$). Let $\{S_l\}_{l\in\mathcal{L}ynX}$ denotes a transcendence basis of the *shuffle algebra* over $\mathbb{C}$ on $X$ and let $\{\check{S}_l\}_{l\in\mathcal{L}ynX}$ denote the Lyndon basis of the *free Lie algebra* over $\mathbb{C}$ on $X$. Each word $u = x_0^{s_1-1}x_1\dots x_0^{s_r-1}x_1$ in $X^*x_1$ can be associated to the multi-index $\mathbf{s} = (s_1, \dots, s_r)$ and it can be itself associated to the word $v = y_{s_1}\dots y_{s_r}$ in the the set of words $Y^*$ over the infinite encoding alphabet $Y = \{y_s\}_{s\geq 1}$. Therefore, the polylogarithms, polyzêtas and multiple harmonic sums can be indexed by the multi-index $\mathbf{s}$, as well as by the words $u$ or $v$:

$$\mathrm{Li}_{\mathbf{s}}(z) = \mathrm{Li}_u(z) = \mathrm{Li}_v(z) = \sum_{n_1>\dots>n_r>0} \frac{z^{n_1}}{n_1^{s_1}\dots n_r^{s_r}}, \tag{4}$$

$$\zeta(\mathbf{s}) = \zeta(u) = \zeta(v) = \sum_{n_1>\dots>n_r>0} \frac{1}{n_1^{s_1}\dots n_r^{s_r}}, \tag{5}$$

$$\mathrm{H}_{\mathbf{s}}(N) = \mathrm{H}_u(N) = \mathrm{H}_v(N) = \sum_{N>n_1>\dots>n_r>0} \frac{1}{n_1^{s_1}\dots n_r^{s_r}}. \tag{6}$$

By considering the noncommutative generating series on $Y$ of these objects (see Section 2.2), in particular

$$\Lambda(z) = \sum_{w\in Y^*} \mathrm{Li}_w(z)\, w \quad \text{and} \quad \mathrm{H}(N) = \sum_{w\in Y^*} \mathrm{H}_w(N)\, w, \tag{7}$$

we show the existence of a noncommutative generating series $S$ of convergent polyzêtas $\{\zeta(w)\}_{w\in Y^*\setminus y_1Y^*}$ such that

$$\lim_{z\to 1} \exp\left[y_1 \log\frac{1}{1-z}\right]\Lambda(z) = \lim_{N\to\infty} \exp\left[\sum_{k\geq 1} \mathrm{H}_{y_k}(N)\frac{(-y_1)^k}{k}\right]\mathrm{H}(N) = S. \tag{8}$$

This enables us to focus on the algebraic combinatoric aspects of the asymptotic behaviour of the coefficients of the formal serial expansion of the output of (2).

# 2 Asymptotic analysis of dynamical system

## 2.1 Multiple harmonic sum and polylogarithm

Let $\{t_i\}_{i\in\mathbb{N}_+}$ be an infinite set of variables. The elementary symmetric functions $\lambda_k$ and the sums of powers $\psi_k$ are defined by

$$\lambda_k(\underline{t}) = \sum_{n_1>\dots>n_k>0} t_{n_1}\dots t_{n_k} \quad \text{and} \quad \psi_k(\underline{t}) = \sum_{n>0} t_n^k. \tag{9}$$

They are respectively coefficients of the following generating functions

$$\lambda(\underline{t}|z) = \prod_{i\geq 1}(1+t_iz) \quad \text{and} \quad \psi(\underline{t}|z) = \sum_{i\geq 1}\frac{t_i}{1-t_iz}. \tag{10}$$

These generating functions satisfy a Newton identity

$$d/dz \log \lambda(\underline{t}|z) = \psi(\underline{t}|-z). \tag{11}$$

The fundamental theorem from symmetric function theory asserts that $\{\lambda_k\}_{k\geq 0}$ are linearly independent, and, putting $\lambda_0 = 1$, gives the remarkable identity:

$$k!\lambda_k = (-1)^k \sum_{\substack{s_1,\ldots,s_k>0 \\ s_1+\ldots+ks_k=k}} \binom{k}{s_1,\ldots,s_k} \left(-\frac{\psi_1}{1}\right)^{s_1} \ldots \left(-\frac{\psi_k}{k}\right)^{s_k} \tag{12}$$

Let $w = y_{s_1} \ldots y_{s_r} \in Y^*$. The quasi-symmetric function $F_w$, of depth $r = |w|$ and of degree (or weight) $s_1 + \ldots + s_r$, is defined by

$$F_w(\underline{t}) = \sum_{n_1>\ldots>n_r>0} t_{n_1}^{s_1} \ldots t_{n_r}^{s_r}. \tag{13}$$

In particular, $F_{y_1^k} = \lambda_k$ and $F_{y_k} = \psi_k$. As a consequence, the functions $\{F_{y_1^k}\}_{k\geq 0}$ are linearly independent and integrating the differential equation (11) shows that functions $F_{y_1^k}$ and $F_{y_k}$ are linked by the formula

$$\sum_{k\geq 0} F_{y_1^k} z^k = \exp\left[-\sum_{k\geq 1} F_{y_k} \frac{(-z)^k}{k}\right]. \tag{14}$$

The remarkable identity (12) can be then seen as

$$k!y_1^k = (-1)^k \sum_{\substack{s_1,\ldots,s_k>0 \\ s_1+\ldots+ks_k=k}} \binom{k}{s_1,\ldots,s_k} \frac{(-y_1)^{\sqcup\!\sqcup s_1}}{1^{s_1}} \sqcup\!\sqcup \ldots \sqcup\!\sqcup \frac{(-y_k)^{\sqcup\!\sqcup s_k}}{k^{s_k}} \tag{15}$$

Every $\mathrm{H}_w(N)$ can be obtained by specializing variables $\{t_i\}_{N\geq i\geq 1}$ at $t_i = 1/i$ and, for $i > N, t_i = 0$ in the quasi-symmetric function $F_w$ [11]. In the same way, when $w \in Y^* \setminus y_1 Y^*$, the convergent polyzêta $\zeta(w)$ can be obtained by specializing variables $\{t_i\}_{i\geq 1}$ at $t_i = 1/i$ in $F_w$ [11]. The notation $F_w$ is extended by linearity to all polynomials over $Y$.

If $u$ (resp. $v$) is a word in $Y^*$, of length $r$ and of weight $p$ (resp. of length $s$ and of weight $q$), $F_{u\sqcup\!\sqcup v}$ is a quasi-symmetric function of depth $r+s$ and of weight $p+q$, and we have $F_{u\sqcup\!\sqcup v} = F_u\, F_v$. Hence, $\mathrm{H}_{u\sqcup\!\sqcup v} = \mathrm{H}_u\, \mathrm{H}_v$ [11]. In the same way, when $u, v \in Y^* \setminus y_1 Y^*$, we also have $\zeta(u\sqcup\!\sqcup v) = \zeta(u)\,\zeta(v)$ [11].

The Chen's iterated integral over $\omega_0, \omega_1$ associated to the word $w = x_{i_1} \cdots x_{i_k}$ over $X$ and along the integration path $z_0 \rightsquigarrow z$ is defined recursively as follows

$$\int_{z_0\rightsquigarrow z} \omega_{i_1} \cdots \omega_{i_k} = \int_{z_0\rightsquigarrow z} \omega_{i_1}(z_1) \int_{z_0\rightsquigarrow z_1} \omega_{i_2} \cdots \omega_{i_k}. \tag{16}$$

In a shortened notation, we denote this integral by $\alpha_{z_0}^z(w)$ and $\alpha_{z_0}^z(\epsilon) = 1$. One can check that the polylogarithm $\mathrm{Li}_{s_1,\ldots,s_r}$ is also the value of the iterated integral over $\omega_0, \omega_1$ and along the integration path $0 \rightsquigarrow z$ :

$$\mathrm{Li}_w(s_1,\ldots,s_r) = \alpha_0^z(x_0^{s_1-1}x_1 \ldots x_0^{s_r-1}x_1). \tag{17}$$

The definition of polylogarithms is extended over $X^*$ by putting, for $k \in \mathbb{N}$, $\mathrm{Li}_{x_0^k}(z) = \log^k z/k!$.

The functions $\{\mathrm{Li}_w\}_{w\in X^*}$ are $\mathcal{C}$-linearly independent [7]. Thus, the functions defined by $\mathrm{P}_w(z) = (1-z)^{-1}\,\mathrm{Li}_w(z)$, for $w \in X^*$, are also $\mathbb{C}$-linearly independent. Since, for any $w \in Y^*, \mathrm{P}_w$ is the ordinary generating function of $\{\mathrm{H}_w(N)\}_{N\geq 0}$ : $\mathrm{P}_w(z) = \sum_{N\geq 0} \mathrm{H}_w(N)\, z^N$, then $\{\mathrm{H}_w(N)\}_{w\in Y^*}$ are also $\mathbb{C}$-linearly independent. On the one hand, $\ker \mathrm{P} = \{0\}$ and $\ker \mathrm{H} = \{0\}$, and on the other hand, P is a morphism for the Hadamard product:

$$\mathrm{P}_u(z) \odot \mathrm{P}_v(z) = \sum_{N\geq 0} \mathrm{H}_u(N)\mathrm{H}_v(N) z^N = \sum_{N\geq 0} \mathrm{H}_{u \sqcup\!\sqcup v}(N) z^N = \mathrm{P}_{u \sqcup\!\sqcup v}(z). \quad (18)$$

**Proposition 1 ([10]).** *Extended by linearity, the application* $\mathrm{P} : u \mapsto \mathrm{P}_u$ *is an isomorphism from polynomial algebra* $(\mathbb{C}\langle Y\rangle, \sqcup\!\sqcup)$ *over Hadamard algebra* $(\mathbb{C}\{\mathrm{P}_w\}_{w\in Y^*}, \odot)$. *Moreover, the application* $\mathrm{H} : u \mapsto \mathrm{H}_u = \{\mathrm{H}_u(N)\}_{N\geq 0}$ *is an isomorphism from* $(\mathbb{C}\langle Y\rangle, \sqcup\!\sqcup)$ *over the algebra* $(\mathbb{C}\{\mathrm{H}_w\}_{w\in Y^*}, .)$.

Studying the equivalence between the action of $\{(1-z)^l\}_{l\in\mathbb{Z}}$ over $\{\mathrm{P}_w(z)\}_{w\in Y^*}$ and the action of $\{N^k\}_{k\in\mathbb{Z}}$ over $\{\mathrm{H}_w(N)\}_{w\in Y^*}$ (see [9]), we have

**Theorem 1.** *The Hadamard* $\mathcal{C}$*-algebra of* $\{\mathrm{P}_w\}_{w\in Y^*}$ *can be identified with that of* $\{\mathrm{P}_l\}_{l\in\mathcal{L}yn Y}$. *Identically, the algebra of harmonic sums* $\{\mathrm{H}_w\}_{w\in Y^*}$ *with polynomial coefficients can be identified with that of* $\{\mathrm{H}_l\}_{l\in\mathcal{L}yn Y}$.

By Identity (15) and by applying the isomorphism H on the set of Lyndon words $\{y_r\}_{1\leq r\leq k}$, we obtain $\mathrm{H}_{y_1^k}$ as a polynomial in $\{\mathrm{H}_{y_r}\}_{1\leq r\leq k}$ (which are algebraically independent), and

$$\mathrm{H}_{y_1^k} = \sum_{\substack{s_1,\ldots,s_k>0\\ s_1+\ldots+ks_k=k}} \frac{(-1)^k}{s_1!\ldots s_k!}\left(-\frac{\mathrm{H}_{y_1}}{1}\right)^{s_1}\ldots\left(-\frac{\mathrm{H}_{y_k}}{k}\right)^{s_k}. \quad (19)$$

Since the coefficient of $z^N$ in the Taylor expansion of $\mathrm{P}_{y_1^k}$ is $\mathrm{H}_{y_1^k}(N)$ then let

$$\mathrm{Mono}(z) = e^{-(x_1+1)\log(1-z)} = \sum_{k\geq 0} \mathrm{P}_{y_1^k}(z)\; y_1^k \quad (20)$$

$$\mathrm{Const} = \sum_{k\geq 0} \mathrm{H}_{y_1^k}\; y_1^k = \exp\left[-\sum_{k\geq 1} \mathrm{H}_{y_k}\frac{(-y_1)^k}{k}\right]. \quad (21)$$

## 2.2 Noncommutative generating series

The noncommutative generating series $\mathrm{L} = \sum_{w\in X^*} \mathrm{Li}_w\; w$ satisfies the Drinfel'd differential equation [1, 2]

$$d\mathrm{L} = (x_0\omega_0 + x_1\omega_1)\mathrm{L} \quad \text{with} \quad \mathrm{L}(\varepsilon) \underset{\varepsilon\to 0^+}{\sim} e^{x_0\log\varepsilon}. \quad (22)$$

This enables us to prove that L is the exponential of a Lie series [7]. So, applying a Ree's theorem, it verifies Friedrichs criterion [7], *i..e* $\mathrm{Li}_{u \sqcup\!\sqcup v} = \mathrm{Li}_u\; \mathrm{Li}_v$, for

$u, v \in X^*$. In particular, when $u, v \in x_0X^*x_1$, we also have $\zeta(u \sqcup\!\sqcup v) = \zeta(u)\,\zeta(v)$. From the factorization of the monoid by Lyndon words, we obtain the factorization of the series L (see[7]) :

$$\mathrm{L}(z) = e^{x_1 \log \frac{1}{1-z}} \mathrm{L}_{\mathrm{reg}}(z) e^{x_0 \log z} \quad \text{with} \quad \mathrm{L}_{\mathrm{reg}} = \prod_{l \in \mathcal{L}yn X \setminus \{x_0, x_1\}}^{\searrow} e^{\mathrm{Li}_{S_l}\, \check{S}_l}. \tag{23}$$

For $l \in \mathcal{L}yn X \setminus \{x_0, x_1\}$, the polynomial $S_l$ is a finite combination of words in $x_0X^*x_1$. So, let $Z = \mathrm{L}_{\mathrm{reg}}(1)$ and let $\mu$ be the monoid endomorphism verifying $\mu(x_0) = -x_1$, $\mu(x_1) = -x_0$. One also has [8]

$$\mathrm{L}(z) = \mu[\mathrm{L}(1-z)]Z = e^{x_0 \log z} \mu[\mathrm{L}_{\mathrm{reg}}(1-z)] e^{-x_1 \log(1-z)} Z. \tag{24}$$

Thus $\mathrm{L}(z) \underset{z \to 0}{\sim} \exp(x_0 \log z)$ and $\mathrm{L}(z) \underset{z \to 1}{\sim} \exp(-x_1 \log(1-z))\ Z$.

Let $\pi_Y : \mathrm{LI}_{\mathcal{C}}\langle\langle X \rangle\rangle \to \mathrm{LI}_{\mathcal{C}}\langle\langle Y \rangle\rangle$ be a projector such that $\pi_Y(f\ wx_0) = 0$, for $f \in \mathrm{LI}_{\mathcal{C}}, w \in X^*$. Then

$$\Lambda(z) = \pi_Y \mathrm{L}(z) \underset{z \to 1}{\sim} \exp\left(y_1 \log \frac{1}{1-z}\right) \pi_Y Z. \tag{25}$$

By (23), the noncommutative generating series of $\{\mathrm{P}_w\}_{w \in X^*}$ is given by

$$\begin{aligned} \mathrm{P}(z) \;=\; (1-z)^{-1}\mathrm{L}(z) \;&=\; e^{-(x_1+1)\log(1-z)} \mathrm{L}_{\mathrm{reg}}(z) e^{x_0 \log z} && (26)\\ &=\; e^{x_0 \log z} \mu[\mathrm{L}_{\mathrm{reg}}(1-z)] e^{-(x_1+1)\log(1-z)} Z && (27)\\ &=\; e^{x_0 \log z} \mu[\mathrm{L}_{\mathrm{reg}}(1-z)] \mathrm{Mono}(z) Z. && (28) \end{aligned}$$

Thus, $\mathrm{P}(z) \underset{z \to 0}{\sim} e^{x_0 \log z}$ and $\mathrm{P}(z) \underset{z \to 1}{\sim} \mathrm{Mono}(z)Z$. Let $\mathrm{H}(N) = \sum_{w \in Y^*} \mathrm{H}_w(N)\ w$ be the noncommutative generating series of $\{\mathrm{H}_w(N)\}_{w \in Y^*}$ [10]. Then

**Proposition 2.** *$\pi_Y \mathrm{P}(z) \underset{z \to 1}{\sim} \mathrm{Mono}(z) \pi_Y Z$ and $\mathrm{H}(N) \underset{N \to \infty}{\sim} \mathrm{Const}(N) \pi_Y Z$.*

As a consequence of (25), (20) and of Proposition 2, one obtains

**Theorem 2.**

$$\lim_{z \to 1} \exp\left(y_1 \log \frac{1}{1-z}\right) \Lambda(z) = \lim_{N \to \infty} \exp\left(\sum_{k \geq 1} \mathrm{H}_{y_k}(N) \frac{(-y_1)^k}{k}\right) \mathrm{H}(N) = \pi_Y Z.$$

Thus, by the knowledge of Taylor expansion in 0 of $\{\mathrm{P}_w(1-z)\}_{w \in X^*}$ we obtain

**Theorem 3 ([9]).** *For all $g \in \mathcal{C}\{\mathrm{P}_w\}_{w \in Y^*}$, there exist algorithmically computable $c_j \in \mathbb{C}, \alpha_j \in \mathbb{Z}, \beta_j \in \mathbb{N}$ and $b_i \in \mathbb{C}, \eta_i \in \mathbb{Z}, \kappa_i \in \mathbb{N}$ such that*

$$\begin{aligned} g(z) \;&\sim\; \sum_{j=0}^{+\infty} c_j (1-z)^{\alpha_j} \log^{\beta_j}(1-z) \quad \textit{for} \quad z \to 1,\\ [z^n]g(z) \;&\sim\; \sum_{i=0}^{+\infty} b_i n^{\eta_i} \log^{\kappa_i}(n) \qquad\qquad \textit{for} \quad n \to \infty. \end{aligned}$$

**Corollary 1 ([9]).** *Let $\mathcal{Z}$ be the $\mathbb{Q}$-algebra generated by convergent polyzêtas and $\mathcal{Z}'$ be the $\mathbb{Q}[\gamma]$-algebra generated by $\mathcal{Z}$. Then there exist algorithmically computable $c_j \in \mathcal{Z}, \alpha_j \in \mathbb{Z}, \beta_j \in \mathbb{N}$ and $b_i \in \mathcal{Z}', \kappa_i \in \mathbb{N}, \eta_i \in \mathbb{Z}$ such that*

$$\begin{aligned} \forall w \in Y^*, \quad \mathrm{P}_w(z) &\sim \sum_{j=0}^{+\infty} c_j (1-z)^{\alpha_j} \log^{\beta_j}(1-z) \quad \textit{for} \quad z \to 1, \\ \forall w \in Y^*, \quad \mathrm{H}_w(N) &\sim \sum_{i=0}^{+\infty} b_i N^{\eta_i} \log^{\kappa_i}(N) \qquad\qquad \textit{for} \quad N \to +\infty. \end{aligned}$$

**Definition 1 ([6]).** *Let $\zeta_{\sqcup\!\sqcup} : (C\langle\langle X\rangle\rangle, \sqcup\!\sqcup) \to (\mathbb{C}, .)$ be the algebra morphism (*i.e. *for $u, v \in X^*, \zeta_{\sqcup\!\sqcup}(u \sqcup\!\sqcup v) = \zeta_{\sqcup\!\sqcup}(u)\zeta_{\sqcup\!\sqcup}(v)$) verifying for all convergent words $w \in x_0X^*x_1, \zeta_{\sqcup\!\sqcup}(w) = \zeta(w)$, and such that $\zeta_{\sqcup\!\sqcup}(x_0) = \zeta_{\sqcup\!\sqcup}(x_1) = 0$.*

Then, the noncommutative generating series $Z_{\sqcup\!\sqcup} = \sum_{w \in X^*} \zeta_{\sqcup\!\sqcup}(w)\, w$ verifies $Z_{\sqcup\!\sqcup} = Z = \mathrm{L}_{\mathrm{reg}}(1)$ [6]. In consequence, $Z_{\sqcup\!\sqcup}$ is the unique Lie exponential verifying $\langle Z_{\sqcup\!\sqcup} | x_0 \rangle = \langle Z_{\sqcup\!\sqcup} | x_1 \rangle = 0$ and $\langle Z_{\sqcup\!\sqcup} | w \rangle = \zeta(w)$, for any $w \in x_0X^*x_1$.

**Definition 2.** *Let $\zeta_{\text{⧢}} : (\mathbb{C}\langle Y\rangle, \text{⧢}) \to (\mathbb{C}, .)$ the algebra morphism (*i.e. *for all convergent words $u, v \in Y^*, \zeta_{\text{⧢}}(u \text{⧢} v) = \zeta_{\text{⧢}}(u)\zeta_{\text{⧢}}(v)$) verifying for $w \in Y^* \setminus y_1Y^*$, $\zeta_{\text{⧢}}(w) = \zeta(w)$ and such that $\zeta_{\text{⧢}}(y_1) = \gamma$.*

**Lemma 1.** *Let $b_{n,k}(t_1, \ldots, t_{n-k+1})$ be the (exponential) partial Bell polynomials in the variables $\{t_l\}_{l \geq 1}$. If $t_m = (-1)^m(m-1)!\zeta_{\text{⧢}}(m)$, for $m \geq 1$ then*

$$\exp\left[\sum_{k \geq 1} \zeta_{\text{⧢}}(k) \frac{(-y_1)^k}{k}\right] = 1 + \sum_{n \geq 1} \left[\sum_{k=1}^{n} b_{n,k}(\gamma, \zeta(2), 2\zeta(3), \ldots)\right] \frac{(-y_1)^n}{n!}.$$

Let us build the noncommutative generating series of $\zeta_{\text{⧢}}(w)$. Taking the constant part of the two members of $\mathrm{H}(N) \underset{N \to \infty}{\sim} \mathrm{Const}(N)\pi_Y Z$, we have

**Theorem 4.** *Let $Z_{\text{⧢}} = \sum_{w \in Y^*} \zeta_{\text{⧢}}(w)\, w$ be the noncommutative generating series of $\zeta_{\text{⧢}}(w)$. Let $b_{n,k}(t_1, \ldots, t_{n-k+1})$ be the (exponential) partial Bell polynomials in the variables $\{t_l\}_{l \geq 1}$. If $t_m = (-1)^m(m-1)!\zeta_{\text{⧢}}(m)$ then*

$$Z_{\text{⧢}} = \left[1 + \sum_{n \geq 1} \left(\sum_{k=1}^{n} b_{n,k}(\gamma, \zeta(2), 2\zeta(3), \ldots)\right) \frac{(-y_1)^n}{n!}\right] \pi_Y Z.$$

Therefore, by identifying the coefficients of $y_1^k w$ in each member and by using the identity $x_1^k x_0 u = \sum_{l=0}^{k} x_1^l \sqcup\!\sqcup (x_0[(-x_1)^{k-l} \sqcup\!\sqcup u])$, for $u \in X^*$ (see [6]), we get

**Corollary 2.** *For $w \in x_0X^*x_1$,* i.e. *$w = x_0u$ and $\pi_Y w \in Y^* \setminus y_1Y^*$, and for $k \geq 0$, the constant $\zeta_{\text{⧢}}(x_1^k w)$ associated with the divergent polyzêta $\zeta(x_1^k w)$ is a polynomial of degree $k$ in $\gamma$ and with coefficients in $\mathcal{Z}$ :*

$$\zeta_{\text{⧢}}(x_1^k w) = \sum_{i=0}^{k} \frac{\zeta(x_0[(-x_1)^{k-i} \sqcup\!\sqcup u])}{i!} \left[\sum_{j=1}^{i} b_{i,j}(\gamma, -\zeta(2), 2\zeta(3), \ldots)\right].$$

*Moreover, for $l = 0, .., k$, the coefficient of $\gamma^l$ is of weight $|w| + k - l$.*

*In particular, for $s > 1$, the constant $\zeta_{\sqcup\!\sqcup}(1, s)$ associated with $\zeta(1, s)$ is linear in $\gamma$ and with coefficients in $\mathbb{Q}[\zeta(2), \zeta(2i + 1)]_{0<i\leq(s-1)/2}$.*

Applying the (surjective) morphism $\zeta_{\sqcup\!\sqcup}$ on the identity (15), we deduce

**Corollary 3.** *The constant $\zeta_{\sqcup\!\sqcup}(x_1^k)$ associated with the divergent polyzêta $\zeta(x_1^k)$ is a polynomial of degree $k$ in $\gamma$ with coefficients in $\mathbb{Q}[\zeta(2), \zeta(2i+1)]_{0<i\leq(k-1)/2}$ :*

$$\zeta_{\sqcup\!\sqcup}(x_1^k) = \sum_{\substack{s_1,\ldots,s_k>0 \\ s_1+\ldots+ks_k=k}} \frac{(-1)^k}{s_1!\ldots s_k!}(-\gamma)^{s_1}\left(-\frac{\zeta(2)}{2}\right)^{s_2}\ldots\left(-\frac{\zeta(k)}{k}\right)^{s_k}.$$

*Moreover, for $l = 0, .., k$, the coefficient of $\gamma^l$ is of weight $k - l$.*

## 2.3 Polysystem and convergence criterium

Here, generalizing a little, suppose that $A$ is a $\mathbb{C}$-algebra and is a complete normed vector space and that a norm is denoted by $\|.\|$. The set of formal power series (resp. polynomials) on $X$, is denoted by $A\langle\langle X\rangle\rangle$ (resp. $A\langle X\rangle$).

**Definition 3.** *Let $\phi, \chi$ be real positive functions over $X^*$. Let $S \in A\langle\langle X\rangle\rangle$.*

1. *$S$ will be said to be* $\phi-$exponentially majored *if it verifies*

$$\exists K \in \mathbb{R}_+, \exists n \in \mathbb{N}, \forall w \in X^{\geq n}, \|\langle S|w\rangle\| \leq K\phi(w)/|w|!.$$

*We denote by $A^{\phi-em}\langle\langle X\rangle\rangle$ the set of formal power series in $A\langle\langle X\rangle\rangle$ which are $\phi-$exponentially majored.*

2. *$S$ verifies the* $\chi-$growth condition *if it verifies*

$$\exists K \in \mathbb{R}_+, \exists n \in \mathbb{N}, \forall w \in X^{\geq n}, \|\langle S|w\rangle\| \leq K\chi(w)|w|!.$$

*We denote by $A^{\chi-gc}\langle\langle X\rangle\rangle$ the set of formal power series in $A\langle\langle X\rangle\rangle$ verifying the $\chi-$growth condition.*

**Definition 4.** *Let $\mathcal{C}l$ be a class of formal power series in $A\langle\langle X\rangle\rangle$. The power series $S \in A\langle\langle X\rangle\rangle$ is said to be* continuous *over $\mathcal{C}l$ if for any $H \in \mathcal{C}l$, $\sum_{w\in X^*}\langle H|w\rangle\langle S|w\rangle$ is normally convergent. In this case, we denote this sum as $\langle H\|S\rangle$.*

*In particular, let $\phi$ be a real positive function defined over $X^*$, $S$ will be said to be* $\phi-$exponentially continuous *if it is continuous over $A^{\phi-em}\langle\langle X\rangle\rangle$. We denote by $A^{\phi-ec}\langle\langle X\rangle\rangle$ the set of formal power series which are $\phi-$exponentially continuous.*

For any real positive function $\phi$ defined over $X^*$, we have $A\langle X\rangle \subset A^{\phi-ec}\langle\langle X\rangle\rangle$. Otherwise, for $\phi = 0$, we get $A\langle X\rangle = A^{0-ec}\langle\langle X\rangle\rangle$. Hence, any formal power series is $0-$exponentially continuous. If $\phi, \chi$ verify $\sum_{x\in X}\chi(x)\phi(x) < 1$ then $\sum_{w\in X^*}\chi(w)\phi(w)$ is normally convergent. If $F \in A^{\chi-gc}\langle\langle X\rangle\rangle, C \in A^{\phi-em}\langle\langle X\rangle\rangle$ then there exists $K_i \in \mathbb{R}_+, n_i \in \mathbb{N}$ such that for $w \in X^{\geq n_i}, i = 1, 2$, one has $\|\langle F|w\rangle\| \leq K_1\chi(w)|w|!$ and $\|\langle C|w\rangle\| \leq K_2\phi(w)/|w|!$. Hence, $\|\langle F|w\rangle\langle C|w\rangle\| \leq K_1K_2\chi(w)\phi(w)$, for $w \in X^*$ and $|w| \geq \max\{n_1, n_2\}$. In other terms,

**Proposition 3 ([4]).** *Suppose two real positive morphisms over $X^*$, $\phi$ and $\chi$ verify the condition $\sum_{x\in X}\chi(x)\phi(x)<1$. Then, for $S\in A^{\chi-gc}\langle\langle X\rangle\rangle$, $S$ is continuous over $A^{\phi-em}\langle\langle X\rangle\rangle$.*

Let $q_1,\ldots,q_n$ be commutative indeterminates over $\mathbb{C}$. The algebra of formal power series (resp. polynomials) over $Q=\{q_1,\ldots,q_n\}$ with coefficients in $\mathbb{C}$ is denoted by $\mathbb{C}[[Q]]$ (resp. $\mathbb{C}[Q]$).

**Definition 5.** *Let $f=\sum_{i_1,\ldots,i_n\geq 0} f_{i_1,\ldots,i_n}q_1^{i_1}\ldots q_n^{i_n}\in\mathbb{C}[[Q]]$. We set*

$$\begin{array}{rcl} E(f) & = & \{\rho\in\mathbb{R}_+^n : \exists C_{f\rho}\in\mathbb{R}_+ \ st\ \forall i_1,\ldots,i_n\geq 0, |f_{i_1,\ldots,i_n}|\rho^{i_1}\ldots\rho^{i_n}\leq C_{f\rho}\}\\ \check{E}(f) & : & \textit{interior of } E(f) \textit{ in } \mathbb{R}^n.\\ \mathrm{CV}(f) & = & \textit{convergence domain of } f \ = \ \{q\in\mathbb{C}^n : (|q_1|,\ldots,|q_n|)\in\check{E}(f)\}. \end{array}$$

*The power series $f$ is said to be* convergent *if $\mathrm{CV}(f)\neq\emptyset$. Let $\mathcal{U}$ be an open domain in $\mathbb{C}^n$ and let $q\in\mathbb{C}^n$. The power series $f$ is said to be convergent on $q$ (resp. over $\mathcal{U}$) if $q\in\mathrm{CV}(f)$ (resp. $\mathcal{U}\subset\mathrm{CV}(f)$). We set $\mathbb{C}^{\mathrm{cv}}[[Q]]=\{f\in\mathbb{C}[[Q]] : \mathrm{CV}(f)\neq\emptyset\}$. Let $q\in\mathrm{CV}(f)$. There exist constants $C_{f\rho},\rho$ and $\check{\rho}$ such that $|q_1|<\check{\rho}<\rho,\ldots,|q_n|<\check{\rho}<\rho$ and $|f_{i_1,\ldots,i_n}|\rho^{i_1+\ldots+i_n}\leq C_{f\rho}$, for $i_1\ldots,i_n\geq 0$. The* convergence module *of $f$ at $q$ is $(C_{f\rho},\rho,\check{\rho})$.*

Suppose that $\mathrm{CV}(f)\neq\emptyset$ and let $q\in\mathrm{CV}(f)$. If $(C_{f\rho},\rho,\check{\rho})$ is a convergence module of $f$ at $q$ then $|f_{i_1,\ldots,i_n}q_1^{i_1}\ldots q_n^{i_n}|\leq C_{f\rho}(\check{\rho}/\rho)^{i_1+\ldots+i_n}$. Hence, at $q$, the power series $f$ is majored term by term by $C_{f\rho}\prod_{k=0}^m(1-\check{\rho}_k/\rho_k)^{-1}$. Therefore, $f$ is uniformly absolutely convergent in $\{q\in\mathbb{C}^n : |q_1|<\check{\rho},\ldots,|q_n|<\check{\rho}\}$ which is an open domain in $\mathbb{C}^n$. Thus, $\mathrm{CV}(f)$ is an open domain in $\mathbb{C}^n$. Since the partial derivation of order $j_1,\ldots,j_n\geq 0$ of $f$ is estimated by $\|D_1^{j_1}\ldots D_n^{j_n}f\|\leq C_{f\rho}\partial^{j_1+\ldots+j_n}/\partial\check{\rho}^{j_1+\ldots+j_n}\prod_{k=0}^m(1-\check{\rho}_k/\rho_k)^{-1}$ then $\mathrm{CV}(f)\subset\mathrm{CV}(D_1^{j_1}\ldots D_n^{j_n}f)$.

Let $f\in\mathbb{C}^{\mathrm{cv}}[[Q]]$ and let $\{A_i\}_{i=0,1}$ be a polysystem. Let $(\rho,\check{\rho},C_f)$ and let $\{(\rho,\check{\rho},C_i)\}_{i=0,1}$ be convergence modules at $q\in\mathrm{CV}(f)\Cap_{i=0,1,j=1,..,n}\mathrm{CV}(A_i^j)$ of $f$ and $\{A_i^j\}_{j=1,..,n}$ respectively. Let $X$ be an alphabet in bijection with $\{A_i\}_{i=0,1}$. We put $A_\epsilon=$ identity and $C_\epsilon=1$. For any word $w=vx_i, x_i\in X, v\in X^*$, we set $A_w=A_vA_i$ and $C_w=C_vC_i$. These notations are extended to $A\langle\langle X\rangle\rangle$ and we will denote by $A_wf_{|q}$ the evaluation at $q$ of $A_wf$. Then [4]

$$|A_wf_{|q}|\leq\frac{C_f}{(1-r)^n}\frac{C_w|w|!}{\binom{n+|w|-1}{|w|}}\left[\frac{n(n+1)}{\tau(1-r)^{n+1}}\right]^{|w|},\tag{29}$$

where, $\tau=\min_{1\leq k\leq n}\rho_k$ and $r=\max_{1\leq k\leq n}\check{\rho}_k/\rho_k$. Therefore,

**Theorem 5 ([4]).** *Let $K=C_f(1-r)^{-n}$ and let $\chi$ be the real positive function defined over $X^*$ by $\chi(x_i)=C_in(n+1)\tau^{-1}(1-r)^{-(n+1)}$, for $i=0,1$. Then the generating series $\sum_{w\in X^*}A_wf_{|q}\,w$ satisfies the $\chi-$growth condition.*

## 2.4 Polysystem and nonlinear differential equation

Here, the observation $f$ of the nonlinear dynamical system (2) is an element of $\mathbb{C}^{\mathrm{cv}}[[Q]]$ and the analytic vector fields $\{A_i\}_{i=0,1}$ (3) constitute a polysystem.

Let $(\rho, \check{\rho}, C_f)$ and let $(\rho, \check{\rho}, C_i)$, for $i = 0, 1$, be convergence modules of $f$ and $\{A_i^j\}_{j=1,..,n}$ respectively at $q \in \mathrm{CV}(f) \Cap_{i=0,1,j=1,..,n} \mathrm{CV}(A_i^j)$.

Let $\sigma f_{|_{q_0}} = \sum_{w \in X^*} A_w f_{|_{q_0}}\, w$ be the Fliess' generating series of (2) satisfying the $\chi-$growth condition and let $S_{z_0 \rightsquigarrow z} = \sum_{w \in X^*} \alpha_{z_0}^z(w)\, w$ be the Chen's generating series of inputs $u_0, u_1$. Let $\varepsilon \in ]0, 1[$ and let $z_i = \varepsilon \exp(\mathrm{i}\theta_i)$, for $i = 0, 1$. We set $\theta = \theta_1 - \theta_0$. Let $\Gamma_0(\varepsilon, \theta)$ (resp. $\Gamma_1(\varepsilon, \theta)$) be the path turning around 0 (resp. 1) in the positive direction from $z_0$ to $z_1$.

By induction on the length $|w|$ of $w$, one has $|\langle S_{\Gamma_i(\varepsilon,\theta)}|w\rangle| = (2\varepsilon)^{|w|_{x_i}} \theta^{|w|}/|w|!$, where $|w|_{x_i}$, denotes the number of occurrences of letter $x_i$ in $w$, for $i = 0, 1$. For $\varepsilon \to 0^+$, these estimation yield $S_{\Gamma_i(\varepsilon,\theta)} = e^{\mathrm{i}\theta x_i} + o(\varepsilon)$. In particular, if $\Gamma_0(\varepsilon)$ (resp. $\Gamma_1(\varepsilon)$) is a circular path of radius $\varepsilon$ turning around 0 (resp. 1) in the positive direction, starting at $z = \varepsilon$ (resp. $1 - \varepsilon$), then, by the noncommutative residu theorem [7, 5], we get $S_{\Gamma_0(\varepsilon)} = e^{2\mathrm{i}\pi x_0} + o(\varepsilon)$ and $S_{\Gamma_1(\varepsilon)} = e^{-2\mathrm{i}\pi x_1} + o(\varepsilon)$. The Chen's generating series $S_{z_0 \rightsquigarrow z}$ also satisfies the following differential equation

$$dS = (x_0\omega_0 + x_1\omega_1)S \quad \text{with} \quad S(0) = 1. \tag{30}$$

Thus, $S_{z_0 \rightsquigarrow z}$ and $\mathrm{L}(z)\mathrm{L}(z_0)^{-1}$ satisfy (22) taking the same value at $z_0$ and $S_{z_0 \rightsquigarrow z} = \mathrm{L}(z)\mathrm{L}(z_0)^{-1}$. Finally, the asymptotic behavior of L gives

**Proposition 4 ([7, 5]).** $S_{\varepsilon \rightsquigarrow 1-\varepsilon} \underset{\varepsilon \to 0^+}{\sim} e^{-x_1 \log \varepsilon} Z e^{-x_0 \log \varepsilon}$.

The Fliess' fundamental formula [3] can be then extended and gives the output $y$ of (2) as follows

**Theorem 6.** $y(z) = \langle \sigma f_{|_{q_0}} \| S_{z_0 \rightsquigarrow z} \rangle = \sum_{w \in X^*} \langle A_w f_{|_{q_0}} | w \rangle \langle S_{z_0 \rightsquigarrow z} | w \rangle$.

Using the factorization indexed by Lyndon words of the Lie exponential series L (23), we deduce some expansions of the Chen's generating series $S_{z_0 \rightsquigarrow z}$ and we obtain finally

**Corollary 4.** *The output $y$ of the nonlinear dynamical system with singular inputs admits then the following functional expansions*

$$\begin{aligned}
y(z) &= \sum_{w \in X^*} g_w(z)\, A_w \circ f_{|q_0}, \\
&= \sum_{k \geq 0} \sum_{n_1, \ldots, n_k \geq 0} g_{x_0^{n_1} x_1 \ldots x_0^{n_k} x_1}(z)\ \mathrm{ad}_{A_0}^{n_1} A_1 \ldots \mathrm{ad}_{A_0}^{n_k} A_1 e^{\log z A_0} \circ f_{|q_0}, \\
&= \prod_{l \in \mathcal{L}yn X} \exp\Big( g_{S_l}(z)\, A_{\check{S}_l} \circ f_{|q_0} \Big), \\
&= \exp\Big( \sum_{w \in X^*} g_w(z)\, A_{\pi_1(w)} \circ f_{|_{q_0}} \Big),
\end{aligned}$$

*where, for any word $w$ in $X^*$, $g_w$ belongs to a polylogarithm algebra and $\pi_1(w)$ is the following Lie series*

$$\pi_1(w) = \sum_{k \geq 1} \frac{(-1)^{k-1}}{k} \sum_{u_1, \cdots, u_k \in X^* \setminus \{\epsilon\}} \langle w | u_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup u_k \rangle\ u_1 \cdots u_k.$$

# 3 Conclusion

By Corollary 4, the output $y$ of the nonlinear dynamical system (2) is a combination of elements $g$ belonging to a completion of the polylogarithm algebra. If its Taylor expansion is given as in (1) then, by Corollary 1, the coefficients of this expansion belong to a completion of the harmonic algebra and there exist algorithmically computable coefficients $a_i \in \mathbb{Z}$, $b_i \in \mathbb{N}$ and $c_i$, belonging to the $\mathbb{C}$-algebra generated by $\mathcal{Z}$ and by Euler's $\gamma$ constant, such that

$$y_n \underset{n\to\infty}{\sim} \sum_{i\geq 0} c_i n^{a_i} \log^{b_i} n. \tag{31}$$

# References


[1] Drinfel'd.– Quasi-Hopf Algebras, *Leningrad Math. J.*, 1, 1419-1457, 1990.

[2] Drinfel'd– On quasitriangular quasi-hopf algebra and a group closely connected with $\mathrm{gal}(\bar{q}/q)$, *Leningrad Math. J.*, 4, 829-860, 1991.

[3] Fliess M.– *Fonctionnelles causales non linéaires et indéterminées non commutatives*, Bull. SMF, 109, pp. 3-40, 1981.

[4] Hoang Ngoc Minh– *Contribution au développement d'outils informatiques pour résoudre des problèmes d'automatique non linéaire*, Thèse, Lille, 1990.

[5] Hoang Ngoc Minh, Jacob G., N.E. Oussous, M. Petitot– Aspects combinatoires des polylogarithmes et des sommes d'Euler-Zagier, *Journal Electronique du Séminaire Lotharingien de Combinatoire B43e*, 2000.

[6] Hoang Ngoc Minh, Jacob G., N.E. Oussous, M. Petitot– De l'algèbre des $\zeta$ de Riemann multivariées à l'algèbre des $\zeta$ de Hurwitz multivariées, *Journal Electronique du Séminaire Lotharingien de Combinatoire*, 44, 2001.

[7] Hoang Ngoc Minh, Petitot, M. and Van der Hoeven, J.– Polylogarithms and Shuffle Algebra, *Proceedings of FPSAC'98*, 1998.

[8] Hoang Ngoc Minh, Petitot, M., and Van der Hoeven, J.– L'algèbre des polylogarithmes par les séries génératrices, *Proceedings of FPSAC'99*, 1999.

[9] Costermans, C., Enjalbert, J.Y. and Hoang Ngoc Minh– Algorithmic and combinatoric aspects of multiple harmonic sums, *Discrete Mathematics & Theoretical Computer Science Proceedings*, 2005.

[10] Hoang Ngoc Minh– Finite polyzêtas, Poly-Bernoulli numbers, identities of polyzêtas and noncommutative rational power series, *Proceedings of* $4^{th}$ *International Conference on Words*, pp. 232-250, 2003.

[11] Hoffman, M.– The algebra of multiple harmonic series, *J. of Alg.*, 1997.

[12] Reutenauer, C.– *Free Lie Algebras*, London Math. Soc. Monographs, New Series-7, Oxford Science Publications, 1993.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# Stability of Equilibria in the Gravitational Field of a Finite Body. 1:2 Resonance

Antonio Elipe[1], Víctor Lanchares[2]

[1] *Grupo de Mecánica Espacial. IUMA. University of Zaragoza. Spain*
[2] *Dpt. Mathematics and Computation. University of La Rioja. Spain*

**Abstract.** We analyze the Lyapunov stability of a particle moving under the gravitational field of a finite body. We complete previous results by determining the stability of 1:2 resonance.

## 1 Introduction

Geostationary points are very important in space missions, since many satellites, especially for communications, are placed in neighborhoods of such points [1]. Besides, there are space missions to other celestial bodies different from the Earth, hence the importance of determining the stability of stationary points for other bodies like planets, natural satellites, etc.

In order to have a general result, it would be desirable to consider several terms of the potential symbolically, and once the theory is elaborated one can obtain the stability for any particular case by replacing the symbols by the specific numerical values of the body considered. The first attempt to work in this direction was done by Deprit and López-Moratalla [2, 3], who formulated the problem from scratch. By considering the first tesserals in the classical expansion in Legendre polynomials of the potential, they found the possible equilibria in a synodic frame and computed the linear stability, showing that among the six possible equilibria (although for the Earth there are only four) four of them are linearly unstable, and hence Lyapunov unstable, whereas the other two (the ones placed on the $y$ axis) are linear stable in some regions of the parametric plane (the parameters depend essentially on the $C_{20}$ $C_{22}$ harmonics, and on the distance of the equilibrium to the origin; for more details, see [4]).

To determine the non-linear stability, Deprit and López-Moratalla [2] followed the scheme given in [5], that is: Expand the Hamiltonian around the equilibrium; next compute the normal form of the expanded Hamiltonian; then apply Arnold theorem [6] to determine whether the equilibrium is stable or unstable in a non-linear sense. However, since the Arnold theorem is not valid for resonances, these cases must be analyzed in a different way.

In [4] the work done in [2] is extended by considering some cases that require higher order normalization and some resonances, although some specific resonance cases are left without analysis. In this communication, we complete the above-mentioned work [4] by analyzing 1:2 resonance.

## 2 Equations of motion and equilibria

Let us consider the motion of a satellite with respect to a synodic frame that is rotating, as the planet does, around the axis of greatest inertia with angular velocity $\nu$. The Hamiltonian for this problem is

$$\mathcal{H} = \frac{1}{2}(X^2 + Y^2 + Z^2) - \nu(xY - yX) + \mathcal{V}(x, y, z), \tag{1}$$

with potential $\mathcal{V}(x, y, z)$ defined by

$$\mathcal{V} = -\frac{\mu}{r}\left[1 + \left(\frac{\oplus}{r}\right)^2 \left\{3\,\Gamma_{2,2}\frac{x^2 - y^2}{r^2} - \tfrac{1}{2}\Gamma_{2,0}\left(1 - 3\frac{z^2}{r^2}\right)\right\}\right], \tag{2}$$

where $\mu$ is the Gaussian constant, $r = \sqrt{x^2 + y^2 + z^2}$ is the radial distance of the satellite, $\oplus$ the planet's equatorial radius and where the harmonic coefficients are such that $\Gamma_{2,0} < 0 < \Gamma_{2,2}$, because of planetary oblateness and our choice of the axes.

Equilibria are found by zeroing the corresponding Hamilton equations. It is shown in [2] that, in general, there are 6 equilibria on the plane $xy$, two on the $x$ axis ($E_1(\pm r_1, 0)$), and four on the $y$ axis ($E_2(\pm r_2, 0)$ and $E_2'(\pm r_2', 0)$).

By expanding the Hamiltonian in Taylor series around the equilibria, one obtains

$$\mathcal{H}^{(j)} = \mathcal{H}_2^{(j)} + \mathcal{H}_3^{(j)} + \mathcal{H}_4^{(j)} + \dots \tag{3}$$

where $\mathcal{H}_n^{(j)}$ is a homogeneous polynomial of degree $n$ in the new variables (variations). The script $j = 1, 2$ depending on the equilibrium considered. The quadratic part is

$$\mathcal{H}_2^{(j)} = \frac{1}{2}(\Xi_j^2 + H_j^2) - \omega(\xi_j H_j - \eta_j \Xi_j) + \frac{1}{2}\omega^2(\alpha_j\,\xi_j^2 + \beta_j\,\eta_j^2) = \frac{1}{2}\boldsymbol{\zeta}_j A_j \boldsymbol{\zeta}_j, \tag{4}$$

and the parameters $\alpha_j$, $\beta_j$ are

$$\alpha_j = 1 - 12\,(-1)^j \Gamma_{22}\,\frac{a_k^3 \oplus^2}{r_j^5} \qquad \text{and} \qquad \beta_j = 2\left(\frac{a_k^3}{r_j^3} - 2\right),$$

with $a_k$ the semimajor axis of a Keplerian ellipse with mean motion equal to $\nu$.

From here on, we will drop the index $j$.

This quadratic Hamiltonian provides the linearized equations of motion $\dot{\boldsymbol{\zeta}} = JA\boldsymbol{\zeta} = B\boldsymbol{\zeta}$, where $J$ is the standard symplectic matrix. The eigenvalues of matrix $B$, i.e. the roots of the characteristic equation

$$\det(\lambda I - B) = \lambda^4 + \nu^2(\alpha + \beta + 2)\lambda^2 + \nu^4(1 - \alpha)(1 - \beta) = 0,$$

will determine the stability character of the equilibria. The roots are

$$\begin{aligned} \omega_1^2 &= (\nu^2/2)\left[-(\alpha + \beta + 2) + \sqrt{(\alpha - \beta)^2 + 8(\alpha + \beta)}\right], \\ \omega_2^2 &= (\nu^2/2)\left[-(\alpha + \beta + 2) - \sqrt{(\alpha - \beta)^2 + 8(\alpha + \beta)}\right]. \end{aligned} \tag{5}$$

As proven in [2], points $E_1$ and $E_2'$ are unstable everywhere, whereas points $E_2$ enjoy linear stability at some specific regions. To determine its Lyapunov stability, that is, the influence of the remaining terms of the Hamiltonian (3), Arnold's theorem [6] is used.

**Theorem 1 (Arnold)** *Consider a two-degrees-of-freedom Hamiltonian system* $\mathcal{H}$, *expressed in real canonical coordinates* $(\Phi_1, \Phi_2, \phi_1, \phi_2)$, *as*

$$\mathcal{H} = \mathcal{H}_2 + \mathcal{H}_4 + \ldots + \mathcal{H}_{2n} + \tilde{\mathcal{H}},$$

*where:*

1. $\mathcal{H}$ *is real analytic in a neighborhood of the origin* $\mathbb{R}^4$,
2. $\mathcal{H}_{2k}$, $1 \leq k \leq n$, *is a homogeneous polynomial of degree* $k$ *in* $\Phi_i$, *with real coefficients. In particular,*

$$\mathcal{H}_2 = \omega_1\Phi_1 - \omega_2\Phi_2, \quad 0 < \omega_1,\, 0 < \omega_2;$$
$$\mathcal{H}_4 = \frac{1}{2}\left(A\Phi_1^2 - 2B\Phi_1\Phi_2 + C\Phi_2^2\right).$$

3. $\tilde{\mathcal{H}}$ *has a power expansion in* $\Phi_i$ *which starts with terms at least of order* $2n+1$.

*Under these assumptions, the origin is a stable equilibrium provided for some* $k$, $2 \leq k < n$, $\mathcal{H}_2$ *does not divide* $\mathcal{H}_{2k}$ *or equivalently, provided* $D_{2k} = \mathcal{H}_{2k}(\omega_2, \omega_1) \neq 0$ *and for* $2 \leq j < k$, $D_{2j} = \mathcal{H}_{2j}(\omega_2, \omega_1) = 0$.

This theorem assumes that there are no resonances in the principal part of the Hamiltonian, and that the Hamiltonian is expressed in its normal form. The study for non-resonant cases was made in [2, 4], and we will not repeat it here. However, we will reproduce the symplectic transformation that converts the quadratic Hamiltonian (4) in the simple form of two oscillators as in Arnold's theorem. This transformation

$$\boldsymbol{w} = (u_1, u_2, U_1, U_2) \longmapsto \boldsymbol{\zeta} = (\xi, \eta, \Xi, H),$$

valid even for resonances, is defined by the linear transformation $\boldsymbol{\zeta} = B\boldsymbol{w}$, with

$$B = \begin{pmatrix} ia_1 & -ia_2 & a_1 & a_2 \\ -b_1 & b_2 & -ib_1 & -ib_2 \\ b_1\nu - a_1\omega_1 & a_2\omega_2 - b_2\nu & -i(a_1\omega_1 - b_1\nu) & -i(a_2\omega_2 - b_2\nu) \\ i(a_1\nu - b_1\omega_1) & -i(a_2\nu - b_2\omega_2) & a_1\nu - b_1\omega_1 & a_2\nu - b_2\omega_2 \end{pmatrix},$$

and $a_1, a_2, b_1, b_2$ given by the expressions

$$a_1^2 = \frac{\omega_1^2 + \omega^2(1-\beta)}{2\omega_1(\omega_1^2 - \omega_2^2)}, \qquad b_1^2 = \frac{\omega_1^2 + \omega^2(1-\alpha)}{2\omega_1(\omega_1^2 - \omega_2^2)},$$
$$a_2^2 = \frac{\omega_2^2 + \omega^2(1-\beta)}{2\omega_2(\omega_1^2 - \omega_2^2)}, \qquad b_1^2 = \frac{\omega_2^2 + \omega^2(1-\alpha)}{2\omega_2(\omega_1^2 - \omega_2^2)}.$$

After a significant amount of symbolic algebra, it can be proved that Eq. (4) becomes

$$\mathcal{H}_2 = i\,\omega_1 u_1\,U_1 + i\,\omega_2 u_2\,U_2, \tag{6}$$

and each term $\mathcal{H}_k$ of the Hamiltonian (3) is a homogeneous polynomial of degree $k$ in the variables $\boldsymbol{w} = (u_1, u_2, U_1, U_2)$.

By making use of the Poincaré variables $(\phi_1, \phi_2, \Phi_1, \Phi_2)$, related with the complex variables above defined by the symplectic transformation

$$\begin{aligned} u_1 &= \sqrt{\Phi_1}\exp(i\phi_1), & u_2 &= \sqrt{\Phi_2}\exp(-i\phi_2),\\ U_1 &= -i\sqrt{\Phi_1}\exp(-i\phi_1), & U_2 &= i\sqrt{\Phi_2}\exp(i\phi_2), \end{aligned} \tag{7}$$

the linearized Hamiltonian (4) reduces to

$$\mathcal{H}_2 = \omega_1\Phi_1 - \omega_2\Phi_2, \tag{8}$$

## 3 1:2 resonance

If there is a commensurability among the frequencies given in (5) we face the problem of *resonances*. For instance if $\omega_1 = 2\omega_2$ we have the resonance 1:2, that is, a resonance of third order. In this case, $\alpha$ and $\beta$ are related by

$$3(\alpha+\beta+2) + 5\sqrt{(\alpha-\beta)^2 + 8(\alpha+\beta)} = 0,$$

or equivalently, by

$$\beta = \frac{1}{8}\left[(-41 + 17\,\alpha) + 5\,\sqrt{73 - 82\,\alpha + 9\,\alpha^2}\right]. \tag{9}$$

The normalized Hamiltonian will be made of the kernel of the Lie derivative $L_2$. In the algebra of homogeneous polynomials in $(u, v, U, V)$,

$$L_2\,(u_1^m U_1^n u_2^p U_2^q) = [i\,\omega_1\,(m-n) + i\,\omega_2\,(p-q)]\,u_1^m U_1^n u_2^p U_2^q.$$

Thus, monomials of the type $(u_1\,U_1)^m (u_2\,U_2)^p$ belong to $\ker L_2$. Besides, due to the resonance $(\omega_1 = 2\omega_2 = \omega)$,

$$L_2\,(u_1^m U_1^n u_2^p U_2^q) = i\,\omega\,[2(m-n) + (p-q)]\,u_1^m U_1^n u_2^p U_2^q,$$

and hence, $\ker L_2$ also contains resonant terms of the type $U_1 u_2^2$ and $u_1 U_2^2$. Thus, the normal form is generated by the following invariants with respect to $\mathcal{H}_2$

$$I_1 = u_1 U_1, \quad I_2 = u_2 U_2, \quad I_3 = u_1 U_2^2, \quad I_4 = U_1 u_2^2.$$

Hence, normalization up to first order only contains the terms in $\mathcal{H}_3$ corresponding to these monomials. In this way, we have

$$\mathcal{H}_3 = c_1(u_1 U_2^2 - i U_1 u_2^2), \tag{10}$$

where

$$c_1 = \frac{1}{2r_2}\left[a_2(-4+5\alpha-\beta)(a_2b_1-2a_1b_2)-(8+7\beta)b_1b_2^2\right],$$

and $\beta$ is given by (9).

To study the stability in the resonant cases, we will use a result given in [8], but before doing so, we need to define the real functions [7, 8]

$$M_1 = \frac{i}{2}(2I_1+I_2), \quad M_2 = \frac{i}{2}(2I_1-I_2), \quad S = \frac{i}{\sqrt{2}}(I_4+iI_3), \quad C = \frac{1}{\sqrt{2}}(I_4-iI_3),$$

that satisfy the relation

$$C^2+S^2 = (M_1+M_2)(M_1-M_2)^2.$$

The normalized Hamiltonian (10) may be expressed as

$$\mathcal{H}_3 = \sqrt{2}\,c_1\,S.$$

Now, we can apply the *geometric criterion* given by the authors in [8]:

**Theorem 2 (Geometric criterion)** *Let us suppose that the Hamiltonian is normalized up to a certain order $N \geq r$, the first term that does not vanish for $M_2 = 0$, and let us consider the two surfaces $\mathcal{G}_1 \equiv C^2+S^2 = M_1^r$ and $\mathcal{G}_2 \equiv H(C,S,M_1;M_2=0) = 0$. Then, if their only common point is the origin, this point is stable. If they transversely intersect each other, the origin is unstable.*

Thus, whenever $c_1 \neq 0$, by the geometric stability criterion, the equilibrium position is unstable. Otherwise, the stability is decided by higher order terms in the normalization.

In Figure 1 we plot coefficient $r_2c_1$ as a function of $\alpha$. By solving the equation $c_1 = 0$ we find the root at $\alpha = \alpha_0 = 0.9823588648766484$. Consequently, at resonance 1:2, the equilibrium is unstable everywhere except (perhaps) at $\alpha_0$. For this particular case, we need to push forward the normal form.

After some computations, for $\alpha = \alpha_0$, the next non-null term of the normalized Hamiltonian, expressed as a function of the invariants $M_1$, $M_2$, $C$ and $S$, is

$$\mathcal{H}_4 = \frac{1}{r_2^2}\left(5.69310754306\,M_1^2 - 6.24186562149\,M_1M_2 + 0.52415122257\,M_2^2\right).$$

Applying the geometric criterion, that is, by setting $M_2 = 0$ in $\mathcal{H}_4 = 0$, the function $\mathcal{G}_2$ is

$$\mathcal{G}_2 \equiv \frac{1}{r_2^2}\left(5.69310754306\,M_1^2\right),$$

which intersects with $\mathcal{G}_1 \equiv C^2+S^2 = M_1^4$ only at the origin. In consequence, at the resonance, we have *instability* except at $\alpha = \alpha_0$, where the equilibrium is stable.

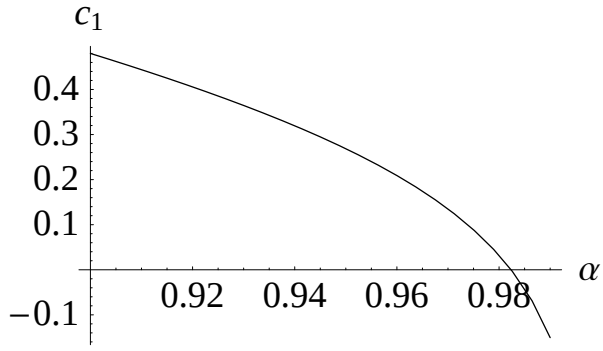



Figure 1: Plot $r_2 c_1$ versus $\alpha$.

It is worth noting that this result could be obtained without any computation of the normal form. In fact, if $c_1 = 0$, $\mathcal{H}_3$ contains no resonant monomial. As a consequence, the first order normalized Hamiltonian coincides with that obtained by means of Birkhoff normalization in [4]. Moreover, the generating function also coincides, provided no resonant terms exist. But, for a third-order resonance, there are no resonant monomials of degree four, and it follows that the second-order normalized Hamiltonian must be given by the expression computed in [4].

This expression provides us $D_4$, the information required by Arnold's theorem to decide the stability. But this is also the information required by the geometric criterion. So, in this case both theorems must give the same information, as stated in [8]. However, $D_4$ is given by the quotient of two polynomials. The polynomial in the numerator vanishes for $\alpha = \alpha_0$, as it was noticed in [4]. But, the denominator also vanishes (indeed it vanishes for every $\alpha$ corresponding to a third-order resonance). Thus, the factor $2\omega_1 - \omega_2$ appears in both numerator and denominator of $D_4$ and it cancels out giving rise to a non-zero value of $D_4$. Consequently, the equilibrium is stable in this particular case.

**Acknowledgments.** Supported by the Spanish Ministry of Science and Technology (Projects # ESP2005-07107 and MTM2005-08595).

# References

[1] Soop, E.M., *Handbook of geostationary orbits*. Space Technology Library Series. Kluwer Academic Pub. Dordrecht, 1994.

[2] Deprit, A. and López-Moratalla, T., "Estabilidad orbital de satélites estacionarios," *Revista Matemática Universidad Complutense*, Vol. 9, 1996, pp. 311–333.

[3] López-Moratalla, T., "Estabilidad orbital de satélites estacionarios," Ph. D. Dissertation (in Spanish). Univ. de Zaragoza. Spain, *Boletín ROA*, Vol. 5/97, 1997.

[4] Elipe, A. and López-Moratalla, T., "On the Lyapunov stability of stationary points around a central body," *J. Guidance, Control and Dynamics*, Vol. 29, 2006, pp. 1376–1383.

[5] Deprit, A. and Deprit-Bartholomé, A., "Stability of the Triangular Lagrangian Points," *Astronomical Journal*, Vol. 72, 1967, pp. 173–179.

[6] Arnold, V.I., "The stability of the equilibrium position of a Hamiltonian system of ordinary differential equations in the general elliptic case." *Soviet Math. Dokl.* Vol. 2, 1961, pp. 247–249.

[7] Elipe, A., Lanchares, V., López-Moratalla, T. and Riaguas, A., "Nonlinear stability in resonant cases: A geometrical approach," *Journal of Nonlinear Sciences*, Vol. 11, 2001, pp. 211–222.

[8] Elipe, A., Lanchares, V. and Pascual, A.I., "On the stability of equilibria in two degrees of freedom Hamiltonian systems," *Journal of Nonlinear Sciences*, Vol. 15, 2005, pp. 305–319.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# An Analytical Model for a Lunar Orbiter

A. Abad[1], A. Elipe[1], J.F. San-Juan[2], M. San-Martín[2]

[1] *Grupo de Mecánica Espacial. IUMA. University of Zaragoza. Spain*
[2] *Dpt. Mathematics and Computation. University of La Rioja. Spain*

**Abstract.** By means of analytical theories based on Lie-Deprit transforms, we obtain families of periodic orbits for the zonal problem of an orbiter around the moon. We recover previous results and find new orbits.

## 1 Introduction

The problem of finding periodic orbits about the moon, natural satellites or asteroids is of interest, since several space missions are planned with the goal of orbiting such bodies (see for instance [11, 14, 12] and references therein). Among the possible orbits, frozen orbits are very useful; indeed, frozen orbits are characterized by constant eccentricity, inclination and pericenter direction on the average; thus, they are very convenient for reconnaissance. Several procedures are used to find frozen orbits, from *brute force*, that is, by zeroing Gauss equations and solving the corresponding system, to more sophisticated methods like the numerical continuation of families of periodic orbits, or by averaging the Hamiltonian and finding the equilibria of the reduced Hamiltonian. This later procedure is the one we will use in this work. It has been successfully used for finding frozen orbits for the zonal problem of an Earth artificial satellite [3, 2] and has yielded insight on so-called *critical inclination* [1].

In general, for orbiters around the moon, it is necessary to consider the third body attraction, that is, the earth's attraction, which means that time appears in the Hamiltonian. However, as proved in [9], for orbits whose altitude is below 100 km, the third body attraction is almost negligible; thus, for a low-altitude satellite, which is the case studied in this note, we will only consider the moon's gravitational effect, and more precisely, the zonal terms.

In Section 3 we give a brief description of the Lie-Deprit transforms we use, namely, the Delaunay normalization and elimination of the parallax. The analytical integration is done in Section 4, and finally, in Section 5, frozen orbits are determined for low-altitude satellites. The case of high-altitude satellites will appear elsewhere.

## 2 Dynamical model

We consider the motion of a Moon orbiter under the Moon's gravitational force.

Let us consider a rotating reference frame $Oxyz$, centered on the Moon and such that the plane $Oxy$ coincides with the Moon's equator and the axis $Ox$ continuously points towards the Earth, which is moving in a circular orbit with radius $a_e$ synchronized with the rotation of the Moon. Let $\boldsymbol{\omega}$ be the angular velocity vector of the Moon; thus, the Hamiltonian of an orbiter is

$$\mathcal{H} = \frac{1}{2}\boldsymbol{X}\cdot\boldsymbol{X} - \boldsymbol{\omega}\cdot(\boldsymbol{x}\times\boldsymbol{X}) + V_m,$$

where $V_m$ is the gravitational potential originated by the Moon. For the Moon we only take into account the zonal contribution; thus,

$$V_m = -\frac{\mu}{r} + \frac{\mu}{r}\sum_{n\geq 2}\frac{\alpha}{r}J_n P_n\left(z/r\right),$$

with $\alpha$ the equatorial radius of the Moon, $r = \|\boldsymbol{x}\|$, $\mu$ the Gaussian constant for the Moon, $J_n$ the Moon harmonics coefficients, and $P_n$ the Legendre polynomial of degree $n$. Consequently, the Hamiltonian of the orbiter may be split into the sum of the Kepler problem $\mathcal{H}_{\mathrm{K}}$ and the zonal Moon potential $\mathcal{H}_{\mathrm{Z}}$

$$\begin{aligned}\mathcal{H}_{\mathrm{K}} &= \frac{1}{2}\boldsymbol{X}\cdot\boldsymbol{X} - \frac{\mu}{r} = \frac{1}{2}\left(R^2 + \frac{\Theta^2}{r^2}\right) - \frac{\mu}{r},\\ \mathcal{H}_{\mathrm{Z}} &= \sum_{n\geq 2}\frac{\mu}{r}\left(\frac{\alpha}{r}\right)^n J_n P_n(\sin i\sin(f+g)),\end{aligned}$$

where $(r,\theta,\nu,R,\Theta,N)$ are the polar-nodal variables, $i$ the inclination, $f$ the true anomaly and $g$ the argument of the periapse.

We take the Hamiltonian as

$$\mathcal{H}_l = \mathcal{H}_0 + \epsilon\mathcal{H}_1 \tag{1}$$

with $\mathcal{H}_0 = \mathcal{H}_{\mathrm{K}}$, $\epsilon\mathcal{H}_1 = \mathcal{H}_{\mathrm{Z}}$, and the small parameter $\epsilon = J_2$. Note that we dropped the Coriolis term since the nodal angle $\nu$ does not appear in the Hamiltonian.

## 3 Lie Transforms for Differential Systems

Let us give here a short description of the general algorithms of the Lie transforms; for more details the reader is referred to the original work of Deprit [4], or to the tutorial [10].

A Lie transform of generator $W$ is a near-identity canonical transformation $\varphi : (\boldsymbol{y},\boldsymbol{Y};\epsilon) \longrightarrow (\boldsymbol{x},\boldsymbol{X})$, such that $\boldsymbol{x}(\boldsymbol{y},\boldsymbol{Y};\epsilon)$ and $\boldsymbol{X}(\boldsymbol{y},\boldsymbol{Y};\epsilon)$ verify

$$\frac{d\boldsymbol{x}}{d\epsilon} = \frac{\partial W}{\partial \boldsymbol{X}}, \qquad \frac{d\boldsymbol{X}}{d\epsilon} = -\frac{\partial W}{\partial \boldsymbol{x}},$$

with initial conditions $\boldsymbol{x}(\boldsymbol{y},\boldsymbol{Y};0)=\boldsymbol{y}$ and $\boldsymbol{X}(\boldsymbol{y},\boldsymbol{Y};0)=\boldsymbol{X}$.

Let us now consider a Hamiltonian $\mathcal{H}$ that is a power series of the small parameter $\epsilon$

$$\mathcal{H}(\boldsymbol{x},\boldsymbol{X};\epsilon)=\sum_{n\geq 0}\frac{\epsilon^n}{n!}\mathcal{H}_n(\boldsymbol{x},\boldsymbol{X}), \tag{2}$$

and a Lie transform whose generating function is the series

$$W(\boldsymbol{x},\boldsymbol{X};\epsilon)=\sum_{n\geq 0}\frac{\epsilon^n}{n!}W_{n+1}(\boldsymbol{x},\boldsymbol{X}).$$

We may ask ourselves how this Hamiltonian is affected by the Lie transform.

By putting $\mathcal{H}_{n,0}=\mathcal{H}_n$, $\forall n\geq 0$, the following recurrence relation holds

$$\mathcal{H}_{i,j}=\mathcal{H}_{i_1,j-1}+\sum_{0\leq k\leq i}\binom{i}{j}\{\mathcal{H}_{k,j-1},W_{i+1-k}\}, \tag{3}$$

for $i\geq 0$ and $j\geq 1$, and where $\{-,-\}$ stands for the Poisson bracket.

This recursive algorithm can be performed by means of the so-called Lie triangle (cfr. Deprit [4]) which simplifies the task of automating the method.

The new Hamiltonian, denoted by $\mathcal{K}$, reads

$$\mathcal{K}(\boldsymbol{y},\boldsymbol{Y};\epsilon)=\sum_{n\geq 0}\frac{\epsilon^n}{n!}\mathcal{K}_n(\boldsymbol{y},\boldsymbol{Y})=\sum_{n\geq 0}\frac{\epsilon^n}{n!}\mathcal{H}_{0,n}(\boldsymbol{y},\boldsymbol{Y}).$$

Usually, the generating function is not known, and it must be determined term by term by means of Eq. (3) in order for the new Hamiltonian to satisfy some predetermined conditions or requirements. The generating function is thus obtained by solving the PDE (3) which may be put in the form of the so-called *homology equation*

$$\mathcal{L}_0(W_n)+\mathcal{K}_n=\tilde{\mathcal{H}}_{0,n}, \tag{4}$$

where $\tilde{\mathcal{H}}_{0,n}$ collects all the terms known from the previous order, and $\mathcal{L}_0(-)$ is the Lie derivative operator, i.e., $\mathcal{L}_0(-)=\{-,\mathcal{H}_0\}$.

Once the generating function $W$ is obtained, it is possible to get both the direct and inverse transformations. For details, see the original work of Deprit [4].

Now, it is necessary to choose the properties the new Hamiltonian must satisfy. The main goal of the analytical theories for artificial satellites is to reduce the original Hamiltonian by means of Lie transformations and in such a way that the new Hamiltonian would be simple enough.

As is seen from Eq. (4), the term $\mathcal{H}_0$ plays a very important *rôle*, since it is necessary to compute the Lie derivative $\mathcal{L}_0$ with respect to it. Thus, the simpler the expression of $\mathcal{H}_0$, the easier it is to compute the Lie derivative.

In the Delaunay map ($\ell$, $g$, $h$, $L$, $G$, $H$), the Keplerian Hamiltonian reads as

$$\mathcal{H}_0=-\frac{\mu^2}{2L^2} \tag{5}$$

The Lie derivative $\mathcal{L}_0 : F \mapsto \{F; \mathcal{H}_0\}$ in the vector field generated by $\mathcal{H}_0$ is very simple indeed: it is the differential operator

$$\mathcal{L}_0 = n\frac{\partial}{\partial \ell} \quad \text{with} \quad n = \frac{\mu^2}{L^3}, \tag{6}$$

and the PDE to be solved, at each order $n$ of the transformation, according to (4) is

$$n\frac{\partial W_n}{\partial \ell} + \mathcal{K}_n = \tilde{H}_{0,n}.$$

The perturbation $\mathcal{R}$ in Delaunay variables is

$$\mathcal{R} \equiv \mathcal{R}(\ell, g, h, L, G, H; \varepsilon) = \sum_{n \geq 1} \frac{\varepsilon^n}{n!} \mathcal{R}_n(\ell, g, h, L, G, H), \tag{7}$$

that is, a power series of the small parameter $\varepsilon$.

It still remains to fix what we want to obtain after a Lie transformation. Ideally, we would like to have a new Hamiltonian independent of some variables, which is known as a *normalization*, but this goal is not always possible and usually it requires expansions in terms of the eccentricity (that is, valid only for almost circular orbits) or a great deal of symbolic computation. Deprit [5] introduced other transformation *simplifications* that do not eliminate the variables, but convert the original Hamiltonian into a simpler Hamiltonian, and to which normalization techniques may be applied; by so doing, the amount of computation is drastically reduced.

## 3.1 Delaunay Normalization

Essentially, a *normalization* consists of removing all the coordinates from the Hamiltonian by means of Lie transforms.

The kernel of the Lie derivative along the Hamiltonian flow of the pure Kepler problem is

$$\ker(\mathcal{L}_0) = \Big\{ F(\ell, g, h, L, G, H) \quad | \quad \mathcal{L}_0(F) = 0 \Big\}.$$

Consequently, as the Delaunay action $L$ is an integral of $\mathcal{H}_0$, it generates an infinitesimal contact transformation [6].

The purpose in that local map is to replace $\mathcal{R}$ by $\mathcal{R}'$ such that $\mathcal{R}' \in \ker(\mathcal{L}_0)$. Then it appears that the symmetrization induced by the new Delaunay action $L'$ is obtained by eliminating $\ell'$ from the transformed Hamiltonian $\mathcal{H}'$. As a side effect, the normalization of (1) allows reducing the number of degrees of freedom by one unit $\mathcal{H} \equiv \mathcal{H}(g', h', G', H')$.

## 3.2 Simplifications. Elimination of parallax.

With a *simplification*, the coordinates from the Hamiltonian are not removed, but the Hamiltonian is reduced to a simpler one.

Given a vector space $\mathcal{A}$ of functions defined over the phase space ($\ell$, $g$, $h$, $L$, $G$, $H$), we say that it is a *Poisson algebra* if, for any $p$, $q \in \mathcal{A}$ the Poisson bracket $\{p;q\} \in \mathcal{A}$.

If we assume that: i) $\mathcal{A}$ is a Poisson algebra of functions over the phase space; ii) $\mathcal{B}$ is a proper vector subspace of $\mathcal{A}$; and iii) $A \in \mathcal{A}$ with $A \notin \mathcal{B}$, then a canonical transformation

$$\varphi : (\ell', g', h', L', G', H') \mapsto (\ell, g, h, L, G, H)$$

is a *simplification* of $A$ if $\varphi(A) \in \mathcal{B}$ (see e.g. [7, 8]).

We emphasize that although a normalization is always a simplification, the converse is not true. While normalizing a perturbed Hamiltonian, in the sense of Delaunay, implies removing the variable $\ell$ from the perturbation and making the transformed system belong to the kernel of $\mathcal{L}_0$, simplifying a Hamiltonian does not necessarily imply reducing the number of degrees of freedom.

The elimination of the parallax invented by Deprit [5] is the paradigm of *simplification*. We borrow from the abstract: *"When the perturbation affecting a Keplerian motion is proportional to $r^{-n}$ $(n \geq 3)$, a canonical transformation of Lie type will convert the system into one in which the perturbation is proportional to $r^{-2}$. Because it removes parallactic factors, the transformation is called the elimination of the parallax."*

The main advantage in performing this transformation is that it drastically reduces computations needed for performing further normalizations.

In Whittaker (polar-nodal) variables, the Keplerian Hamiltonian becomes

$$\mathcal{H}_0 = \frac{1}{2}\left(R^2 + \frac{\Theta^2}{r^2}\right) - \frac{\mu}{r}.$$

The state functions

$$\begin{aligned} p &\equiv p(\Theta) = \Theta^2/\mu, \\ C &\equiv C(r,\theta,R,\Theta) = \left(\frac{p}{r} - 1\right)\cos\theta + \frac{p}{\Theta} R \sin\theta, \\ S &\equiv S(r,\theta,R,\Theta) = \left(\frac{p}{r} - 1\right)\sin\theta - \frac{p}{\Theta} R \cos\theta \end{aligned} \tag{8}$$

play a prominent role in the course of eliminating parallax.

In Delaunay variables, these functions become

$$C \equiv C(g,L,G) = e\cos g, \quad S \equiv S(g,L,G) = e\sin g, \tag{9}$$

and because neither $C$ nor $S$ depend on the mean anomaly, it follows that $\mathcal{L}_0(C) = 0$, $\mathcal{L}_0(S) = 0$ and $\mathcal{L}_0(p) = 0$,

The key to the elimination of parallax is the property of functions of the family

$$\mathcal{F} \equiv \left\{ F = \sum_{j\geq 0} \left(C_j \cos j\theta + S_j \sin j\theta\right) \,\middle|\, C_j, S_j \in \ker(\mathcal{L}_0) \right\}. \tag{10}$$

Indeed, it can be proven that $\forall F \in \mathcal{F}$,

$$\mathcal{L}_0 \left[ \sum_{j\geq 1} \frac{1}{j} \left( C_j \sin j\theta - S_j \cos j\theta \right) \right] + \frac{\Theta}{r^2} C_0 = \frac{\Theta}{r^2} F.$$

We may identify this relation with the homology equation (4) by putting

$$F = \frac{r^2}{\Theta} \tilde{\mathcal{H}}_{n,0}, \quad \mathcal{H}_{0,n} = \frac{\Theta}{r^2} C_0, \quad \mathcal{W}_n = \sum_{j\geq 1} \frac{1}{j} \left( C_j \sin j\theta - S_j \cos j\theta \right).$$

Thus, it is sufficient to start with the function $r^2 \tilde{\mathcal{H}}_n/\Theta$, and express it as a function of the family (10) by replacing $1/r$ and $R$ with

$$\frac{1}{r} = \frac{1}{p}(1 + C\cos\theta + S\sin\theta), \quad R = \frac{\Theta}{p}(C\sin\theta - S\cos\theta), \tag{11}$$

and then, select from this expression $C_0$, i.e. those terms which do not depend on $\theta$. With this, and after a suitable rearranging of terms, we will build up the generating function $W_n$.

# 4 Analytical integration

As explained in the above section, once the elimination of parallax is performed, the new Lie transforms are easier to implement. Thus, our first step consists of applying the elimination of parallax to the Hamiltonian. After the elimination of parallax, the new Hamiltonian, when expressed in the Delaunay chart, depends on the mean anomaly $\ell$, and hence, our second step is to make a Delaunay normalization. Note that the zero order Hamiltonian is the Kepler problem, $H_0 = \mathcal{H}_K$. The Lie derivative is the one given in (6), and the new Hamiltonian chosen is the averaged Hamiltonian over the mean anomaly $\ell$,

$$\mathcal{K}_n = \frac{1}{2\pi} \int_0^{2\pi} \widetilde{\mathcal{H}}_{0,n}\, d\ell; \tag{12}$$

hence, the homology equation gives

$$W_n = \frac{1}{n} \int \left( \widetilde{\mathcal{H}}_{0,n} - \mathcal{K}_n \right) d\ell. \tag{13}$$

The Delaunay normalization reduces by 1 the degrees of freedom since it makes the angle $\ell$ cyclic. Hence, its conjugate action $L$ is an integral. After the Dealunay normalization, the Hamiltonian corresponding to low orbit has only one degree of freedom; it contains only one angle, the pericenter $g$, and we stop the process of applying more Lie transforms to it. In short, after several transformations and an adequate rearranging of terms, we arrive at the 1-DOF Hamiltonian $\mathcal{K}_Z = \mathcal{K}(-, g, -; L, G, H)$.

A detailed analysis of the qualitative behavior of this system for several models of the Moon potential shows that the influence of harmonics $J_2$ and $J_7$ is dominant with respect to other harmonics. This agrees with previous results obtained numerically in [11], [15] and [13]. Hence, in order to have a simpler qualitative analysis, from now on we shall consider only these two harmonics in our study. Thus, we will deal with the following Hamiltonian

$$\begin{aligned}\mathcal{K}_{\mathrm{Z}} = \quad & \frac{\mu J_2 \alpha^2}{4}\frac{3s^2-2}{a^3\eta^3} \\ & + \frac{3\mu J_7 \alpha^7}{16384}\frac{es}{a^8\eta^{13}}\Big[231e^4s^4(13s^2-12)\sin 5g \\ & -105e^2s^2(3e^2+8)(143s^4-220s^2+80)\sin 3g \\ & +70(5e^4+20e^2+8)(429s^6-792s^4+432s^2-64)\sin g\Big].\end{aligned} \tag{14}$$

As usual, $c = \cos i$, $s = \sin i$ and $\eta^2 = 1 - e^2$.

The equations of motion of the Hamiltonian (14) are

$$\begin{aligned}\frac{dg}{dt} &= m_{00} + m_{01}\sin g + m_{03}\sin 3g + m_{04}\sin 5g, \\ \frac{dG}{dt} &= m_{10}\cos g + m_{12}\cos 3g + m_{13}\cos 5g,\end{aligned} \tag{15}$$

where $m_{i,j}$ are coefficients depending on $a$, $e$, $i$, $\mu$, $\alpha$, $\omega$, $J_2$ and $J_7$, but once the fundamental constants are fixed, the coefficients only depend on the three orbital elements, $a$, $e$ and $i$.

# 5 Analysis for low-altitude orbits

Frozen orbits correspond to equilibria of a system. In the case of low-altitude orbits, the system under consideration is given by Equations (15). Thus, we have to make zero the right hand part of these equations and solve the corresponding system

$$\begin{aligned}& m_{00} + m_{01}\sin g + m_{03}\sin 3g + m_{04}\sin 5g = 0, \\ & m_{10}\cos g + m_{12}\cos 3g + m_{13}\cos 5g = 0,\end{aligned} \tag{16}$$

The second equation holds when $\cos g = 0$, that is, for $g = \pi/2, 3\pi/2$; thus, by replacing these values in the first equation of (16), there results one equation depending on three variables that represents a 3-D surface in the variables $i$, $e$ and $a$, whose graphics appear in Figure 1. Points on this surface correspond to equilibria of the system (16), that is, to frozen orbits.

In this figure we see that for sections $a =$ constant, we obtain very similar curves; thus, to explore details more deeply, we choose a semi-major axis $a = \alpha + 100$ km, which indeed corresponds to low orbits. For this value of $a$, we plot the curve $i$ versus $e$ such as it appears in Figure 2.

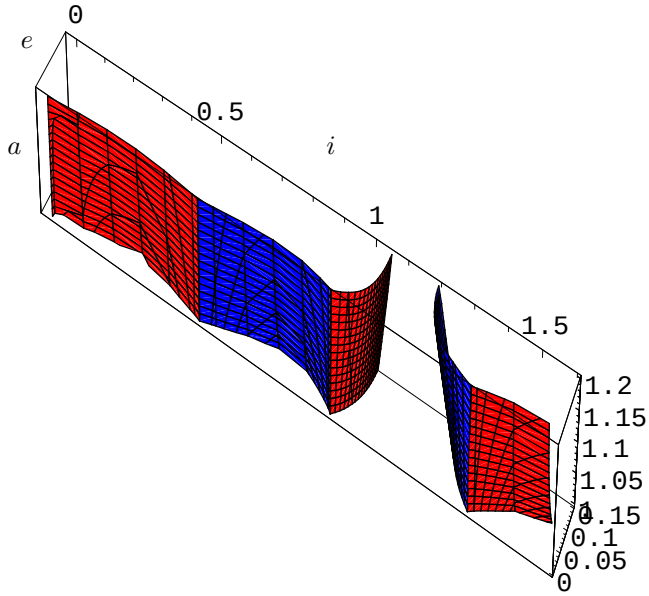



Figure 1: Surface of frozen orbits for values of $i$, $e$ and $a$.

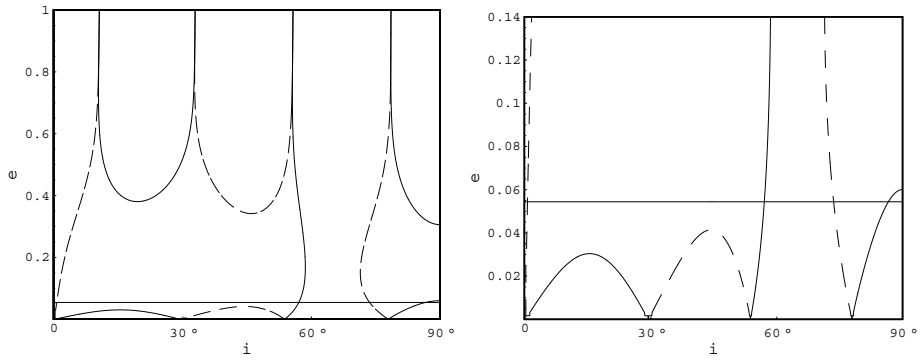



Figure 2: Frozen orbits for $a = \alpha$+100 km. The horizontal line $e = 0.054407$ is the limit for impact orbits. Right plot is a part of the left one, showing realistic orbits.

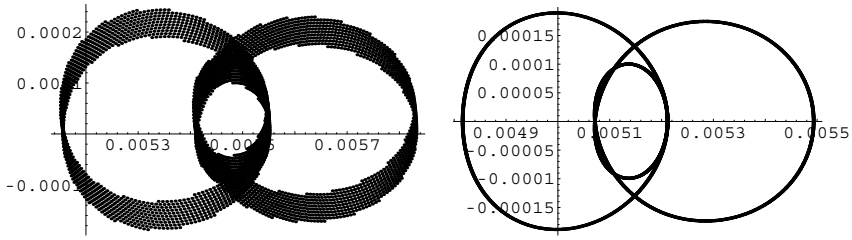



Figure 3: Evolution after 20 days of a frozen orbit $(e\sin\omega, e\cos\omega)$ of the original Hamiltonian. Initial conditions for the left are averaged elements. Initial conditions for the right are osculating elements, obtained by the inverse of the Lie transforms.

The horizontal line at $e = 0.054407$ is the limit for impact orbits. The higher the eccentricity, the shorter the distance at the pericenter, and for eccentricities bigger than $e = 0.054407$ the orbiter would impact the moon surface. Both graphics in Fig. 2 represent the same section, but the one on the right shows fewer values of the eccentricity. The solid line corresponds to $g = \pi/2$, while the dashed line corresponds to $g = 3\pi/2$.

When comparing this map of frozen orbits with the one given by Elipe and Lara [11, Fig. 1], we see that they are almost identical; the only exception is that in Fig. 2 we detect a new family of almost equatorial frozen orbits, - represented as the line very close to the vertical axis, - that was not detected in the quoted paper.

Note that these frozen orbits correspond to the averaged Hamiltonian. As an illustration on how these orbits evolve in the original Hamiltonian, we take from the graphics one a set of initial conditions of a frozen orbit. Then, we implement the initial conditions in an orbit propagator with the initial system and we plot the evolution of the orbit for 200 revolutions (almost 20 days) in the left part of Figure 3. We clearly see that the frozen character degenerates for long time integration.

However, since we obtained the averaged Hamiltonian through Lie transforms, we have the generating functions, and by means of the inverse transformation we convert the averaged initial conditions into osculating elements. Then, we again compute the orbit with the propagator for the same time intervals as in the above case as is shown in the right part of Figure 3. For this new set of initial conditions, the orbit remains very close to being periodic.

It still remains to analyze the case $\cos g \neq 0$ in the system (15). This case will appear in a forthcoming work.

**Acknowledgments.** Supported by the Spanish Ministry of Science and Technology (Projects # ESP2005-07107 and # MTM2006-06961).

# References

[1] Coffey, S.L., Deprit, A., Miller, B., "The critical inclination of the artificial satellite theory," *Celestial Mechanics*, **39**, 1986, 365–406.

[2] Coffey, S.L., Deprit, A., Deprit, E., and Healy, L., "Painting the phase space portrait of an integrable dynamical system," *Science*, **247**, 1990, 833–836.

[3] Coffey, S.L., Deprit, A., Deprit, E., "Frozen Orbits for satellites close to an Earth-like planet," *Celestial Mechanics and Dynamical Astronomy*, **59,** 1994, 37–72.

[4] Deprit, A., "Canonical transformations depending on a small parameter," *Celestial Mechanics*, **1**, 1969, 12-30.

[5] Deprit, A., "The Elimination of the Parallax in Satellite Theory," *Celestial Mechanics*, **24**, 1981, 111–153.

[6] Deprit, A., "Delaunay Normalisations," *Celestial Mechanics*, **26**, 1982, 9–21.

[7] Deprit, A. Ferrer, S., "Simplifications in the theory of artificial satellites," *Journal of the Astronautical Sciences*, **37**, 1989. 451–463.

[8] Deprit, A., Miller, B., "Simplify or perish," *Celestial Mechanics and Dynamical Astronomy*, **45**, 1989. 189–200.

[9] Folta, D. and Quinn, D., "Lunar frozen orbits," *2006 AIAA/AAS Astrodynamics Specialist Conference*, Keystone, CO.

[10] Meyer, K. R. and Hall, G. R., *Introduction to Hamiltonian Dynamical Systems and the N–Body Problem,* in "Applied Mathematical Sciences," 1992, Vol. 90, Springer–Verlag, New York.

[11] Elipe, A., and Lara, M., "Frozen Orbits About the Moon," *Journal of Guidance, Control and Dynamics*, **26**, No. 2, 2003. 238–243.

[12] M.E., Paskowitz and Scheeres, D., "Design of Science Orbits About Planetary Satellites: Application to Europa," *Journal of Guidance, Control and Dynamics*, **29**, No. 5, 2006. 1147-1158.

[13] Ramanan, R.V. and Adimurthy, V., "An analysis of Near-Circular Lunar Mapping Orbits", *J. Earth Syst. Sci.*, **114**, 2005. 619–626.

[14] Russell, R.P. and Lara, M., "Repeat Ground Track Lunar Orbits in the Full-Potential Plus Third-Body Problem", *AIAA 2006-6750*, 2006.

[15] San Juan, J.F., "Lunar orbiter: The influence of $J_7$", *Actas IX Jornadas de Mecánica Espacial*, Jaca, Spain. 2006.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# On the Distance Between Two Quasi-Keplerian Orbits

Luis Floría

*University of Zaragoza. Space Mechanics Group.*
*Faculty of Sciences (Mathematics). E– 50 009 Zaragoza, Spain.*
lfloria@unizar.es

**Abstract.** A class of perturbed Keplerian systems (the so-called *generalized quasi-Keplerian systems*) is considered, and the problem of finding the distance (in the sense of Set Theory) between two generalized quasi-Keplerian orbits (with a common focus) is formulated in terms of *universal functions*, irrespective of the specific kind of perturbed conic-section orbits at issue.
**Key words and expressions:** Perturbed Keplerian systems, (perturbed) conic-section orbits, distance function, uniform treatment of two-body motion, universal functions, universal orbital variables and parameters.

## 1 Introduction

Within the framework of Hamiltonian Celestial Mechanics, Deprit ([4], §4) considered certain reducing canonical transformations (*quasi-Delaunay transformations* and *torsions*, according to his own terminology) for the contraction of the simplest model of *quasi-Keplerian systems* onto Keplerian Hamiltonians. Deprit himself (*ibid.* §7, §9) resorted to torsion-type transformations to reduce and solve the *generalized* quasi-Keplerian Hamiltonian characterizing his first-order radial intermediary for the Main Problem in Artificial Satellite Theory. A similar treatment can be applied to any member of the related class of conservative generalized quasi-Keplerian systems defining his chain of radial intermediaries (integrable approximations for the said non-integrable Main Oblateness Perturbation Problem), in which the perturbation terms are proportional to $r^{-2}$.

A more general class of perturbed Keplerian Hamiltonians (see Floría [5]), giving rise to our concept of *generalized quasi-Keplerian systems*, was reduced and solved under the assumption of perturbed elliptic-type orbital motion.

Although most of Deprit's analysis seems to be restricted to the consideration of bound orbits within the attraction field of quasi-Keplerian potentials, some of his developments can be extended to the case of *any kind of (perturbed) orbital motion under the said quasi-Keplerian models.*

Our approach is based on an extended phase-space formulation and treatment of our *generalized* quasi-Keplerian systems (namely, those previously studied in Floría [5]), whose canonical reduction and solution can be achieved with

the help of appropriate generalized TR-mappings (Deprit, [3]), in terms of the corresponding sets of generalized canonical orbital TR-elements that are *uniformly applicable to any kind of (not necessarily bound) two-body orbital motion.*

Baluyev and Kholshevnikov [1] have investigated in detail the problem of finding the (minimal) distances between two *arbitrary* Keplerian orbits that share a common center of force (focus), taking into account every possible combination of *ordered pairs of confocal conic-section solutions* to the unperturbed gravitational two-body problem.

Thus, with the closest approach between two orbiting particles in mind, they established a normalized, adimensional, squared Keplerian distance function and a set of two simultaneous equations corresponding to the *necessary conditions* for its *critical points* (relative extrema), for all types of confocal Keplerian solutions. In each case of interest, they parameterized the orbital motion with an adequate anomaly-like variable. With appropriate notations, their discussion is performed in terms of combinations of (trigonometric, hyperbolic and/or algebraic) "polynomial-like" expressions involving the angular variables parameterizing the two paths.

The present paper is devoted to the problem of the minimum separation distance between pairs of point masses moving along their respective *generalized quasi-Keplerian orbits about a common center of force.* The aforesaid (see [1]) exhaustive taxonomy and treatment of the diverse combinations of cases is *systematized and reformulated* into a compact, unified approach, in terms of *universal functions* and *universal parameters* (see, e.g., Stiefel and Scheifele [7], §11; Battin [2], §4.5 - §4.7). The squared distance function between two perturbed Keplerian-like orbits with a "common focus", the two bivariate components of its gradient, and the subsequent discussion of stationarity conditions can be translated into a *universal formulation*; the parametric equations of the orbits are given by using the respective *universal eccentric-like anomaly*, introduced through Stumpff's generalization of Sundman's differential transformation of the time variable. Such pseudo-times are involved in the argument of universal functions occurring in the solutions.

## 2 On the Concept of Quasi-Keplerian Systems

*Keplerian systems* are governed by the Newtonian Gravitation-Law Potential

$$V_\kappa(r) \;=\; -K/r\,, \quad K \text{ being a positive constant.}$$

In turn, *quasi-Keplerian* systems (Deprit [4], §4, §7, §9) are one-degree-of-freedom perturbed Keplerian systems governed by potentials of the type

$$\mathcal{Q}_2(r) \;=\; V_\kappa(r) + \mathcal{V}_2(r) \;=\; -K/r \;+\; \mathcal{D}/r^2 \;\;,$$

where $\mathcal{D}$ might be either an absolute constant (e.g., Manev's model) or some function of conserved quantities (e.g., Deprit-type radial intermediaries in Artificial Satellite Theory). As an extension of this idea, *generalized quasi-Keplerian*

*systems* (Floría [5], §2) are one-degree-of-freedom perturbed Keplerian systems with perturbing potentials containing a linear combinations of terms proportional to $r^{-j}$ , $j = 0, 1, 2$:

$$\mathcal{V}(r) = V_\kappa(r) + \mathcal{V}_0(r) + \mathcal{V}_1(r) + \mathcal{V}_2(r) \ , \quad \mathcal{V}_j(r) = \mathcal{J}/r^j \ ,$$

each coefficient $\mathcal{J}$ being an absolute constant or some function of conserved quantities. These potentials are *compatible* with a basic geometrical and dynamical Keplerian-like pattern.

# 3 Distance Between Two Keplerian Orbits

In this Section, we closely follow and summarize the considerations of Baluyev and Kholshevnikov in [1]. The idea of *distance between two orbits* is taken in the sense of Set Theory: minimal value of distances between two points belonging to two given curves, $\mathcal{C}\,, \mathcal{C}\,' \subset R^3$, with respective parametric representations $\mathcal{C} \equiv \vec{r} = \vec{r}\,(u)$ and $\mathcal{C}\,' \equiv \vec{r}\,' = \vec{r}\,'\,(u')$. Accordingly, one is interested in *critical points* of the function

$$d^2\,(u, u') \ = \ ||\,\vec{r} - \vec{r}\,'\,||^2 \ .$$

Such critical points must satisfy the *necessary conditions* for extrema:

$$\nabla_{(u,u')}\, d^2 \ = \ \left(\partial\, d^2/\partial u\,,\, \partial\, d^2/\partial u'\right) \,=\, (\,0\,,\,0\,) \ .$$

In particular, let $\mathcal{C} \subset R^3$ be an *arbitrary* (conic-section) *solution* to the unperturbed *Kepler problem* governed by the vector differential equation

$$\frac{d^2\,\vec{r}}{d\,t^2} \,+\, \frac{\mu}{r^3}\,\vec{r} \ = \ \vec{0}\,, \tag{1}$$

where $\mu$ stands for the gravitational coupling parameter of this Keplerian system. For future use, let $(\,a, q, p, e, I, \omega, \Omega\,)$ be a set of Keplerian orbital elements of $\mathcal{C}$, and $h_k$ its Keplerian energy. The *metric elements* $q$ (distance of the pericenter) and $p$ (semi-latus rectum) are given by

$$-\,h_k = \mu\,(\,1 \,-\, e\,)/(\,2\,q\,) \ , \quad p \,=\, q\,(\,1 \,+\, e\,) \ .$$

In terms of an *adequate* eccentric-like anomaly, the position along the (planar) curve $\mathcal{C}$ is expressed with the help of elementary circular or hyperbolic functions:
• Elliptic orbit:

$$\begin{aligned} \vec{r} \ &= \ \vec{r}\,(\,u\,) \,=\, a\,(\cos u \,-\, e\,)\,\vec{P} \,+\, \left(a\,\sqrt{1 \,-\, e^{\,2}}\,\sin u\right)\,\vec{Q}\,, \\ \vec{S} \ &= \ \sqrt{1 \,-\, e^{\,2}}\,\vec{Q}\,. \end{aligned}$$

• Hyperbolic orbit:

$$\begin{aligned} \vec{r} \ &= \ \vec{r}\,(\,u\,) \,=\, a\,(\,e \,-\, \cosh u\,)\,\vec{P} \,-\, \left(a\,\sqrt{e^{\,2} \,-\, 1}\,\sinh u\right)\,\vec{Q}\,, \\ \vec{S} \ &= \ \sqrt{e^{\,2} \,-\, 1}\,\vec{Q}\,. \end{aligned}$$

• Parabolic orbit (as a limiting case of an ellipse):

$$\begin{aligned} \vec{r} &= \vec{r}(\sigma) = q\left(1-\sigma^2\right)\vec{P} + q\,(2\sigma)\,\vec{Q}\,, \quad q \equiv \text{ pericentric distance,} \\ \sigma &= \tan\left(\frac{f}{2}\right), \quad f \equiv \text{ true anomaly.} \end{aligned}$$

These expressions can be considered to be particular instances of a *more general situation*, namely, the parametric representation of a planar curve, in terms of a *parameter* $u$, with the help of two *orthogonal* unit vectors $\vec{P}$ and $\vec{Q}$:

$$\vec{r} = \vec{r}(u) = \varphi(u)\,\vec{P} + \psi(u)\,\vec{Q}\,.$$

In cases of interest in Orbital Mechanics, vector $\vec{P}$ points towards the *pericenter* of the orbit $\mathcal{C}$ (which is unambiguously determined, except for circular motion, $e = 0$); once given $\vec{P}$, vector $\vec{Q}$ is uniquely determined, except for rectilinear trajectories ($p = q = 0$, $\psi(u) = 0$). In all cases, components of vectors $\vec{P}$ and $\vec{Q}$ can be expressed in terms of orbital elements $(I, \omega, \Omega)$:

$$\begin{aligned} P_x &= \cos\omega\cos\Omega - \cos I\sin\omega\sin\Omega\,, \\ P_y &= \cos\omega\sin\Omega + \cos I\sin\omega\cos\Omega\,, \\ P_z &= \sin I\sin\omega\,, \\ Q_x &= -\sin\omega\cos\Omega - \cos I\cos\omega\sin\Omega\,, \\ Q_y &= -\sin\omega\sin\Omega + \cos I\cos\omega\cos\Omega\,, \\ Q_z &= \sin I\cos\omega\,. \end{aligned}$$

Let $\mathcal{C}'$ denote *another solution* to the Kepler problem given in Eq. (1):

$$\vec{r}' = \varphi'(u')\,\vec{P}' + \psi'(u')\,\vec{Q}'\,.$$

(Note: Quantities pertaining to curve $\mathcal{C}'$ are marked with a *stroke* ($'$)).
The squared distance function $\delta(u, u') = ||\,\vec{r} - \vec{r}'\,||^2 = r^2 + r'^2 - 2\,\vec{r}\cdot\vec{r}'$:

$$||\,\varphi(u)\vec{P} + \psi(u)\vec{Q} - \varphi'(u')\vec{P}' - \psi'(u')\vec{Q}'\,||^2,$$

$$\begin{aligned} r^2 &= \varphi^2(u) + \psi^2(u)\,, \quad r'^2 = \varphi'^2(u') + \psi'^2(u')\,, \\ \vec{r}\cdot\vec{r}' &= \varphi(u)\varphi'(u')\vec{P}\vec{P}' + \varphi(u)\psi'(u')\vec{P}\vec{Q}' \\ &+ \psi(u)\varphi'(u')\vec{Q}\vec{P}' + \psi(u)\psi'(u')\vec{Q}\vec{Q}'. \end{aligned}$$

And critical points of $\delta$ satisfy the system of equations

$$\partial\,\delta(u, u')/\partial u = \partial\,\delta(u, u')/\partial u' = 0\,.$$

A thorough analysis of this function $\delta$ and its critical points for *all types of Keplerian solutions* is performed by Baluyev and Kholshevnikov in [1]. For arbitrary pairs of Keplerian conic-section solutions to the two-body problem, they

consider all possible *combinations of ordered pairs* of orbits about a *common center of force*. There are *nine* main cases, $\{\,\mathcal{K}_{j\,k}\,;\;j\,,\,k\,=\,1\,,\,2\,,\,3\,\}$, corresponding to the three main types of orbits (ellipse, hyperbola, parabola):

$$1 \;\longrightarrow\; \text{ellipse}\,, \quad 2 \longrightarrow \text{hyperbola}\,, \quad 3 \longrightarrow \text{parabola}\,.$$

Notice that the ellipse-hyperbola $\mathcal{K}_{1\,2}$ and hyperbola-ellipse $\mathcal{K}_{2\,1}$ cases are *not equivalent*, since the analysis involves *elimination* of the variable $u'$ parameterizing the position along the *second* curve. For practical applications, the four cases $\mathcal{K}_{1\,1}$ , $\mathcal{K}_{1\,2}$ , $\mathcal{K}_{2\,1}$ , and $\mathcal{K}_{2\,2}$ are the most important.

# 4 On Universal Functions in Orbital Mechanics

Some members of certain families of special functions (Stiefel and Scheifele [7], §11; Battin [2], §4.5 - §4.7) can be used for a *compact* representation and treatment of analytical solutions of the two-body problem.

The *Stumpff $c_n$-functions* solve, under a unified treatment, the second-order linear differential equation with constant coefficients

$$d^2y/ds^2 + \varrho\, y = 0 \quad (\varrho \text{ is a } \textit{real parameter}).$$

For $z = \varrho\, s^2$ , the solution can be given by means of a "linear combination" of $c_0(z)$ and $c_1(z)$ , *irrespective of the sign and value of* $\varrho$ :

$$y(s) = y(0)\, c_0\left(\varrho\, s^2\right) \,+\, y'(0)\, s\, c_1\left(\varrho\, s^2\right).$$

• *Stumpff $c_n$-functions* are defined by means of the power series:

$$c_n(z) \;\;=\;\; \sum_{k=0}^{\infty} (-1)^k\, \frac{z^{\,k}}{(\,2\,k\,+\,n\,)\,!}\,, \quad n\,=\,0,1,2,...$$

and have the following properties:
– The series are absolutely convergent for all complex values of $z$ .
– They are real-valued for real $z$ .
– If $\varrho$ is a real parameter, and $z = \varrho\, s^{\,2}$ , the series is convergent for all $s$ regardless of $\varrho$ .
– They are slowly varying functions.

• *Alternative Universal Functions $U_n$* :

$$U_n\,(s,\varrho) \;\;=\;\; s^{\,n}\, c_n\left(\varrho\, s^2\right) \;=\; \sum_{k=0}^{\infty} (-1)^k\, \frac{\varrho^{\,k}\, s^{\,2\,k+n}}{(\,2\,k\,+\,n\,)\,!}\,, \quad n = 0,1,2,...$$

Some *properties and relations*, satisfied by these functions, are:

$$\begin{aligned}
\grave{U}_0 \;\;&=\;\; d\,U_0/d\,s \;=\; -\,\varrho\, U_1\,; \quad \grave{U}_n \;=\; d\,U_n/d\,s \;=\; U_{n\,-\,1}\,, \quad n \,\in\, N\,; \\
1 \;\;&=\;\; {U_0}^{\,2} \,+\, \varrho\, {U_1}^{\,2}\,, \quad 1 \;=\; U_0 \,+\, \varrho\, U_2\,, \\
{U_1}^{\,2} \;\;&=\;\; U_2\,(1 \,+\, U_0)\,, \quad {U_1}^{\,2} \;=\; 2\,U_2 \,-\, \varrho\, {U_2}^{\,2}\,.
\end{aligned}$$

In many practical applications to the *Kepler problem*, $\varrho$ is related to the orbital energy $h_k$ of the Keplerian systems, and one usually takes:
$\mu \equiv$ gravitational bodycentric parameter (Keplerian coupling parameter),
$e \equiv$ eccentricity of the Keplerian orbit, $q \equiv$ distance of the pericenter,
$L \equiv$ negative of the Keplerian energy, $L = -h_k$,

$$\varrho = 2L\,, \quad L = \frac{\mu\,(1-e)}{2q}\,, \quad p = q\,(1+e)\,, \quad s^n c_n\left(2Ls^2\right) = U_n\,(s,2L)\,.$$

These considerations lead to a *universal formulation of and solution to* the Kepler problem. In terms of $s$ as the argument of universal functions, the two-body problem admits a *compact closed-form analytical solution* in the orbital plane:

$$\begin{aligned}
x &= x_1 = q - \mu s^2 c_2\left(2Ls^2\right) = q - \mu U_2\,(s,2L)\,,\\
y &= x_2 = \sqrt{\mu q\,(1+e)}\,s c_1\left(2Ls^2\right) = \sqrt{\mu q\,(1+e)}\,U_1\,(s,2L)\,,\\
r &= \|\,\vec{r}\,\| = q + \mu e s^2 c_2\left(2Ls^2\right) = q + \mu e U_2\,(s,2L)\,,\\
dr &= \mu e s c_1\left(2Ls^2\right) ds = \mu e U_1\,(s,2L)\,ds\,,\\
dt &= r\,ds \ \text{(Sundman's transformation)},\\
t &= qs + \mu e U_3\,(s,2L)\,, \ \text{Kepler's equation.}
\end{aligned}$$

The fictitious time $s$ (proportional to the classical eccentric anomaly in the cases of elliptic and hyperbolic motion) is a *universal* eccentric-like *anomaly*, given via Stumpff's universal generalization of Sundman's regularizing transformation $dt = r\,ds$. It vanishes at a chosen reference time which usually corresponds to the pericenter: $r = q \longrightarrow s_0 = 0$.

The dependence of $s$ on the physical time $t$ is given by the Kepler equation.

This universal description of motion can be supplemented with an additional representation in polar coordinates. Consider the *polar equation of a conic section* in the orbital plane, in terms of polar coordinates $(r, f)$:

$$r\,(f) = p/\,(\,1 + e\,\cos f\,)\,, \quad p = q\,(\,1 + e\,)\,.$$

The *universal parameters* $s$ and $f$ are related to each other via the expressions

$$\begin{aligned}
\sqrt{r}\,\cos\,(\,f/2\,) &= \sqrt{q}\;U_0\,(\,s/2\,,\,2L\,)\,,\\
\sqrt{r}\,\sin\,(\,f/2\,) &= \sqrt{\mu\,(1+e)}\;U_1\,(\,s/2\,,\,2L\,)\,,\\
r\,\sin f &= \sqrt{\mu\,q\,(1+e)}\;U_1\,(\,s\,,\,2L\,)\,.
\end{aligned}$$

As a consequence (see also notations in Section 3), a *Keplerian orbit* $\mathcal{C}$, solution to Eq. (1), can be represented *in its orbital plane* by means of

$$\vec{r} = \vec{r}\,(\,s\,) = [\,q - \mu\,U_2\,(s,2L)\,]\;\vec{P} + \left[\sqrt{\mu\,q\;(\,1+e\,)}\,U_1\,(s,2L)\right]\vec{Q}\,,$$

$$\vec{S} = \sqrt{1+e}\;\vec{Q}\,, \quad \|\,\vec{P}\,\| = 1\,, \ \|\,\vec{Q}\,\| = 1\,, \quad \vec{P} \perp \vec{Q}\,.$$

Another Keplerian orbit $\mathcal{C}\,' \equiv \vec{r}\,' = \vec{r}\,'(\,s\,')$ would admit an analogous representation in its orbital plane. As a matter of notation, remember the convention in Section 3, and denote

$$U_n \;\equiv\; U_n\,(s, 2L)\;, \qquad U_n\,' \;\equiv\; U_n\,'\,(s\,', 2L\,')\;.$$

# 5 Distance Function in Universal Formulation

In this Section we propose (in terms of universal functions) a *unified approach* to the problem of obtaining the *distance between two arbitrary Keplerian orbits*. Taking advantage of our previous considerations, from the Euclidean distance $||\,\vec{r} - \vec{r}\,'\,||^2 = r^2 + r'^2 - 2\,\vec{r}\,\vec{r}\,'$, one defines the *adimensional distance function*

$$\begin{aligned}
\rho\,(s, s') \;&=\; ||\,\vec{r} - \vec{r}\,'\,||^2\,/(2\,q\,q\,') \;=\; \delta\,(s, s')\;, \qquad \text{with} \\
r^2 \;&=\; [\,q + \mu e U_2\,]^2 \;=\; q^2 + \mu^2 e^2 U_2^2 + 2\mu q e U_2\;, \\
\vec{r}\,\vec{r}\,' \;&=\; (qq' - q\mu' U_2' - \mu q' U_2 + \mu\mu' U_2 U_2')\,PP' \\
&\quad + \left(q\sqrt{\mu' q'}\,U_1' - \mu\sqrt{q'\mu'}\,U_2 U_1'\right) PS' \\
&\quad + (q'\sqrt{\mu q}\,U_1 - \mu'\sqrt{q\mu}\,U_1 U_2')\,P'S \;+\; \sqrt{\mu q \mu' q'}\,U_1 U_1' SS'\;.
\end{aligned}$$

After introducing the notations and abbreviations

$$\begin{aligned}
M \;&=\; \mu/q\;, \qquad \vec{\Sigma} \;=\; \sqrt{M}\;\vec{S}\;, \qquad \alpha \;=\; q/q\,'\;, \\
M' \;&=\; \mu'/q\,'\;, \qquad \vec{\Sigma}\,' \;=\; \sqrt{M'}\;\vec{S}\,'\;, \quad \alpha' \;=\; q'/q\;,
\end{aligned}$$

one has

$$\begin{aligned}
\rho\,(s, s\,') \;&=\; \frac{||\,\vec{r} - \vec{r}\,'\,||^2}{2q\,q\,'} \;=\; \delta\,(s, s\,') \;=\; \left(\frac{\alpha}{2} + \frac{\alpha\,'}{2} - PP'\right) \\
&\quad + \frac{\alpha}{2}\,M^2\,e^2\,{U_2}^2 + \frac{\alpha\,'}{2}\,M'^2\,e'^2\,U_2'^2 \;+\; (\alpha M e + MPP')\,U_2 \\
&\quad + (\alpha' M' e' + M' PP')\,U_2\,' \;-\; MM' PP' U_2 U_2\,' \;-\; \Sigma\Sigma' U_1\,U_1\,' \\
&\quad - P\Sigma' U_1\,' \;+\; MP\Sigma' U_2\,U_1\,' \;-\; P'\Sigma\,U_1 \;+\; M' P'\Sigma\,U_1\,U_2\,'\;.
\end{aligned}$$

Notice that function $\rho$ *involves* functions $U_1\,(s)\,,\, U_2\,(s)\,,\, U_1\,'\,(s\,')\,, U_2\,'\,(s\,')$:

$$\rho \;=\; \text{Funct}\,(\,U_1\,(s)\,,\, U_2\,(s)\,,\, U_1\,'\,(s\,')\,,\, U_2\,'\,(s\,'))\,.$$

Accordingly, function $\partial\rho/\partial s$ *involves* $U_1\,(s)\,,\, \grave{U}_1\,(s) = U_0\,(s)\,,\, U_2\,(s)\,,\, \grave{U}_2\,(s) = U_1\,(s)\,,\, U_1\,'\,(s\,')\,,\, U_2\,'\,(s\,')$. Contributions of $U_0\,(s)$ can be expressed *in terms of* $U_2\,(s)$, since $U_0\,(s) = 1 \,-\, (\,2\,L\,)\,U_2\,(s)$.
An *analogous dependence* is found for $\partial\rho/\partial s\,'$. And function $U_0\,'\,(s\,')$ can be eliminated with the help of the relation $U_0\,'\,(s\,') = 1 \,-\, (\,2\,L\,')\,U_2\,'\,(s\,')$. Subsequent calculations can be reduced to *expressions in terms of* functions

$$U_1\,(s)\,,\, U_2\,(s)\,,\, U_1\,'\,(s\,')\,, U_2\,'\,(s\,')\,.$$

Then, *necessary conditions for extrema*,

$$\nabla_{(s\,,\,s')}\,\rho\,(s,s') \;=\; (\,\partial\,\rho/\partial\,s\;,\;\partial\,\rho/\partial\,s')\;=\;\vec{0}\,,\qquad \text{yield}:$$

$$\begin{aligned}
\partial\,\rho/\partial\,s \;&=\; 0 \;\Longrightarrow\; A\,(s)\,U_1{}' + B\,(s)\,U_2{}' \;=\; C\,(s)\;,\\
\partial\,\rho/\partial\,s' \;&=\; 0 \;\Longrightarrow\; \mathcal{M}\,(s)\,U_1{}' + \mathcal{N}\,(s)\,U_2{}' \;=\; \mathcal{K}\,U_1{}'\,U_2{}' + \widehat{\mathcal{N}}\,(s)\;,
\end{aligned}$$

where the functions depending on the universal variable $s$ read

$$\begin{aligned}
A\,(s) \;&=\; MP\Sigma'U_1 \,-\, \Sigma\,\Sigma'U_0 \;=\; MP\Sigma'U_1 \,+\, (2L)\,\Sigma\,\Sigma'U_2 \,-\, \Sigma\,\Sigma'\,,\\
B\,(s) \;&=\; M'P'\,\Sigma U_0 \,-\, MM'\,PP'\,U_1\\
&=\; M'P'\,\Sigma \,-\, (2L)\,M'P'\,\Sigma U_2 \,-\, MM'\,PP'\,U_1\,,\\
C\,(s) \;&=\; (\,B\,(s)/M'\,) \,-\, \alpha\,M\,e\,U_1\,(\,1 + M\,e\,U_2)\;,\\
\mathcal{M}\,(s) \;&=\; MM'PP'\,U_2 \,-\, M'P'\,\Sigma U_1 \,-\, M'PP' \,-\, \alpha'M'e'\,,\\
\widehat{\mathcal{N}}\,(s) \;&=\; M\,P\,\Sigma'U_2 \,-\, \Sigma\,\Sigma'U_1 \,-\, P\,\Sigma'\,,\quad \mathcal{N}\,(s) = (\,2\,L'\,)\,\widehat{\mathcal{N}}\,(s)\,,\\
\mathcal{K} \;&=\; \alpha'\,M'^{\,2}\,e'^{\,2}\;,\;\text{ independent of }\;s\,.
\end{aligned}$$

All these expressions can be reformulated (cf. Gronchi [6], §3) into a polynomial form by means of the change of variable $s \longrightarrow \tau$ (Battin [2], §4.7):

$$U_1\,(s,2L) \;=\; (\,2\,\tau\,)/\left[1 \,+\, (2\,L)\,\tau^{\,2}\right]\;,\quad U_2\,(s,2L) \;=\; (2\,\tau^{\,2})/\left[1 \,+\, (2\,L)\,\tau^{\,2}\right]\,.$$

Analogous formulae replace functions of $s'$ with functions of $\tau'$.

# 6 Case of Generalized Quasi-Keplerian Orbits

Consider the canonical set (see Deprit [3]; [4], §2; Floría [5], §2) of the *Hill-Whittaker polar-nodal variables* $(\,r\,,\,\theta\,,\,\nu\,;\,p_r\,,\,p_\theta\,,\,p_\nu\,)$, enlarged with the canonically conjugate pair $(\,t\,;\,p_0\,)$, where $t$ is the physical time (as an additional coordinate), and $p_0$ denotes its conjugate momentum (negative of the total energy of the system), to coordinatize the extended, 8-dimensional phase space.

The Hamiltonian of the basic *quasi-Keplerian system* (according to the concept due to Deprit), as a perturbation of a standard Keplerian system, reads

$$\begin{aligned}
\mathcal{Q} \;&\equiv\; \mathcal{Q}\,(\,r\,,\,-\,,\,-\,;\,p_r\,,\,p_\theta\,,\,-\,;\,\alpha\,) \,=\mathcal{H}_k\,(\,r\,,\,-\,,\,-\,;\,p_r\,,\,p_\theta\,,\,-\,) \,-\, \frac{\mu\,\alpha}{2\,r^{\,2}}\\
&=\; \frac{1}{2}\left[\,p_r^2 \,+\, \frac{p_\theta^2}{r^{\,2}}\,\right] \,-\, \frac{\mu}{r} \,-\, \frac{\mu\,\alpha}{2\,r^{\,2}}\;,
\end{aligned}$$

where $\mathcal{H}_k$ stands for the Hamiltonian of a conventional Kepler problem.

The extended phase-space formulation of the homogeneous Hamiltonian of a *generalized quasi-Keplerian system*, in extended polar nodal variables, yields

$$\begin{aligned}
\mathcal{H} \;&\equiv\; \mathcal{H}\,(\,r\,,\,-\,,\,-\,,\,-\,;\,p_r\,,\,p_\theta\,,\,p_\nu\,,\,p_0\,) \;=\; p_0 \,+\, \mathcal{H}_k\,(\,r\,,\,-\,,\,-\,;\,p_r\,,\,p_\theta\,,\,-\,)\\
&\quad + V_0\,(\,p_\theta\,,\,p_\nu\,,\,p_0\,) \,+\, \frac{1}{r}\,V_1\,(\,p_\theta\,,\,p_\nu\,,\,p_0\,) \,+\, \frac{1}{r^{\,2}}\,V_2\,(\,p_\theta\,,\,p_\nu\,,\,p_0\,)\;.
\end{aligned}$$

Certain "radial intermediaries" for the Main Oblateness Perturbation Problem in Artificial Satellite Theory fit into this model (Floría [5]). To solve this Hamiltonian, we perform a *generalized TR-like canonical transformation* (Deprit [3]), from polar nodal variables to new variables $( q_\Phi , q_L , q_G , q_N ; \Phi , L , G , N )$, and a *reparametrizing* transformation $t \longrightarrow \tau^*$ of the independent variable, given by a *generalized Sundman-type* differential relation

$$d t \;=\; \widetilde{f}\, d\tau^* , \quad \widetilde{f} \;=\; \left(2\, r^{\,2}\right) / \left(2\, G - \Phi\right) ,$$

such that function $\mathcal{H}$ is contracted onto a Keplerian-like Hamiltonian $\mathcal{K}(\Phi) = \Phi$, with a simple *canonical solution* in terms of $\tau^*$: $q_\Phi = \tau^* + \text{const.}$, while the remaining new variables are constant. Now, let $Q$ denote the function

$$Q \;\equiv\; Q ( r ; \Phi , L , G , N ) \,=\, \frac{2\,(\mu - V_1)}{r} \,-\, 2\,(L + V_0) \,-\, \frac{\left(\gamma^{\,2} + 2\,V_2\right)}{r^{\,2}} ,$$

where one understands that here $V_j \equiv V_j ( L , G , N )$ and $\gamma = G - \Phi$ are functions of the new momenta. And introduce *subsidiary quantities* $\mu^*$, $q$, $p$, $e$, $\Gamma$, $\Lambda$:

$$\begin{aligned}
\mu^* \;&=\; \mu - V_1 , \quad \Lambda \;=\; L + V_0 , \\
q \;&=\; \frac{\mu^* (1 - e)}{2\,(L + V_0)} \,=\, \frac{\mu^* (1 - e)}{2\,\Lambda} , \\
\Gamma^{\,2} \;&=\; \gamma^{\,2} + 2\,V_2 \,=\, \mu^* q (1 + e) \,=\, \mu^* p , \quad p = q (1 + e) ,
\end{aligned}$$

and *auxiliary anomaly-like universal* variables $f$ and $s$ such that

$$r(f) = p / (1 + e \cos f) ,$$

simulating the polar equation of a conic section in the orbital plane, in terms of polar coordinates $(r, f)$, and

$$\begin{aligned}
\sqrt{r}\, \cos ( f/2 ) \;&=\; \sqrt{q}\; U_0 ( s/2 , 2\Lambda ) , \\
\sqrt{r}\, \sin ( f/2 ) \;&=\; \sqrt{\mu^* (1+e)}\; U_1 ( s/2 , 2\Lambda ) , \\
r \sin f \;&=\; \sqrt{\mu^* q (1+e)}\; U_1 ( s , 2\Lambda ) , \\
r(s) \;&=\; q \,+\, \mu^* e\, U_2 ( s , 2\,\Lambda ) , \quad d r = \mu^* e\, U_1 ( s , 2\,\Lambda )\, d s , \\
d f \;&=\; (\Gamma / r)\; d s .
\end{aligned}$$

With the additional notations

$$I_0 \;=\; q\, s \,+\, \mu^* e\, U_3 ( s , 2\,\Lambda ) , \quad I_1 \,=\, s , \quad I_2 \,=\, f / \Gamma ,$$

a *Keplerian-like solution to* $\mathcal{H}$ in terms of the universal parameters $f$ and $s$ is

$$\begin{aligned}
r \;&=\; p / (1 + e \cos f) \;=\; q + \mu^* e U_2 (s, 2\Lambda) ; \\
p_r \;&=\; \sqrt{Q} \;=\; (\mu^* / p)\, e \sin f \;=\; \mu^* e U_1 (s, 2\Lambda) / r ; \\
p_\theta \;&=\; G, \; p_\nu \,=\, N, \; p_0 \,=\, L \text{ are constant;}
\end{aligned}$$

$$\begin{aligned} q_\Phi &= \gamma\, I_2 = \tau^* + \text{const.}, \\ \theta &= q_G + \frac{\partial V_0}{\partial G}\, I_0 + \frac{\partial V_1}{\partial G}\, I_1 + \left[\gamma + \frac{\partial V_2}{\partial G}\right] I_2, \\ \nu &= q_N + \frac{\partial V_0}{\partial N}\, I_0 + \frac{\partial V_1}{\partial N}\, I_1 + \frac{\partial V_2}{\partial N}\, I_2, \\ t &= q_L + \left[1 + \frac{\partial V_0}{\partial L}\right] I_0 + \frac{\partial V_1}{\partial L}\, I_1 + \frac{\partial V_2}{\partial L}\, I_2. \end{aligned}$$

After this *universal formulation* and *canonical reduction* (in extended phase-space) of the generalized quasi-Keplerian system characterized by $\mathcal{H}$, the preceding considerations concerning the treatment of the distance between two arbitrary (unperturbed) Keplerian orbits can be *translated* into the case of two arbitrary (perturbed) *quasi-Keplerian orbits* $\mathcal{C}$ and $\mathcal{C}'$ *generated by* $\mathcal{H}$. Formally, a simple "change of notations" suffices to readily achieve this goal.

The subsequent investigation of the systems of equations obtained from the conditions for extrema, the algebraic elimination process, the search for positive roots of polynomial equations, etc. will require the adequate use of Computer Algebra systems.

**Acknowledgments.** This research was supported by MCYT (DGI) of Spain, Project MTM2006-06961.

# References

[1] Baluyev, R, V., and Kholshevnikov, K. V. - Distance Between Two Arbitrary Unperturbed Orbits. *Celestial Mechanics & Dynamical Astronomy* Vol. 91 ( 3-4), 2005, 287-300.

[2] Battin, R. H. - *An Introduction to the Mathematics and Methods of Astrodynamics* (§4.5 - §4.7, pp. 174 - 190). AIAA Education Series, American Institute of Aeronautics and Astronautics, New York, 1987.

[3] Deprit, A. - A Note Concerning the TR-Transformation. *Celestial Mechanics* Vol. 23 (4), 1981, 299-305.

[4] Deprit, A. - The Elimination of the Parallax in Satellite Theory. *Celestial Mechanics* Vol. 24 (2), 1981, 111-153.

[5] Floría, L. - Canonical Elements and Keplerian-like Solutions for Intermediary Orbits of Satellites of an Oblate Planet. *Celestial Mechanics & Dynamical Astronomy* Vol. 57 (1-2), 1993, 203-223.

[6] Gronchi, G. F. - On the Stationary Points of the Squared Distance Between Two Ellipses With a Common Focus. *SIAM Journal on Scientific Computing* Vol. 24 (1), 2002, 61-80.

[7] Stiefel, E. L., and Scheifele, G - *Linear and Regular Celestial Mechanics* (§11, pp. 42 - 51). Springer Verlag, Berlin-Heidelberg-New York, 1971.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# Symmetry Analysis of Discrete Dynamical Systems

Vladimir V. Kornyak
*Laboratory of Information Technologies*
*Joint Institute for Nuclear Research*
*141980 Dubna, Russia*
kornyak@jinr.ru

**Abstract.** We outline a C program for symmetry analysis of discrete dynamical systems and lattice models in statistical mechanics. In particular, the program constructs and investigates *phase portraits* of discrete dynamical systems *modulo groups* of their symmetries, searches dynamical systems possessing specific properties, e.g.,*reversibility*, computes microcanonical *partition functions* and searches *phase transitions* in mesoscopic systems.

## 1 Introduction

An appealing feature of symmetry analysis of finite discrete systems is its completeness, in contrast to the case of continuous systems where only a negligible small part of all thinkable symmetries — point and contact Lie, Bäcklund and Lie–Bäcklund, and some sporadic instances of so-called *non-local* symmetries — can be studied. Furthermore, there are many philosophical and physical arguments that discreteness is more fundamental than continuity, which arises only as a logical limit in considering large collections of discrete structures.

Recently [1, 2] we have shown that any collection of discrete points taking values in finite sets possesses some kind of *locality*. More specifically, let us consider a collection $\delta = \{x_1, \ldots, x_N\}$ of $N$ "points". Each $x_i$ takes value in its own set of values $Q_i = \left\{s_i^1, \ldots, s_i^{q_i}\right\}$ or using the standard notation $Q_i = \{0, \ldots, q_i - 1\}$. Adopting $Q^\delta$ as the symbolical notation for the Cartesian product $Q_1 \times \cdots \times Q_N$, we define a *relation* on $\delta$ as an arbitrary subset $R^\delta \subseteq Q^\delta$. Then we define a *consequence* of relation $R^\delta$ as an *arbitrary* superset $S^\delta \supseteq R^\delta$ and a *proper consequence* as a consequence which can be represented in the form $P^\alpha \times Q^{\delta\setminus\alpha}$, where $P^\alpha$ is a *nontrivial* relation (i.e., $P^\alpha \neq Q^\alpha$) on the proper subset $\alpha \subset \delta$. Any relation $R^\delta$ allows a decomposition in terms of its proper consequences and this decomposition imposes naturally a structure of an *abstract simplicial complex* which is one of the mathematical abstractions of locality.

These *discrete relations on abstract simplicial complexes* in special cases correspond to *systems of polynomial equations* (if all points $x_i$ take values in

the same set $Q$ with cardinality being a power of a prime $|Q| = p^k$) and to *cellular automata* (if domain $\delta$ allows decomposition into congruent simplices with the same relation on the simplices and this *local* relation is *functional*). The notion of discrete relations also covers discrete dynamical systems more general than cellular automata. The lattice models in statistical mechanics can be treated in this framework too by considering *ensembles* of discrete relations on abstract simplicial complexes.

In this paper we study the dependence of the behavior of discrete dynamical systems on graphs — one-dimensional simplicial complexes — on symmetries of the graphs. We describe our C program for discrete symmetry analysis and results of its application to cellular automata and mesoscopic lattice models.

## 2 Symmetries of Lattices and Orbits of Functions on Lattices

The space of a discrete dynamical system is a ***lattice*** $L$ represented by a $k$-valent graph $G_L$. By a ***symmetry*** of lattice $L$ we mean the automorphism group $\mathrm{Aut}(G_L)$ of the graph of $L$. In applications one often assumes that the lattice $L$ is embedded in some continuous space. In this case the notion of 'dimension' of lattice makes sense. Note that the same graph can be embedded regularly in different continuous spaces as is clear from Fig. 1. Thus, the group $\mathrm{Aut}(G_L)$ is usually larger than the symmetry group of the lattice placed in a space.

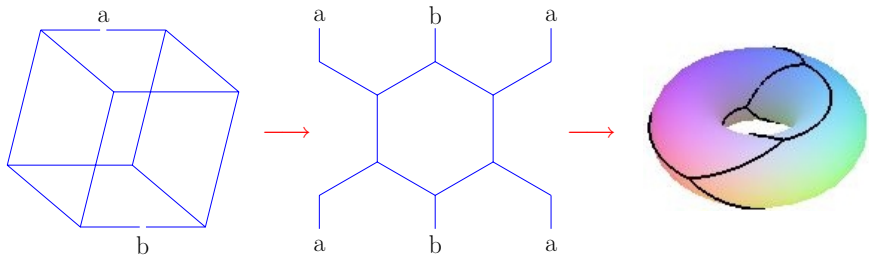



Figure 1: The same graph forms a 4-gonal (6 tetragons) lattice in sphere $\mathbb{S}^2$ and a 6-gonal (4 hexagons) lattice in torus $\mathbb{T}^2$.

The ***automorphism group*** of a graph with $n$ vertices may have up to $n!$ elements. Nevertheless, currently the most efficient algorithm designed by B. McKay [4] determines graph automorphisms by constructing a compact set (no more than $n-1$ elements, but usually much less) of generators of the group.

To study the symmetry properties of a dynamical system on a lattice $L$ we should consider the action of $\mathrm{Aut}(G_L)$ on the space $\Sigma = Q^L$ of ***Q-valued functions*** on $L$, where $Q = \{0, \ldots, q-1\}$ is the set of vertex values. We shall call the elements of $\Sigma$ *states* (or later in Sect. 5 *microstates*). $\mathrm{Aut}(G_L)$ acts non-transitively on the space $\Sigma$ splitting $\Sigma$ into disjoint orbits of different sizes.

Burnside's lemma counts the total number of orbits in $\Sigma$

$$N_{orbits} = \frac{1}{|\mathrm{Aut}(G_L)|} \sum_{g \in \mathrm{Aut}(G_L)} q^{N^g_{cycles}} \ ,$$

where $N^g_{cycles}$ is the number of cycles in a group element $g$.

A large symmetry group allows us to represent the dynamics on the lattice in more compact form. For example, the symmetry group of an icosahedron, dodecahedron and buckyball[1] is $\mathrm{S}_5$, and the information about behavior of any dynamical system on these lattices can be compressed nearly in proportion to $|\mathrm{S}_5| = 120$.

In Tab. 1 we collect some ***quantitative information*** about the lattices shown in Fig.2 and their automorphism groups, namely, *number of vertices* $V(G_L)$, *size of automorphism group* $|\mathrm{Aut}(G_L)|$, total *number of states* $\Omega = |\Sigma| \equiv q^{V(G_L)}$ (here $q = 2$) and *number of group orbits* $N_{orbits}$ in the space of states.

Table 1: Lattices, groups, orbits: quantitative characteristics.

| Lattice | $V(G_L)$ | $\lvert\mathrm{Aut}(G_L)\rvert$ | $\Omega = q^{V(G_L)}$ | $N_{orbits}$ |
|---|---|---|---|---|
| Tetrahedron | 4 | 24 | 16 | 5 |
| Hexahedron | 8 | 48 | 256 | 22 |
| Icosahedron | 12 | 120 | 4096 | 82 |
| Dodecahedron | 20 | 120 | 1048576 | 9436 |
| Graphene 3×4 Torus | 24 | 48 | 16777216 | 355353 |
| Graphene 3×4 Klein bottle | 24 | 16 | 16777216 | 1054756 |
| Triangular 4×6 | 24 | 96 | 16777216 | 180070 |
| Square 5×5 | 25 | 200 | 33554432 | 172112 |
| Buckyball | 60 | 120 | 1152921504606846976 $\approx 10^{18}$ | 9607679885269312 $\approx 10^{16}$ |

Note that the lattices marked in Fig. 2 as "Graphene 3×4", "Triangular 4×6" and "Square 5×5" can be closed by identifications of opposite sides of rectangles in several different ways, in particular, forming embeddings in the torus and in the Klein bottle. Computation shows that the Klein bottle arrangement (as well as others except for embeddings in the torus) leads to *non-transitive* lattices. For example, the Klein bottle arrangement of the hexagonal lattice "Graphene 3×4" has a 16-element symmetry group splitting the set of vertices into two orbits of sizes 8 and 16. Since non-transitivity of points contradicts to our usual concept of space, we shall not consider such non-transitive lattices.

[1]Traditionally, the icosahedral group $\mathrm{I}_h = \mathrm{A}_5$ is considered as the symmetry group for these polyhedra. $\mathrm{A}_5$ is a 60-element discrete subgroup of SO(3). Adding reflections to $\mathrm{A}_5$ we get a twice as large (and hence more efficient for our purposes) group $\mathrm{S}_5$.

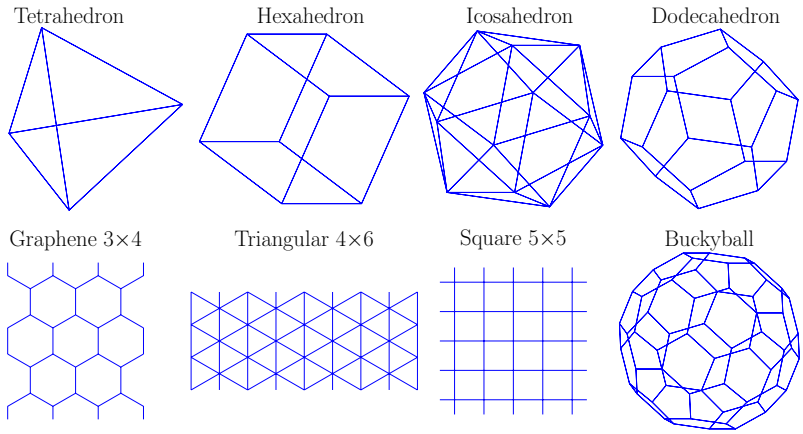



Figure 2: Lattices of Tab. 1.

# 3 Computer Program

A C program exploiting symmetries for study different properties of deterministic and statistical lattice systems takes the following input elements:

- Graph of lattice $G_L = \{N_1, \ldots, N_n\}$. $N_i$ is the neighborhood of the $i$th vertex.
- *Cellular automata branch:*
  A set of local rules $R = \{r_1, \ldots, r_m\}$. $r_i$ is the bit representation of $i$th rule.
- *Statistical models branch:*
  Hamiltonian of the model.
- Some control parameters.

The program computes the automorphism group $\mathrm{Aut}(G_L)$ and

- in the case of cellular automata the program constructs *phase portraits* of automata *modulo* $\mathrm{Aut}(G_L)$ for all rules from $R$.
  Manipulating the above-mentioned control parameters we can
  - select automata with specified properties, for example, *reversibility*;
  - search automata whose phase portraits contain specified structures, for example, limit cycles of a given length, *Gardens of Eden* [5], or, more generally, isolated cycles, etc.
- in the case of a lattice model the program computes the partition function and other characteristics of the system, and searches phase transitions.

***Example of timing.*** The full run of all 136 symmetric 3-valent binary cellular automata on the dodecahedron takes $\approx 40$ sec on a 1133MHz Pentium III PC.

# 4 Deterministic Systems

The splitting of the space $\Sigma$ of functions on a lattice into group orbits of different sizes imposes a ***universal restriction on the behavior of any dynamical system*** whatever particular law governs the evolution of the system. Namely, the dynamical trajectories can obviously be directed only from *larger orbits* to *smaller orbits* or to *orbits of the same size*, and *periodic trajectories* must lie *within the orbits of equal size*. This is an analog of the *Second Law of Thermodynamics*: any isolated system may only lose information in its evolution.

As a specific class of discrete dynamical systems, we consider the 'one-time-step' ***cellular automata*** on $k$-valent lattices with ***local rules symmetric*** with respect to all permutations of $k$ outer vertices of the neighborhood. This symmetry property is an immediate discrete analog of the general local diffeomorphism invariance of fundamental physical theories based on the continuum of space and time — the diffeomorphism group of manifold $M$ is very special subgroup of the infinite symmetric group of $M$, i.e., $\mathrm{Diff}(M) \subset \mathrm{Sym}(M)$.

As we demonstrate in [3], in the binary case ($q = 2$) the automata with symmetric local rules are completely equivalent to the generalized Conway's "Game of Life" automata [5] and, hence, their rules can be formulated in terms of "Birth"/"Survival" lists.

Adopting the convention that the outer points and the root point of the neighborhood are denoted $x_1, \ldots, x_k$ and $x_{k+1}$, respectively, we can write a *local rule* determining one time-step evolution of the root in the form

$$x'_{k+1} = f\left(x_1, \ldots, x_k, x_{k+1}\right). \tag{1}$$

The total number of rules (1) symmetric with respect to permutations of points $x_1, \ldots, x_k$ is equal to $q^{\binom{k+q-1}{q-1}q}$. For the case of our interest ($k = 3$, $q = 2$) this number is 256.

Observing that the rules which can be obtained from each other by permuting $q$ elements of the set $Q$ are equivalent (such a permutation means nothing but a renaming of values), we can reduce the number of rules of interest. This reduced number can be easily counted via *Burnside's lemma* as the number of orbits of rules (1) under the action of the group $\mathrm{S}_q$. The concrete expression depends on the cyclic structure of elements of $\mathrm{S}_q$. For the case $q = 2$ we have the number of non-equivalent rules

$$N_{rules} = 2^{2k+1} + 2^k.$$

Thus, studying 3-valent binary case, we have to consider 136 different rules.

### 4.0.1 Examples of Phase Portraits. Cellular Automaton 86.

As an example, let us consider rule 86 on a tetrahedron, cube and dodecahedron. The number 86 is the "little endian" representation of the bit string 01101010 taken from the last column of the rule table with $\mathrm{S}_3$-symmetric combinations of values for $x_1, x_2, x_3$

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_4'$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 1 |
| 1 | 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 | 1 |
| 1 | 1 | 0 | 1 | 0 |
| 1 | 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 1 | 0 |

The rule can also be represented in the "Birth"/"Survival" notation as **B123/S0**, or as polynomial over the Galois field $\mathbb{F}_2$ (see [3]):

$$x_4' = x_4 + x_1x_2x_3 + x_1x_2 + x_1x_3 + x_2x_3 + x_1 + x_2 + x_3 \ .$$

The phase portraits on a tetrahedron and a cube are sketched in Fig. 3. A number within a circle representing a state orbit is the size of the orbit.

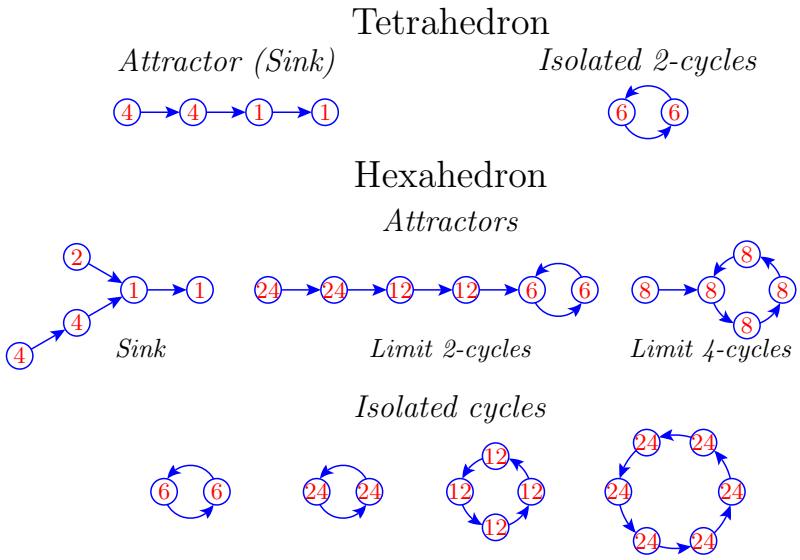



Figure 3: Rule 86. Equivalence classes of trajectories on tetrahedron and cube.

As for the dodecahedron case, the phase portrait consists of 3774 classes of trajectories and is too large to be pictured. Instead, a concise form of output with information about numbers and periods of isolated and limit cycles is presented below:

```
======================== Basic values ========================
Valence of lattice                          VALENCE_K =       3
Number of lattice vertices                    GRAPH_V =      20
Order of lattice automorphism group    AUTOMORPHISM_N =     120
Number of vertex values                       VALUE_Q =       2
Number of states                               StateN = 1048576
Number of state orbits                   StateOrbitN =    9436
Number of automata under study AUTOMATON_LOCAL_RULE_N =       1
==============================================================
==================== Automaton  86 =====================
Isolated cycles modulo automorphisms
Period Number
     2     25
     4     25
     6     17
    10      1
Limit cycles modulo automorphisms
Period Number
     1      1
     2      8
     4     24
     6     24
    10      5
    12      4
    20      1
    22      2
    30      1
There are 3636 separate influxes
==========================================================
===== Max Period =  30   Max Trajectory Length =  41 =====
===== Number of Trajectory Classes = 3774            =====
======================= Statistics =======================
Time: 0.64 sec
```

### 4.0.2 Reversibility.

The program is able to select automata with properties specified at input. One such important property is *reversibility*.

In this connection we would like to mention recent works by G. 't Hooft. One of the difficulties of Quantum Gravity is a conflict between the irreversibility of Gravity — information loss (dissipation) at the black hole horizon — with the reversibility and unitarity of standard Quantum Mechanics. One of the approaches to reconcile both theories proposed by 't Hooft (see, e.g., [6, 7]) is based on the following concepts

- physical systems have *discrete degrees of freedom* at tiny distance scales;
- the states of these degrees of freedom form the *primordial* basis of a (nonunitary) Hilbert space;
- primordial states form *equivalence classes*: two states are equivalent if they evolve into the same state after some lapse of time;
- the equivalence classes form *by construction* the basis of a *unitary* Hilbert space and now evolution is described by the time-reversible Schrödinger equation.

In our terminology this corresponds to transition to limit cycles: after a short time of evolution the limit cycle becomes physically indistinguishable from a reversible isolated cycle. Computation confirms that such behavior is typical for dynamical systems, whereas reversibility is a rather rare property tending to disappear with the growth of complexity of a system. Moreover, reversible systems revealed in our computation are trivial, as one can see below.

Applying our program to all 136 symmetric 3-valent automata we have the following two rules trivially reversible on all lattices:

- $85 \sim \mathrm{B0123/S} \sim x'_4 = x_4 + 1$,
- $170 \sim \mathrm{B/S0123} \sim x'_4 = x_4$.

Besides these uninteresting rules there are 6 reversible rules on a *tetrahedron*:

- $43 \sim \mathrm{B0/S012} \sim x'_4 = x_4(x_2x_3 + x_1x_3 + x_1x_2 + x_3 + x_2 + x_1) + x_1x_2x_3 + x_2x_3 + x_1x_3 + x_1x_2 + x_3 + x_2 + x_1 + 1$;
- $51 \sim \mathrm{B02/S02} \sim x'_4 = x_3 + x_2 + x_1 + 1$;
- $77 \sim \mathrm{B013/S1} \sim x'_4 = x_4(x_2x_3 + x_1x_3 + x_1x_2 + x_3 + x_2 + x_1 + 1) + x_1x_2x_3 + x_2x_3 + x_1x_3 + x_1x_2 + 1$;
- $178 \sim \mathrm{B2/S023} \sim x'_4 = x_4(x_2x_3 + x_1x_3 + x_1x_2 + x_3 + x_2 + x_1 + 1) + x_1x_2x_3 + x_2x_3 + x_1x_3 + x_1x_2$;
- $204 \sim \mathrm{B13/S13} \sim x'_4 = x_3 + x_2 + x_1$;
- $212 \sim \mathrm{B123/S3} \sim x'_4 = x_4(x_2x_3 + x_1x_3 + x_1x_2 + x_3 + x_2 + x_1) + x_1x_2x_3 + x_2x_3 + x_1x_3 + x_1x_2 + x_3 + x_2 + x_1$.

Two already listed of the above rules, 51 and 204, are reversible on a *cube*. There are no nontrivial reversible rules on a *dodecahedron* or a *3×4 graphene*.

# 5 Mesoscopic Lattice Models

The state of a deterministic dynamical system at any point of time is determined uniquely by previous states of the system. A Markov chain — for which transition from any state to any other state is possible with some probability — is a typical example of a *non-deterministic* dynamical system. In this section we apply a symmetry approach to the lattice models in statistical mechanics. These models can be regarded as special instances of Markov chains. *Stationary distributions* of these Markov chains are studied by the methods of statistical mechanics.

The main tool of conventional statistical mechanics is the Gibbs ***canonical ensemble*** – an imaginary collection of identical systems placed in a huge thermostat with temperature $T$. The statistical properties of a canonical ensemble are encoded in the *canonical partition function*

$$Z = \sum_{\sigma \in \Sigma} \mathrm{e}^{-E_\sigma / k_B T} \quad . \tag{2}$$

Here $\Sigma$ is the set of microstates, $E_\sigma$ is energy of microstate $\sigma$, and $k_B$ is Boltzmann's constant. The canonical ensemble being essentially an asymptotic concept — its formulation is based on an approximation called the "thermodynamic limit" — is applicable only to large (strictly speaking, infinite) homogeneous systems.

Nowadays much attention is paid to study systems which are too large for a detailed microscopic description but too small for essential features of their behavior to be expressed in terms of classical thermodynamics. A discipline — often called ***mesoscopy*** — to study such systems covers a wide range of applications from nuclei, atomic clusters, and nanotechnological structures to multi-star systems [8, 9, 10]. To study *mesoscopic* systems one should use the more fundamental ***microcanonical ensemble*** instead of the canonical one. A microcanonical ensemble is a collection of identical isolated systems at fixed energy. Its definition does not include any approximating assumptions. In fact, the only key assumption of a microcanonical ensemble is that all its microstates are equally probable. This leads to the *entropy* formula

$$S_E = k_B \ln \Omega_E \ , \tag{3}$$

or, equivalently, to the *microcanonical partition function*

$$\Omega_E = \mathrm{e}^{S_E/k_B} \ . \tag{4}$$

Here $\Omega_E$ is the number of microstates at fixed energy $E$. We omit further Boltzmann's constant assuming $k_B = 1$. Note that in the thermodynamic limit the microcanonical and canonical descriptions are equivalent and the link between them is provided by the Laplace transform. On the other hand, mesoscopic systems demonstrate observable experimentally and in computation peculiarities of behavior, such as heat flowing from cold to hot, negative specific heat, and "convex intruders" in the entropy versus energy diagram, etc. These anomalous – from the canonical thermostatistics point of view – features have natural explanation within microcanonical statistical mechanics [10]. Note also that finite and long-range interacting infinite systems demonstrate similar statistical behavior. Such systems are called *nonextensive* in statistical mechanics.

The symmetry approach to study ***mesoscopic lattice models*** is based on exact enumeration of group orbits of microstates. Since statistical studies are based essentially on simplifying assumptions, it is important to control these assumptions by exact computation, wherever possible. Moreover, exact computation, hopefully, might reveal subtle details in behavior of considered systems.

As a typical example, let us consider the ***Ising model***. The model consists of *spins* placed on a lattice. The set of vertex values is $Q = \{-1, 1\}$ and the interaction Hamiltonian is given by

$$H = -J \sum_{(i,j)} s_i s_j - B \sum_i s_i \ , \tag{5}$$

where $s_i, s_j \in Q$; $J$ is a coupling constant ($J > 0$ and $J < 0$ correspond to *ferromagnetic* and *antiferromagnetic* cases, respectively); the first sum runs

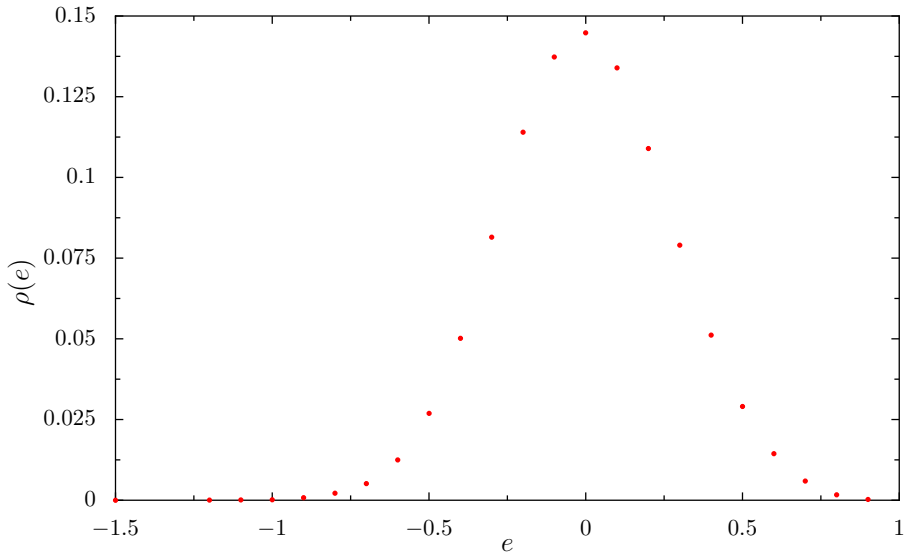



Figure 4: Microcanonical density of states $\rho(e) = \Omega_E/\Omega$ versus energy per vertex $e = E/V(G_L)$ for the Ising model on a dodecahedron.

over all edges $(i, j)$ of the lattice; $B$ is an external "magnetic" field. The second sum $M = \sum_i s_i$ is called the *magnetization*. To avoid technical details we will consider only the case $J > 0$ (we take $J = 1$) and $B = 0$ in what follows.

Since Hamiltonian and magnetization are constants on the group orbits, we can count the number of microstates corresponding to particular values of these functions – and hence compute all needed statistical characteristics – simply by summation of sizes of appropriate orbits.

Fig. 4 shows the microcanonical partition function for the Ising model on a dodecahedron. Of course, other characteristics of the system can be computed easily via counting sizes of group orbits, e.g., the magnetization is shown in Fig. 5.

The needs of nanotechnological science and nuclear physics focus special attention on ***phase transitions in finite systems***. Unfortunately, classical thermodynamics and the rigorous theory of critical phenomena in homogeneous infinite systems fails at the mesoscopic level. Several approaches have been proposed to identify phase transitions in mesoscopic systems. The most accepted of them is search for *"convex intruders"* [11] in the entropy versus energy diagram. In standard thermodynamics there is a relation

$$\left.\frac{\partial^2 S}{\partial E^2}\right|_V = -\frac{1}{T^2}\frac{1}{C_V} \ , \tag{6}$$

where $C_V$ is the specific heat. We see from (6) that $\left.\frac{\partial^2 S}{\partial E^2}\right|_V < 0$ and hence the entropy versus energy diagram must be concave. Nevertheless, in mesoscopic

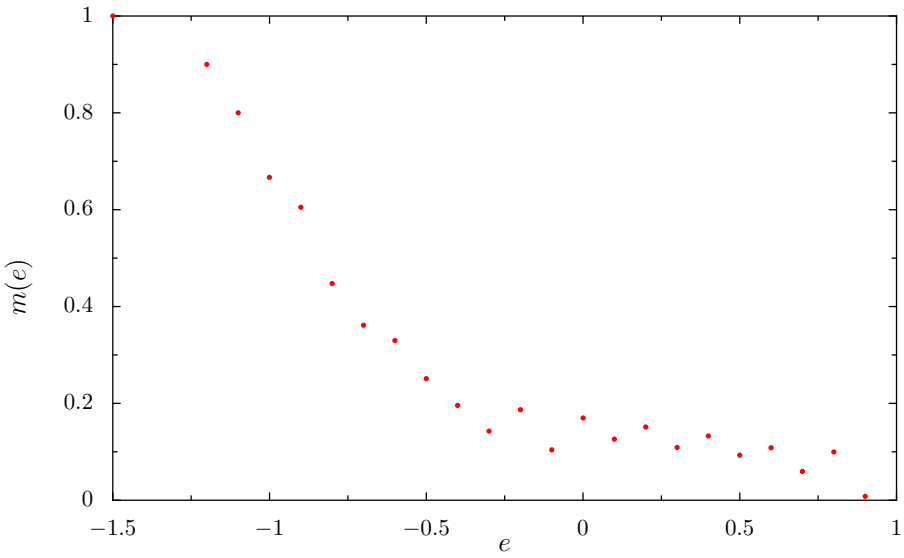



Figure 5: Specific magnetization $m(e) = M(E)/V(G_L)$ vs. energy per vertex $e$ for the Ising model on a dodecahedron.

systems there might be intervals of energy where $\left.\frac{\partial^2 S}{\partial E^2}\right|_V > 0$. These intervals correspond to first-order phase transitions and are called ***convex intruders***. From the point of view of standard thermodynamics one can speak about phenomenon of *negative heat capacity*, of course, if one accepts that it makes sense to define the variables $T$ and $C_V$ as temperature and the specific heat, respectively, in these circumstances. In [12] it was demonstrated via computation with exactly solvable lattice models that the convex intruders flatten and disappear in the models with local interactions as the lattice size grows, while in the case of long-range interaction these peculiarities survive even in the infinite limit.

A convex intruder can be found easily by computer for the discrete systems we discuss here. Let us consider three adjacent values of energy $E_{i-1}, E_i, E_{i+1}$ and corresponding numbers of microstates $\Omega_{E_{i-1}}, \Omega_{E_i}, \Omega_{E_{i+1}}$. In our discrete case the ratio $(E_{i+1} - E_i) / (E_i - E_{i-1})$ is always a rational number $p/q$ and we can write the convexity condition for entropy in terms of numbers of microstates as the easily computed inequality

$$\Omega_{E_i}^{p+q} < \Omega_{E_{i-1}}^{p} \Omega_{E_{i+1}}^{q} \quad . \tag{7}$$

As a rule $E_{i+1} - E_i = E_i - E_{i-1}$ and the inequality (7) takes the form

$$\Omega_{E_i}^{2} < \Omega_{E_{i-1}} \Omega_{E_{i+1}} \quad .$$

This means that within a convex intruder the number of states with energy $E_i$ is less than the *geometric mean* of the number of states at neighboring energies.

Fig. 6 shows the entropy vs. energy diagram for the Ising model on a dodecahedron. The diagram has an apparent convex intruder in the energy

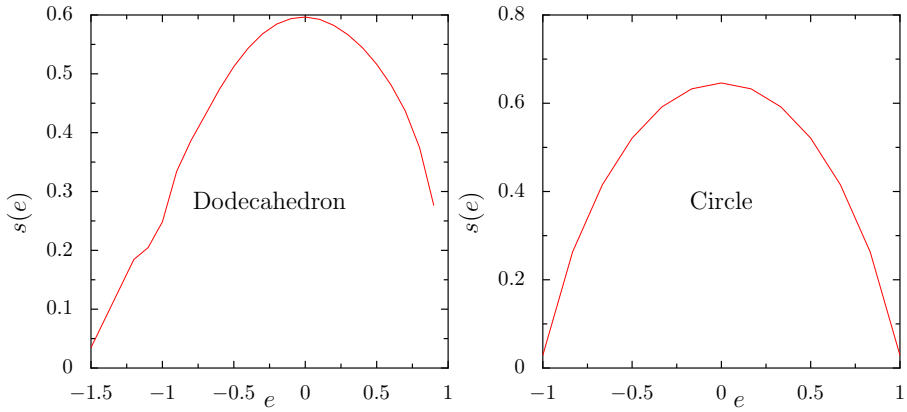



Figure 6: Specific microcanonical entropy $s(e) = \ln(\Omega_E)/V(G_L)$ vs. energy per vertex $e$ for the Ising model on a dodecahedron and on a circle of length 24. The left diagram contains a distinct convex intruder in the interval $-1.2 \leq e \leq -0.9$ and subtle one in the interval $-0.8 \leq e \leq -0.6$. The right diagram is fully concave: a one-dimensional Ising model has no phase transitions.

interval $[-24, -18]$. Exact computation also reveals a subtle convex intruder in the interval $[-16, -12]$. In terms of specific energy, as in Fig. 6, these intervals are $[-1.2, -0.9]$ and $[-0.8, -0.6]$, respectively. It is known that a one-dimensional Ising model has no phase transitions. To illustrate the difference between the diagrams for the cases with and without phase transitions, we place also in Fig. 6 the diagram for the Ising model on the 1D circle lattice with 24 vertices.

# 6 Summary

- A C program for symmetry analysis of finite discrete dynamical systems has been created.
- We pointed out an obvious but important restrictions on possible trajectories of deterministic dynamical systems with non-trivial symmetry.
- Some computational results for cellular automata with symmetric local rules are presented. For these dynamical systems we showed in particular that reversibility is a rare property.
- We demonstrated the capability of exact computing based on symmetries in search of phase transitions for mesoscopic models in statistical mechanics.

**Acknowledgments.** This work was supported in part by grants 07-01-00660 from the Russian Foundation for Basic Research and 5362.2006.2 from the Ministry of Education and Science of the Russian Federation.

## References

[1] Kornyak, V.V.: On Compatibility of Discrete Relations. In: *Computer Algebra in Scientific Computing 2005*, LNCS **3718**, V.G. Ganzha, E.W. Mayr, E.V. Vorozhtsov (Eds.), Springer-Verlag Berlin Heidelberg, 272–284 (2005),
http://arXiv.org/abs/math-ph/0504048

[2] Kornyak, V.V.: Discrete Relations On Abstract Simplicial Complexes. *Programming and Computer Software* **32** No. 2, 84–89 (2006)

[3] Kornyak, V.V.: Cellular Automata with Symmetric Local Rules. In: *Computer Algebra in Scientific Computing 2006*, LNCS **4194**, V.G. Ganzha, E.W. Mayr, E.V. Vorozhtsov (Eds.), Springer-Verlag Berlin Heidelberg, 240–250 (2006),
http://arXiv.org/abs/math-ph/0605040

[4] McKay, B.D.: Practical Graph Isomporphism. *Congressus Numerantium* **30**, 45–87 (1981), http://cs.anu.edu.au/ bdm/nauty/PGI

[5] Gardner, M.: On Cellular Automata Self-reproduction, the Garden of Eden and the Game of Life. *Sci. Am.* **224**, 112-117 (1971)

[6] 't Hooft, G.: Quantum Gravity as a Dissipative Deterministic System. SPIN-1999/07, gr-qc/9903084; *Class. Quant. Grav.* **16**, 3263 (1999); also published in: *Fundamental Interactions: from symmetries to black holes* (Conference held on the occasion of the "Eméritat" of François Englert, 24-27 March 1999, ed. by J.-M. Frère et al, Univ. Libre de Bruxelles, Belgium, 221–240 (1999)

[7] 't Hooft, G.: The mathematical basis for deterministic quantum mechanics. ITP-UU-06/14, SPIN-06/12, quant-ph/0604008, 1–17 (2006)

[8] Imry, Y.: *Introduction to Mesoscopic Physics (Mesoscopic Physics and Nanotechnology, 2).* Oxford University Press, USA, 2002, 256 p.

[9] Gross, D.H.E.: *Microcanonical thermodynamics: Phase transitions in "Small" Systems.* World Scientific, Singapore, 2001, 269 p.

[10] Gross, D.H.E.: A New Thermodynamics from Nuclei to Stars. *Entropy*, **6**, 158-179 (2004)

[11] Gross, D.H.E., Votyakov, E.V.: Phase Transitions in "Small" Systems. *Eur. Phys. J. B*, **15**, 115-126 (2000)

[12] Ispolatov, I., Cohen, E. G. D.: On First-order Phase Transitions in Microcanonical and Canonical Non-extensive Systems, *Physica A* **295**, 475–487 (2001)

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# On Computer Algebra-aided Stability Analysis of Difference Schemes Generated by Means of Gröbner Bases

Vladimir P.Gerdt[1], Yury A.Blinkov[2]

[1] *Laboratory of Information Technologies*
*Joint Institute for Nuclear Research*
*141980 Dubna, Russia*
gerdt@jinr.ru
[2] *Department of Mathematics and Mechanics*
*Saratov State University*
*410071 Saratov, Russia*
BlinkovUA@info.sgu.ru

**Abstract.** Recently we developed an algorithmic method to generate finite difference schemes for linear partial differential equations. The method is based on difference elimination by means of Gröbner bases. To analyze the stability of a generated difference scheme one can compute its differential approximation which is also called the modified equation. There are whole classes of difference schemes for which their stability properties can be studied by this method. However, in many cases, the underlying computation can be done with modern computer algebra software. In this paper we briefly describe the method of differential approximation for Burger's equation and demonstrate both generation of a difference scheme and computation of its differential approximation with *Maple*.

# 1 Introduction

## 1.1 Finite difference approach

The *finite difference approach* [1, 2, 3, 4, 5] with finite elements and finite volumes is the most popular discretization technique for solving ordinary or partial differential equations (PDEs) numerically. In this approach derivatives are approximated by finite differences and the resulting algebraic system – *a difference scheme* – is solved numerically.

In [6] we described an *algorithmic method to generate finite difference schemes* for linear PDEs. The method is based on enlargement of the equations in their integral or differential form by extra integral relations between unknown functions and their derivatives, and on discretization of the obtained system. The

structure of the last system depends on numerical approximation methods for the added integrals. As a result, a system of linear difference equations is derived for the unknown functions and their partial derivatives. A difference scheme is constructed by elimination of all the partial derivatives. The elimination can be achieved by selecting a proper elimination ranking and by computing a Gröbner basis of the linear difference ideal generated by the polynomials in the discrete system.

Sometimes Gröbner bases can be computed even for nonlinear difference systems, as we show in [6] for the Falkowitch-Karman equation describing transonic flow in gas dynamics, derived from discretization of the original PDEs and related integral equations. In this case nonlinear difference schemes can also be generated by our method.

### 1.2 Stability of difference schemes

A difference scheme, to be of practical interest, must be stable. The stability study of difference schemes exploits symbolic mathematical operations. Thus it can be analyzed with help of computer algebra methods and software [7].

To analyze stability one can use a *differential approximation* that is often called the modified equation(s) of difference scheme. There are wide classes of difference schemes whose stability properties can be investigated with the aid of the differential approximation [8]. For all that, in many cases, the computation can be done by means of modern computer algebra software.

In this paper we shall demonstrate how *Maple* can be used for this purpose and present a *Maple* program that computes differential approximations for difference schemes. Together with the *Maple* package for constructing Gröbner bases for linear difference systems [9] the program allows one to generate schemes possessing stability properties.

## 2 Difference schemes for hyperbolic equations

### 2.1 Difference Cauchy problem

Consider the following Cauchy problem

$$\frac{\partial u}{\partial t} = Au, \quad -\infty < x < \infty, \quad t > 0\,, \tag{1}$$

$$u(x,0) = u_0(x), \quad -\infty < x < \infty \tag{2}$$

where $x$ is the spatial variable, $t$ is the temporal variable, $A$ is a linear differential operator and $u_0(x)$ is a given function.
We approximate the Cauchy problem (1)-(2) by the following *difference Cauchy problem*:

$$\frac{u_j^{n+1} - u_j^n}{\tau} = \Lambda_1 u_j^{n+1} + \Lambda_2 u_j^n, \quad j = 0, \pm 1, \pm 2, \ldots; \quad n = 0, 1, 2, \ldots \tag{3}$$

$$u_j^0 = u_0(x_j)\,.$$

### 2.2 Notion of approximation for the initial problem

Let $L$ be the differential operator in the initial equation (1), i.e.

$$Lu = \frac{\partial u}{\partial t} - Au, \tag{4}$$

and let $L_h$ be a difference operator defined in accordance to (4) as

$$L_h u = \frac{u(x, t+\tau) - u(x,t)}{\tau} - \Lambda_1 u(x, t+\tau) - \Lambda_2 u(x,t)\,.$$

Suppose the solution $u(x,t)$ of the Cauchy problem (1)-(2) is smooth enough. If

$$\|Lu - L_h u\| \le C_1 h^{k_1} + C_2 \tau^{k_2}$$

where $k_1 > 0$, $k_2 > 0$ and constants $C_1$ and $C_2$ do not depend on $\tau$ and $h$, then it is said that difference scheme (3) approximates equation (1) and has the order of approximation $k_1$ w.r.t. $h$ and the order $k_2$ w.r.t. $\tau$.

## 3 First differential approximation of difference scheme

The *first differential approximation* (FDA) of difference scheme (3) is the partial differential equation which is obtained from (3) substituting the Taylor expansion for the grid function, while keeping the main terms [8].

One distinguishes *hyperbolic* and *parabolic* forms of a FDA [7]. To obtain the parabolic form of a FDA one uses differential consequences of the initial PDE

$$\frac{\partial u}{\partial t} = Au$$

obtained by differentiating both sides of the PDE w.r.t. the independent variables.

Discretization of PDEs implies that their difference solutions do not satisfy the original PDEs. Deviation of a difference solution from the exact one is called an *error of difference scheme*. Study and classification of errors is based on representing the solution by a trigonometric Fourier series, detecting the variation in amplitude and phase of each harmonic in one step in time, and comparing it with the variation of the exact solution (of PDEs) on the same time interval.

If the harmonic amplitude decreases faster then that for the exact solution, then this effect is called the *amplitude error* of the scheme caused by an extra diffusion inherent to the scheme – *numerical viscosity*. The phase variation of the difference solution distinct from that for the exact solution is called the *phase error* caused by distinction in the phase velocities of the harmonic propagation – *numerical dispersion*.

### 3.1 Example: Lax scheme for Burger's equation

Consider Burgers' equation [1, 2, 5]

$$u_t + f_x = \nu\, u_{xx}, \quad \nu = \text{const} \tag{5}$$

where we denoted $u^2$ by $f$ in order to deal with linear expressions. One of the difference schemes for equation (5) has the form [6]

$$\frac{2u^{n+1}_{j+2} - (u^n_{j+3} + u^n_{j+1})}{2\tau} + \frac{f^n_{j+3} - f^n_{j+1}}{2h} = \nu \frac{u^n_{j+4} - 2u^n_{j+2} + u^n_j}{4h^2}\,. \tag{6}$$

Its differential approximation at the point $(n, j+2)$ is given by

$$\begin{aligned} &\overbrace{u_t + f_x - \nu u_{xx}} - \tfrac{1}{2}u_{xx}\tfrac{h^2}{\tau} + \\ &\quad + \tfrac{1}{2}u_{tt}\tau + (\tfrac{1}{6}f_{xxx} - \tfrac{1}{3}\nu u_{xxxx})h^2 - \tfrac{1}{24}u_{xxxx}\tfrac{h^4}{\tau} + \\ &\qquad + \tfrac{1}{6}u_{ttt}\tau^2 + (\tfrac{1}{120}f_{xxxxx} - \tfrac{2}{45}\nu u_{xxxxxx})h^4 - \tfrac{1}{720}u_{xxxxxx}\tfrac{h^6}{\tau} + \\ &\qquad\quad + \ldots = 0\,. \end{aligned} \tag{7}$$

From (7) it follows that scheme (6) does not approximate equation (5) at $\mathrm{O}(h^2/\tau) \sim 1$. It is an example of a *conditionally convergent scheme.*

For a more detailed analysis of scheme (6) one can construct a *parabolic form* of the FDA for $f = u^2/2$:

$$\begin{aligned} &u_t + uu_x - \overbrace{\left(\nu + \frac{h^2}{2\tau} + (2\nu\tau + \frac{h^2}{2})u_x - \frac{\tau}{2}u^2\right)}\, u_{xx} + \\ &\quad + (u\tau)u_x - (\nu u\tau + \tfrac{h^2}{3}u)u_{xxx} + (\tfrac{\nu^2\tau}{2} + \tfrac{\nu h^2}{6} + \tfrac{h^4}{12\tau})u_{xxxx} = 0. \end{aligned} \tag{8}$$

Formula (8) shows that to approximate the initial PDE the expression marked by $\overbrace{\ldots}$ is to be of order $\nu$ whereas the remaining terms which do not occur in PDE are to be small (vanishing). For equation (5) and zero viscosity ($\nu = 0$) the expression marked by $\overbrace{\ldots}$ must be positive. Otherwise the diffusion coefficient becomes negative and the boundary-value problem becomes incorrect.

For some classes of PDEs one can use the *equivalence theorem* to relate the scheme stability with its differential approximation.

## 4 Generation of difference schemes

In [6] we suggested an algorithmic approach to generation of difference schemes. For instance, consider Burger's equation (5) and add the integral relations together with their numerical approximation as follows

$$\begin{aligned} &\int u_t dt = u, \quad u_t\tau = u^{n+1}_j - \frac{u^n_{j+2} + u^n_j}{2}, \\ &\int f_x dx = f, \quad 2\,h(f_x)^n_{j+1} = f^n_{j+2} - f^n_j, \\ &\int u_x dx = u, \quad 2\,h(u_x)^n_{j+1} = u^n_{j+2} - u^n_j, \\ &\int u_{xx} dx = u_x, \quad 2\,h(u_{xx})^n_{j+1} = (u_x)^n_{j+2} - (u_x)^n_j. \end{aligned}$$

Computation of a Gröbner basis for an elimination ranking $u_{xx} \succ u_t \succ u_x \succ f_x \succ u \succ f$ derives scheme (6).

We call such a method of integration in time as a Lax-type one and use different numerical quadrature formulae for integration over $x$. For the midpoint or the trapezoidal rule for these integrals 8 different schemes are generated.

Here, a question arises: *How closely related are the properties of these 8 schemes?* A partial answer can be obtained by using the differential approximation technique described in the next section.

# 5 Construction of differential approximation

## 5.1 Hyperbolic form

Since an order (for instance, in $h$) of the ratio $\frac{h}{\tau}$ is not known, one cannot specify a linear order w.r.t. divisibility in $\tau$ and $h$ for terms in the differential approximation for a PDE. Equation (7) multiplied by $\tau$ can be partitioned into three groups determining the hyperbolic form:

$$\begin{aligned}&\big[(u_t + f_x - \nu u_{xx})\tau - \tfrac{1}{2}u_{xx}h^2,\\ &\tfrac{1}{2}u_{tt}\tau^2 + (\tfrac{1}{6}f_{xxx} - \tfrac{1}{3}\nu u_{xxxx})\tau h^2 - \tfrac{1}{24}u_{xxxx}h^4,\\ &\tfrac{1}{6}u_{ttt}\tau^3 + (\tfrac{1}{120}f_{xxxxx} - \tfrac{2}{45}\nu u_{xxxxxx})\tau h^4 - \tfrac{1}{720}u_{xxxxxx}h^6\big]\end{aligned}$$

We see that

- the first group does not have divisors;
- the second group has divisors occurring in the first group;
- the third group has divisors occurring in the first and the second groups.

In so doing, a truncation order for the Taylor expansion is to be specified to provide a correct partition into groups.

## 5.2 Parabolic form

To construct a FDA in parabolic form, the derivatives w.r.t. $t$ in the second group are replaced by their values from the first group so that we can compare scheme properties. It can be achieved by the sequential substitution of derivatives in accordance to the lexicographic order $u_{tt} \succ u_{tx} \succ u_t$. Thereby we obtain

$$\begin{aligned}&\Big[(u_t + uu_x - \nu u_{xx})\tau - u_{xx}\tfrac{h^2}{2},\\ &+(-\nu u u_{xxx} - 2\nu u_x u_{xx} + \tfrac{1}{2}\nu^2 u_{xxxx} + \tfrac{1}{2}u^2 u_{xx} + u u_x^2)\tau^2\\ &+(\tfrac{1}{6}u_{xxxx}\nu - \tfrac{1}{3}u_{xxx}u - \tfrac{1}{2}u_{xx}u_x)\tau h^2 + \tfrac{1}{6}u_{xxxx}h^4\Big]\ .\end{aligned}$$

## 5.3 Implementation in *Maple*

We implemented the above described method in *Maple* as a package, FDA (First Differential Approximation). To illustrate the work of the package for Burger's equation we use also the *Maple* package, LDA (Linear Differential Algebra) [9] which allows to compute a Janet basis for an input system of linear difference equations and thereby to generate a difference scheme. The *Maple* session below shows both generation of the difference scheme for Burger's equation and computation of its parabolic form:

```
>restart;

>libname:=libname, "/usr/local/lib/lda", "/usr/local/lib/Janet", "/usr/local/lib/fda":

>with(LDA);
```

[*AddRelation*, *AffEqn*, *AppShiftOp*, *AssertJanetBasis*, *CartanCharacter*, *CompCond*, *CompCondBasis*, *Diff2Shift*, *HF*, *HP*, *HilbertFunction*, *HilbertPolynomial*, *HilbertSeries*, *IndexRegularity*, *InvReduce*, *JanetBasis*, *LDAOptions*, *LDAStats*, *LeadingShift*, *Pol2Shift*, *RemoveLowerOrderTerms*, *ResidueClassBasis*, *ResidueClassRelations*, *ReverseShift*, *Shift2Diff*, *Shift2Op*, *Shift2Pol*, *ShiftGroebnerBasis*, *ShiftOpRepres*, *ShiftRepres*, *ShiftTabVar*, *WeightedHilbertSeries*, *ZeroSets*]

```
>with(FDA);
```

[*DForm*, *PForm*]

```
>L:=[ut(n,j)+fx(n,j)-nu*uxx(n,j),
>ut(n,j+1)*tau-(u(n+1,j+1)-(u(n,j+2)+u(n,j))/2),
>2*fx(n,j+1)*h-(f(n,j+2)-f(n,j)),
>2*ux(n,j+1)*h-(u(n,j+2)-u(n,j)),
>2*uxx(n,j+1)*h-(ux(n,j+2)-ux(n,j))];
```

$$
\begin{aligned}
&[ut\,(n,j)+fx\,(n,j)-\nu\,uxx\,(n,j)\,,\,ut\,(n,j+1)\,\tau-u\,(n+1,j+1)+\tfrac{1}{2}\,u\,(n,j+2)+\\
&\tfrac{1}{2}\,u\,(n,j)\,,\;2\,fx\,(n,j+1)\,h-f\,(n,j+2)+f\,(n,j)\,,\\
&\;2\,ux\,(n,j+1)\,h-u\,(n,j+2)+u\,(n,j)\,,\\
&\;2\,uxx\,(n,j+1)\,h-ux\,(n,j+2)+ux\,(n,j)]
\end{aligned}
$$

```
>JanetBasis(L, [n,j], [uxx,ux,ut,fx,u,f],2);
```

$$[[-4\,\frac{h^2u(n+1,j+2)}{\nu\,\tau}+u\,(n,j)+u\,(n,j+4)+2\,\frac{h^2u(n,j+3)}{\nu\,\tau}-2\,u\,(n,j+2)+2\,\frac{h^2u(n,j+1)}{\nu\,\tau}$$
$$-2\,\frac{hf(n,j+3)}{\nu}+2\,\frac{hf(n,j+1)}{\nu},2\,fx\,(n,j+1)\,h-f\,(n,j+2)+f\,(n,j)\,,$$
$$ut\,(n,j+1)\,\tau-u\,(n+1,j+1)+\tfrac{1}{2}\,u\,(n,j+2)+\tfrac{1}{2}\,u\,(n,j)\,,$$
$$ux\,(n,j)+2\,\frac{hu(n+1,j+1)}{\nu\,\tau}-\tfrac{1}{2}\,\frac{u(n,j+3)}{h}-\frac{hu(n,j+2)}{\nu\,\tau}$$
$$+\tfrac{1}{2}\,\frac{u(n,j+1)}{h}-\frac{hu(n,j)}{\nu\,\tau}+\frac{f(n,j+2)}{\nu}-\frac{f(n,j)}{\nu},$$
$$ut\,(n,j)+fx\,(n,j)-\nu\,uxx\,(n,j)],[n,j],[a,uxx,ux,ut,fx,u,f]]$$

```
>collect(%[1,1]*nu/(4*h^2),[tau,h,nu]);
```

$$\frac{-1/2\,f(n,j+3)+1/2\,f(n,j+1)}{h}+\frac{(1/4\,u(n,j+4)-1/2\,u(n,j+2)+1/4\,u(n,j))\nu}{h^2}$$
$$+\frac{1/2\,u(n,j+3)+1/2\,u(n,j+1)-u(n+1,j+2)}{\tau}$$

```
>a:=-DForm(%,[u,f],[[n,tau,t],[j,h,x]],[0,2],2);
```

$$[D_2\,(f)\,(t,x)-D_{2,2}\,(u)\,(t,x)\,\nu+D_1\,(u)\,(t,x)-\tfrac{1}{2}\,\frac{D_{2,2}(u)(t,x)h^2}{\tau},$$
$$\tfrac{1}{2}\,D_{1,1}\,(u)\,(t,x)\,\tau-\left(-\tfrac{1}{6}\,D_{2,2,2}\,(f)\,(t,x)+\tfrac{1}{3}\,D_{2,2,2,2}\,(u)\,(t,x)\right)\nu h^2$$
$$-\tfrac{1}{24}\,\frac{D_{2,2,2,2}(u)(t,x)h^4}{\tau},$$
$$\tfrac{1}{6}\,D_{1,1,1}\,(u)\,(t,x)\,\tau^2-\left(-\tfrac{1}{120}\,D_{2,2,2,2,2}\,(f)\,(t,x)+\tfrac{2}{45}\,D_{2,2,2,2,2,2}\,(u)\,(t,x)\right)\nu h^4$$
$$-\tfrac{1}{720}\,\frac{D_{2,2,2,2,2,2}(u)(t,x)h^6}{\tau}]$$

```
>f:=u^2/2:
>PForm(a);
```

$$[D_2\,(u)\,(t,x)\,u\,(t,x)-D_{2,2}\,(u)\,(t,x)\,\nu+D_1\,(u)\,(t,x)-\tfrac{1}{2}\,\frac{D_{2,2}(u)(t,x)h^2}{\tau},$$
$$(-\nu\,D_{2,2,2}\,(u)\,(t,x)\,u\,(t,x)-2\,\nu\,D_{2,2}\,(u)\,(t,x)\,D_2\,(u)\,(t,x)$$
$$+\tfrac{1}{2}\,\nu^2D_{2,2,2,2}\,(u)\,(t,x)+\tfrac{1}{2}\,\left(u\,(t,x)\right)^2D_{2,2}\,(u)\,(t,x)\,+u\,(t,x)\,(D_2\,(u)\,(t,x))^2\big)\,\tau$$
$$-\left(\tfrac{1}{3}\,D_{2,2,2}\,(u)\,(t,x)\,u\,(t,x)+\tfrac{1}{2}\,D_{2,2}\,(u)\,(t,x)\,D_2\,(u)\,(t,x)-\tfrac{1}{6}\,D_{2,2,2,2}\,(u)\,(t,x)\,\nu\right)h^2$$
$$+\tfrac{1}{12}\,\frac{D_{2,2,2,2}(u)(t,x)h^4}{\tau}]$$

In the input set of difference equations the midpoint rule is used for the integral relations. One can also use the trapezoidal rule for spatial integration. This derives other schemes. By selecting either the midpoint or the trapezoidal rule for the spatial integrals, we obtain 8 possible schemes [6]. Two of these 8 schemes coincide, and, hence, there are 7 different schemes. Apart from scheme 6 they are

$$\frac{2(u^{n+1}_{j+2}+u^{n+1}_{j+1})-(u^n_{j+3}+u^n_{j+2}+u^n_{j+1}+u^n_j)}{4\tau}+\frac{(f^n_{j+3}+f^n_{j+2})-(f^n_{j+1}+f^n_j)}{4h}=$$
$$=\nu\frac{(u^n_{j+3}-u^n_{j+2})-(u^n_{j+1}-u^n_j)}{2h^2},$$
$$\frac{2u^{n+1}_{j+1}-(u^n_{j+2}+u^n_j)}{2\tau}+\frac{f^n_{j+2}-f^n_j}{2h}=\nu\frac{u^n_{j+2}-2u^n_{j+1}+u^n_j}{h^2},$$
$$\frac{2(u^{n+1}_{j+3}+2u^{n+1}_{j+2}+u^{n+1}_{j+1})-(u^n_{j+4}+2u^n_{j+3}+2u^n_{j+2}+2u^n_{j+1}+u^n_j)}{8\tau}+\frac{(f^n_{j+4}+2f^n_{j+1})-(2f^n_{j+1}+f^n_j)}{8h}=$$
$$=\nu\frac{u^n_{j+3}-2u^n_{j+2}+u^n_{j+1}}{h^2},$$
$$\frac{2(u^{n+1}_{j+3}+u^{n+1}_{j+2})-(u^n_{j+4}+u^n_{j+3}+u^n_{j+2}+u^n_{j+1})}{4\tau}+\frac{f^n_{j+3}-f^n_{j+2}}{h}=$$
$$=\nu\frac{((u^n_{j+5}+u^n_{j+4})-2u^n_{j+3})-(2u^n_{j+2}-(u^n_{j+1}+u^n_j))}{8h^2},$$
$$\frac{2(u^{n+1}_{j+2}+u^{n+1}_{j+1})-(u^n_{j+3}+u^n_{j+2}+u^n_{j+1}+u^n_j)}{4\tau}+\frac{f^n_{j+2}-f^n_{j+1}}{h}=\nu\frac{(u^n_{j+3}-u^n_{j+2})-(u^n_{j+1}-u^n_j)}{2h^2},$$
$$\frac{2(u^{n+1}_{j+3}+2u^{n+1}_{j+2}+u^{n+1}_{j+1})-(u^n_{j+4}+2u^n_{j+3}+2u^n_{j+2}+2u^n_{j+1}+u^n_j)}{8\tau}+\frac{f^n_{j+3}-f^n_{j+1}}{2h}$$
$$=\nu\frac{u^n_{j+3}-2u^n_{j+2}+u^n_{j+1}}{h^2}.$$

The second scheme from the bottom is obtained twice in the course of generating eight schemes.

Computation of differential approximations for scheme (6) and for the above 6 schemes with $f=u^2$ gives

$$u_t+u_xu-\nu\,u_{xx}=(-\tfrac{1}{2}\,\nu^2u_{xxxx}+(u_{xxx}u+2\,u_{xx}u_x)\nu-u_x^2u-\tfrac{1}{2}\,u^2u_{xx})\tau$$
$$+(*)h^2+(**)\frac{h^2}{\tau}+\ldots=0.$$

where contributions of the order $h^2$ and $h^4/\tau$ read

$$\begin{aligned}
[\ldots &= (\ldots)\tau &&+(\tfrac{1}{6}u_{xxxx}\nu-\tfrac{1}{3}u_{xxx}u-\tfrac{1}{2}u_{xx}u_x)h^2 &&+\tfrac{1}{8}u_{xxxx}\tfrac{h^4}{\tau}]\\
[\ldots &= (\ldots)\tau &&+(\tfrac{5}{12}u_{xxxx}\nu-\tfrac{1}{2}u_{xxx}u-\tfrac{1}{2}u_{xx}u_x)h^2 &&+\tfrac{3}{16}u_{xxxx}\tfrac{h^4}{\tau}]\\
[\ldots &= (\ldots)\tau &&+(\tfrac{5}{12}u_{xxxx}\nu-\tfrac{1}{3}u_{xxx}u-\tfrac{1}{2}u_{xx}u_x)h^2 &&+\tfrac{1}{8}u_{xxxx}\tfrac{h^4}{\tau}]\\
[\ldots &= (\ldots)\tau &&+(\tfrac{2}{3}u_{xxxx}\nu-\tfrac{1}{3}u_{xxx}u-\tfrac{1}{2}u_{xx}u_x)h^2 &&+\tfrac{1}{4}u_{xxxx}\tfrac{h^4}{\tau}] \qquad (9)\\
[\ldots &= (\ldots)\tau &&+(\tfrac{1}{6}u_{xxxx}\nu-\tfrac{7}{12}u_{xxx}u-\tfrac{5}{4}u_{xx}u_x)h^2 &&+\tfrac{3}{16}u_{xxxx}\tfrac{h^4}{\tau}]\\
[\ldots &= (\ldots)\tau &&+(\tfrac{5}{12}u_{xxxx}\nu-\tfrac{7}{12}u_{xxx}u-\tfrac{5}{4}u_{xx}u_x)h^2 &&+\tfrac{3}{16}u_{xxxx}\tfrac{h^4}{\tau}]\\
[\ldots &= (\ldots)\tau &&+(\tfrac{2}{3}u_{xxxx}\nu-\tfrac{7}{12}u_{xxx}u-\tfrac{5}{4}u_{xx}u_x)h^2 &&+\tfrac{1}{4}u_{xxxx}\tfrac{h^4}{\tau}]
\end{aligned}$$

From differential approximations (9) it can be easily seen that all schemes have similar properties, and three of them have identical differential approximations.

By inspection of the schemes we see that they have

- the same order of approximation;
- identical dissipative properties;
- very close dispersion properties with some small distinctions in the rational coefficients at derivatives in the terms of order $h^2$ and $h^4/\tau$.

### 5.4 Two-Step Lax-Wendroff Schemes

Denoting the values of functions on the intermediate time level by $\overline{u}$ and $\overline{f}$ we obtain the following difference system:

$$\left\{\begin{array}{rcl} u_t{}_j^n + f_x{}_j^n & = & \nu\, u_{xx}{}_j^n \\ u_t{}_j^n\,\tau & = & \overline{u}{}_j^{n+1} - \frac{u_{j+2}^n + u_j^n}{2} \\ 2 f_x{}_{j+1}^n\, h & = & f_{j+2}^n - f_j^n \\ 2 u_x{}_{j+1}^n\, h & = & u_{j+2}^n - u_j^n \\ 2 u_{xx}{}_{j+1}^n\, h & = & u_x{}_{j+2}^n - u_x{}_j^n \\ \overline{u}_t{}_j^n + \overline{f}_x{}_j^n & = & \nu\, \overline{u}_{xx}{}_j^n \\ \overline{u}_t{}_j^n\,\tau & = & u_j^{n+1} - u_j^n \\ 2\overline{f}_x{}_{j+1}^n\, h & = & \overline{f}_{j+2}^n - \overline{f}_j^n \\ 2\overline{u}_x{}_{j+1}^n\, h & = & \overline{u}_{j+2}^n - \overline{u}_j^n \\ 2\overline{u}_{xx}{}_{j+1}^n\, h & = & \overline{u}_x{}_{j+2}^n - \overline{u}_x{}_j^n. \end{array}\right.$$

For the elimination ranking

$$u_{xx} \succ \overline{u}_{xx} \succ u_x \succ \overline{u}_x \succ u_t \succ \overline{u}_t \succ f_x \succ \overline{f}_x \succ f \succ u \succ \overline{f} \succ \overline{u}$$

a Gröbner basis contains the Lax-Wendroff scheme

$$\left\{\begin{array}{l} \frac{\overline{u}_{j+2}^{n+1} - (u_{j+3}^n + u_{j+1}^n)}{2\,\tau} + \frac{f_{j+3}^n - f_{j+1}^n}{2\,h} = \nu \frac{u_{j+4}^n - 2u_{j+2}^n + u_j^n}{4h^2}, \\ \frac{u_{j+3}^{n+1} - \overline{u}_{j+2}^n}{2\,\tau} + \frac{\overline{f}_{j+3}^n - \overline{f}_{j+1}^n}{2\,h} = \nu \frac{\overline{u}_{j+4}^n - \overline{u}_{j+2}^n + \overline{u}_j^n}{4\,h^2}. \end{array}\right. \tag{10}$$

With all possible combinations of the trapezoidal and midpoint rules one can obtain 49 different Lax-Wendroff which are similar to scheme (10).

## 6 Conclusions

As shown in [6] and in the present paper, Gröbner bases provide a tool for algorithmic construction of finite difference schemes for linear PDEs. Having a difference scheme constructed, the *method of differential approximation (modified equation) allows one to study the stability of schemes* for a wide class of PDEs. In particular, the first differential approximation (FDA) plays an important role in stability analysis.

For linear and some quasilinear PDEs *differential approximations can be constructed algorithmically*, and the underlying algorithms for computing parabolic and hyperbolic forms of FDA have been implemented in *Maple.*

*The methods and software designed were applied to many different PDEs*, for example, to Burgers' equation.Several difference schemes were generated and their stability properties were studied using the method of differential approximation.

**Acknowledgements.** The research presented in this paper was partially supported by grant 07-01-00660 from the Russian Foundation for Basic Research and by grant 5362.2006.2 from the Ministry of Education and Science of the Russian Federation.

# References

[1] A.Quarteroni and A.Valli. *Numerical Approximation of Partial Differential Equations*, 2nd Edition, Springer-Verlag, Berlin, 1997.

[2] J.W.Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*, 2nd Edition. Springer-Verlag, New York, 1998.

[3] A.A.Samarskii. *The Theory of Difference Schemes*. Marcel Dekker, New York, 2001.

[4] J.C.Strikwerda. *Finite difference schemes and partial differential equations*. Philadelphia, SIAM, 2004.

[5] K.W.Morton and D.F.Mayers. *Numerical Solution of Partial Differential Equations*. 2nd Edition, Cambridge University Press, Cambridge, 2005.

[6] V.P.Gerdt, Yu.A.Blinkov and V.V.Mozzhilkin. Gröbner bases and generation of difference schemes for partial differential equations. *Symmetry, Integrability and Geometry: Methods and Applications*, 2:26, 2006.

[7] V.G.Ganzha and E.V.Vorozhtsov. *Computer-aided analysis of difference schemes for partial differential equations*. New York, Wiley-Interscience, 1996.

[8] Yu.I.Shokin and N.N.Yanenko. *Method of differential approximation. Application to gas dynamics.* Nauka, Siberian Division, 1985 (in Russian).

[9] V.P.Gerdt and D.Robertz. A Maple package for computing Gröbner bases for linear recurrence relations. *Nuclear Instruments and Methods in Physics Research*, A559:215219, 2006.

*Computer Algebra and Differential Equations*
*Acta Academiae Aboensis, Ser. B, Vol. 67, no. 2, 2007*
*A. Mylläri, V. Edneral and N. Ourusoff, eds.*

# REDLOG Online Resources for Applied Quantifier Elimination

Thomas Sturm
*Universität Passau, Germany*

**Abstract.** This paper consists of two parts. In the first part we give an overview of the capabilities of the computer logic system REDLOG, an integral part of the computer algebra system REDUCE since 1999. REDLOG implements quantifier elimination over various domains including, among many others, the reals, the integers, and differential algebras. Our focus here is on previously published work on quantifier elimination methods within the field of differential equations. This comprises real quantifier elimination methods and a quantifier elimination method for differentially closed fields. In the second part we announce the new REDLOG website. This website provides regular updates of precompiled REDLOG binaries. In addition, it features the REDLOG Example Management and Information System (REMIS). This is an online database containing REDLOG input files for computation examples discussed in scientific publications over the years. Moreover, REMIS covers the corresponding publications themselves, and provides all relevant links between REDLOG input examples on the one hand and publications on the other hand. Our references at the end of this article are quite comprehensive in order to provide a good starting point to researchers interested in applying quantifier elimination techniques.

## 1 The REDLOG Computer Logic System

Based on experimental implementations beginning in 1992, the author started the realization of REDLOG in early 1994. In April 1995, the system was redesigned by the author together with A. Dolzmann [9], who from then on equally contributed to the realization of REDLOG. REDLOG 1.0 [7] was published on the Web in October 1996. This first REDLOG distribution was quite a success. In October 1996 both the author and A. Dolzmann joined the REDUCE development group, with the result that REDLOG is now developed not as a separate contributed package but as an integral part of the REDUCE system itself. REDLOG 2.0 [10] appeared as part of the distribution of REDUCE 3.7 issued in April 1999. REDLOG 3.0 appeared as part of the current distribution of REDUCE 3.8 in April 2004. Regular updates of REDLOG are now available on the Web. We are going to describe this in more detail in Section 4.

REDLOG stands for REDUCE *logic system* [9, 10]. It provides an extension of the computer algebra system REDUCE to a computer logic system implementing symbolic algorithms on first-order formulas with respect to temporarily fixed first-order languages and theories. Such a choice of language and theory is called a *domain* or, more technically, a *context*.

We briefly summarize the currently existing domains together with short names and alternative names, which are supported for backward compatibility:

BOOLEAN, B, IBALP The class of Boolean algebras with two elements. These algebras are uniquely determined up to isomorphisms. BOOLEAN comprises quantified propositional calculus [35].

COMPLEX, C, ACFSF The class of algebraically closed fields such as the complex numbers over the language of rings.

DIFFERENTIAL, DCFSF A domain for computing over differentially closed fields. There is no natural example for such a field, but the methods can well be used for obtaining relevant and also interpretable results for reasonable differential fields [11].

INTEGERS, Z, PASF The theory of the integers as an ordered ring with congruences. This comprises Presburger arithmetic and recent work on so-called weak quantifier elimination on the full linear theory and certain special cases of nonlinear formulas [26, 27].

PADICS, DVFSF The discretely-valued fields of $p$-adic numbers for some prime $p$ with abstract divisibility relations encoding order between values. All PADICS algorithms are optionally uniform in $p$ [43].

QUEUES, QQE A (two-sided) queue is a finite sequence of elements of some basic type. There are two sorts of variables, one for the basic type and one for the queue type. Accordingly, there is first-order quantification possible for both sorts. So far, the implementation is restricted to the reals as the basic type [39].

REALS, R, OFSF The class of real closed fields such as the real numbers with ordering. This domain was the original motivation for REDLOG. It is still the most important and most comprehensive one [13].

TERMS, TALP Free Malcev-type term algebras. The available function symbols and their arity can be freely chosen. [46].

The idea of REDLOG is to combine methods from computer algebra with first-order logic, thus extending the computer algebra system REDUCE to a computer logic system. In this extended system both the algebraic side and the logic side greatly benefit from each other in numerous ways. REDLOG began with the implementation of real quantifier elimination. Successfully applying such methods to both academic and real-world problems, the authors have developed a large collection of formula manipulation tools, many of which are interesting in their own right.

## 2 Real Quantifier Elimination

For real quantifier elimination, it is common to use the language of ordered rings $L = (0, 1, +, -, \cdot, \leq)$. Certainly, there are integer coefficients admitted in the input. They are formally an abbreviated notation for terms $\pm(1 + \cdots + 1)$. Consider a first-order formula like the following one originally suggested by H. Hong:

$$\varphi \equiv \forall x \exists y (x^2 + xy + b > 0 \wedge x + ay^2 + b \leq 0).$$

The notion *first*-order refers to the fact that the quantifiers $\forall x$ and $\exists y$ range over real *numbers*. In higher order formulas, quantifiers would possibly range over relations as well. There are two kinds of variables in this formula: The first kind are the quantified variables $x$ and $y$. The second kind are free variables or parameters $a$ and $b$. A quantifier elimination procedure takes such a formula as an input. Its output is an equivalent quantifier-free formula. That is, the output will not contain any quantifiers and, consequently, also not the variables $x$ and $y$. In other words we obtain a necessary and sufficient condition in the parameters $a$ and $b$ for validity of the given formula. We would like to encourage the readers to find the answer for our rather simple example themselves. We are going to present the solution computed with REDLOG at the end of this section.

First, we summarize some successful applications of REDLOG within various fields of science and engineering. We start with applications by the research group around the REDLOG developers:

- parametric and nonlinear optimization [13]
- transportation problems [28, 6]
- circuit analysis, design, and diagnosis [41]
- generalized scheduling problems [13, 6]
- real implicitation [13, 5]
- automated theorem proving [12, 40, 14]
- computational geometry [45, 40]
- solid modeling [44, 40, 42]
- robot motion planning [13, 48, 49, 15]
- guarded expressions [8].

Even more interesting are the following third-party applications. They demonstrate that REDLOG is a useful, stable, and well-documented tool, which is accepted and actively used by a considerable community of scientists:

- automatic loop parallelization [16]
- hybrid control theory [22, 1, 25]

- hydraulic network diagnosis (industrial cooperation project)
- theoretical mechanics [21, 20, 19]
- software security [37, 38].

Our list continues with further third-party applications in the area of differential equations. These are probably of particular interest to the CADE community:

- deciding Hopf bifurcations [23]
- equilibrium point analysis [24, 4]
- stability of differential equations [18]
- deciding ellipticity [36].

We are not going to discuss here any of these applications in detail. The interested reader will find that our citations mostly refer to well-established and widely available journals and conference proceedings. We would like to point in particular at the excellent recent survey by Weber on real quantifier elimination and differential equations [47].

In the next section we are going to discuss quantifier elimination for differentially closed fields. It is noteworthy, however, that all relevant applications of REDLOG to differential equations so far, which are listed above, have used real quantifier elimination.

To close the present section, we give the solution of our real quantifier elimination puzzle: REDLOG computes in about 200 ms that the formula $\forall x \exists y(x^2 + xy + b > 0 \wedge x + ay^2 + b \leq 0)$ discussed above is equivalent to $a < 0 \wedge 16b^2 + 36b + 27 > 0 \wedge b > 0$. This can obviously be simplified to $a < 0 \wedge b > 0$.

# 3 Quantifier Elimination in Differential Fields

Differential algebra is usually attributed to Ritt [30]. Ritt had the idea of treating differential equations to a large extent in a purely algebraic framework and developed such framework. As a major result, he proved his differential Nullstellensatz, which is a perfect analogue to Hilbert's Nullstellensatz for algebraically closed fields [17].

The first major algorithmic contribution in differential algebra was Seidenberg's elimination theory [34]. It provided an elimination theorem for 1-primitive formulas with a perfectly algorithmic proof. Using standard techniques this can be easily extended to arbitrary first-order input formulas. The output is, however, correct only in some differential extension field of the considered field, which depends on the input. Hence Seidenberg did not really provide a quantifier elimination procedure for any fixed structure or theory.

On the basis of Seidenberg's work, Robinson introduced in 1959 the notion of a *differentially closed field* [31]. He axiomatized the class of differentially closed fields by combining the following sets of axioms:

1. The field axioms.
2. The Leibniz axioms for the derivative.
3. For each existential formula, the equivalence between this formula and the corresponding quantifier-free formula obtained according to Seidenberg.

Thus while Seidenberg provided a dynamic process in the sense that equivalence holds in differential extension fields, Robinson switched from the outset to a sufficiently large field such that no such extension would ever become necessary. From a model-theoretic point of view, these differentially closed fields are perfect analogues of algebraically closed fields. Unfortunately, there is no natural example for such fields which could play the role that the complex numbers play for algebraically closed fields. Robinson's main result was the *model completeness* of the class of differentially closed fields. It is obvious that Seidenberg's procedure is an effective quantifier elimination procedure for the class of differentially closed fields. Consequently, beyond model completeness, differentially closed fields have even stronger property of *substructure completeness*, which is equivalent to the existence of quantifier elimination. Interestingly, Robinson who had just one year before discussed this stronger phenomenon [32]—without introducing the notion of substructure completeness, however—did not address this fact at all.

In 1968 Blum reanalyzed Seidenberg's proof w.r.t. the assumptions on the differential extension fields made there [2, 3]. By indirect model theoretic methods, viz. saturated models, she found, in contrast to Robinson's pragmatic collection of all possible results of Seidenberg's procedure, a natural axiomatization of differentially closed fields:

3′. For each pair $f$, $g$ of univariate differential polynomials with $\mathrm{ord}(f) > \mathrm{ord}(g)$ there is a $c$ in the field such that $f(c) = 0$ and $g(c) \neq 0$.

Note that this is still an infinite set of axioms. In contrast to Robinson's, however, these axioms are very natural. In fact, they nicely resemble the axiomatization of algebraically closed fields. At that time the notion of substructure completeness had been introduced by Sacks [33], and the scientific community was generally aware of the fact that differentially closed fields admit quantifier elimination via Seidenberg's procedure.

It is a straightforward idea to come full circle by reformulating Seidenberg's elimination procedure in such a way that exactly Blum's axioms become explicit there. This has actually been done by Weispfenning in 1973. This work has became part of his university lectures on differential algebra during the 1980's. An outline of the procedure can be found in [11], to which we refer the reader for more detailed information.

We have already mentioned that there is no natural example at all for a differentially closed field. That is, quantifier elimination will certainly never take place in structures that users actually have in their minds. It rather takes place in a differentially closed extension field, where there generally exist elements that cannot be interpreted as functions.

Nevertheless, quantifier elimination results will also, for certain first-order formulations of natural questions, provide information on the differential field actually under consideration. In the first place, this applies to input formulas that are either purely existential or purely universal.

**Example 1 (Solvability Conditions for parametric systems)** *One example for a purely existential question is that for the solvability of a parametric system of differential equations*

$$\psi \equiv f_1(x_1, \ldots, x_n, u_1, \ldots, u_m) = 0 \wedge \cdots \wedge f_k(x_1, \ldots, x_n, u_1, \ldots, u_m) = 0,$$

*where the* $f_1, \ldots, f_k \in \mathbb{Z}\{x_1, \ldots, x_n, u_1, \ldots, u_m\}$ *are differential polynomials. We are interested in conditions on the parameters* $u_1, \ldots, u_m$ *for the solvability of the system w.r.t.* $x_1, \ldots, x_n$*. A corresponding first-order formulation is given by* $\varphi \equiv \exists x_1 \ldots \exists x_n \psi$*.*

For such existential problems, quantifier elimination yields a quantifier-free formula $\hat{\varphi}$ such that for any differentially closed field $\bar{K}$, we have $\bar{K} \models \varphi \longleftrightarrow \hat{\varphi}$. In other words, $\hat{\varphi}$ is a necessary and sufficient condition in the parameters $u_1$, $\ldots$, $u_m$ for the solvability of $\psi$ in the differentially closed field $\bar{K}$. From this point of view we have in particular that $\hat{\varphi}$ is a necessary condition: $\bar{K} \models \varphi \longrightarrow \hat{\varphi}$, alternatively $\bar{K} \models \forall u_1 \ldots \forall u_m(\varphi \longrightarrow \hat{\varphi})$, which can in turn be rewritten as follows:

$$\begin{aligned}
\forall u_1 \ldots \forall u_m(\varphi \longrightarrow \hat{\varphi}) \quad &\Longleftrightarrow \quad \forall u_1 \ldots \forall u_m(\exists x_1 \ldots \exists x_n(\psi) \longrightarrow \hat{\varphi}) \\
&\Longleftrightarrow \quad \forall u_1 \ldots \forall u_m\big(\neg\exists x_1 \ldots \exists x_n(\psi) \vee \hat{\varphi}\big) \\
&\Longleftrightarrow \quad \forall u_1 \ldots \forall u_m\big(\forall x_1 \ldots \forall x_n(\neg\psi) \vee \hat{\varphi}\big) \\
&\Longleftrightarrow \quad \forall u_1 \ldots \forall u_m \forall x_1 \ldots \forall x_n(\neg\psi \vee \hat{\varphi}).
\end{aligned}$$

Note first that the fact that $\hat{\varphi}$ is a *necessary* condition can be expressed as a universal sentence. It thus holds in all subfields of $\bar{K}$, and in particular, in the field actually under consideration by the user.

Applied to our Example 1, the quantifier elimination result $\hat{\varphi}$ will thus in addition to valid choices possibly identify choices of parameters for which the considered system has no solutions in the field actually under consideration. Note second that we nevertheless expect the case distinctions on the parameters made in $\hat{\varphi}$ to be typical for the input problem rather than for the considered differential field. They would then provide a certain structural insight into the problem modeled by the parametric system $\psi$.

The first point is a fact, which we have proved above. The second point though not mathematically precise, can hopefully be substantiated by empirical data in the future.

**Example 2 (Conditions on the solutions of systems)** [1] *As an example for a purely universal question consider the differential equation* $x'^2 + x = 0$*. By*

[1] This very instructive problem has been suggested by E. V. Pankratiev at the MEXMAT faculty of Moscow State University during the author's stay there.

*taking the derivative we obtain*

$$0 = (x'^2 + x)' = 2x'x'' + x' = x'(2x'' + 1).$$

*It is thus necessary for solutions $x$ that $x' = 0$ or $x'' = -1/2$. We can ask for such conditions by means of a universal formula:*

$$\varphi \equiv \forall x \psi, \quad \textit{where} \quad \psi \equiv x'^2 + x = 0 \longrightarrow x' = a \vee x'' = b.$$

*Our procedure delivers $\hat{\varphi} \equiv a = 0 \wedge 2b + 1 = 0$ as a quantifier-free equivalent after automatic simplification.*

As for Example 1 we formally have $\bar{K} \models \varphi \longleftrightarrow \hat{\varphi}$. This time the implication $\bar{K} \models \hat{\varphi} \longrightarrow \varphi$ corresponds to a universal sentence:

$$\begin{aligned} \forall a \forall b(\hat{\varphi} \longrightarrow \varphi) \quad &\Longleftrightarrow \quad \forall a \forall b(\hat{\varphi} \longrightarrow \forall x \psi) \\ &\Longleftrightarrow \quad \forall a \forall b(\neg\hat{\varphi} \vee \forall x \psi) \\ &\Longleftrightarrow \quad \forall a \forall b \forall x(\neg\hat{\varphi} \vee \psi). \end{aligned}$$

That is, the quantifier-free condition $\hat{\varphi}$ on the parameters is *sufficient* for $\varphi$ in all subfields of $\bar{K}$, in particular in the field actually under consideration.

Applying this observation to Example 2 is a bit puzzling at first: We ask for a necessary condition on $a = x'$ and $b = x''$ for being a solution of the considered equation $x'^2 + x = 0$. According to the discussion above we may, however, only conclude that the obtained result is—in any reasonable differential field—a *sufficient* condition on $a = x'$ and $b = x''$ for being a *necessary* condition as required in the formulation of the input formula. We see that, in general, it requires a certain intuition about mathematical logic to deal with the results of our procedure. Here, the situation can be resolved as follows:[2] From $\bar{K} \models \varphi \longleftrightarrow \hat{\varphi}$ it follows that in particular $\bar{K} \models (\varphi \longleftrightarrow \hat{\varphi})\big[a/0, b/-\tfrac{1}{2}\big]$. That is,

$$\bar{K} \models \forall x\Big(x'^2 + x = 0 \longrightarrow x' = 0 \vee x'' = -\frac{1}{2}\Big) \longleftrightarrow \Big(0 = 0 \wedge 2 \cdot \Big(-\frac{1}{2}\Big) + 1 = 0\Big).$$

Equivalently, $\bar{K} \models \forall x\big(x'^2 + x = 0 \longrightarrow x' = 0 \vee x'' = -\frac{1}{2}\big)$, which as a universal formula holds also in the subfield of $\bar{K}$ actually under consideration.

# 4 Online Resources

REDLOG is a package of the computer algebra system REDUCE. For running REDLOG, there is thus a REDUCE license necessary. REDUCE is distributed by two different providers on the basis of correspondingly different implementations of Standard Lisp [29], which are called CSL and PSL. Detailed information on obtaining REDUCE can be found on `www.reduce-algebra.com`.

The REDLOG website is located at `www.redlog.eu`. It provides various services to the user community including

[2] This pragmatic treatment for the considered example has been suggested by A. Seidl at the University of Passau.

- up-to-date system documentation,
- regular system updates as both source tree and precompiled binaries for various architectures.
- the online database REMIS containing REDLOG application examples and literature on the system itself, its theoretical foundations, and its applications.

In the following subsections we address these items one by one.

## 4.1 Documentation

From the very beginning the REDLOG user manual had been written using the GNU Texinfo system. This system was designed to produce from one source printed output using TEX as well as an online documentation for the GNU info system, which is also integrated into the Emacs editor.

Our current policy is to use the Texinfo sources to produce on the one hand PDF for download and on the other hand an HTML version of the content for online browsing. For the HTML version we have decided to produce one single web page for the entire document. This allows use of the search facility of the web browser to find the desired information.

The printed documentation for previous REDLOG releases had been published as technical reports of the University of Passau [7, 10]. Users should be aware that these manuals, the latest of which is now 8 years old, are considerably out of date. The current PDF version on the web provides much more recent documentation in the exactly the same format.

In order to allow users to "quickly-start" without having to read any manual first, we provide one real quantifier elimination example online with a detailed description and a screenshot. This example is exactly the one by Hong, discussed in Section 2.

## 4.2 Updates

Until and including the current release REDUCE 3.8 in 2004, new versions of REDUCE 3 used to appear at least every four years. We had considered it sufficient at that time to simply synchronize REDLOG releases with releases of REDUCE. According to this policy REDLOG 1.0 had been published on the web as a supplement for REDUCE 3.6 in 1996, REDLOG 2.0 had been distributed as part of REDUCE 3.7 in 1999, and REDLOG 3.0 has been distributed as part of REDUCE 3.8 since 2004.

Recently there is a consensus in the developer community to publish new REDUCE releases less frequently. The reason is that a new release actually refers to a revision of the underlying Lisp system. This is an enormous effort for the maintainers, which is not really necessary frequently anymore since the systems have become sufficiently stable over the years. Certainly, REDUCE development

is continuing. The system is maintained and regularly updated using the support for patch files of the respective Lisp systems.

So the idea is to disconnect releases of REDUCE packages from releases of the REDUCE system itself. Our work described here is one step into that direction.

In the download section of the REDLOG website there are, besides the source tree, precompiled versions of the current REDLOG available for the two major REDUCE distributions: CSL-based and PSL-based. For CSL there is a distinction between 32-bit architectures and 64-bit architectures. For PSL, which compiles to native code, we support the most-requested architectures. Users of other architectures are welcome to contact `support@redlog.eu` if they require support for compilation of the sources.

The release numbers of the online updates of REDLOG are constructed from the release dates in such a way that they become monotonic. For instance REDLOG 3.070328 refers to the release of March 28, 2007. From REDLOG 3.061113 on, there is a function `rlabout()`, which displays a banner message including the release number.

## 4.3 REMIS

We have discussed in the previous sections that over the years there have been numerous REDLOG applications documented in the literature. Their number at the time of writing this note is considerably larger than 100. The applications cover eight different domains of computation. This domains are cited widely and possibly reconsidered in follow-up publications. In that case, they are often modified such that variants come into existence, possibly under the same name.

For the sake of reproducibility and coherent research we consider it crucial to organize all this information in a central online database. This database is called REMIS, i.e., REDLOG *example management and information system.* It is accessible via a web frontend on the REDLOG homepage.

There are two types of entries in REMIS: computation examples and publications. With each computation example, we regularly store the following information:

- links to citing papers in the database
- links to related papers in the database
- links to related examples in the database
- REDUCE source files for computing the examples.

With each publication we regularly store the following information:

- a paper classification, which we discuss in more detail below
- links to contained examples in the database
- links to related examples in the database

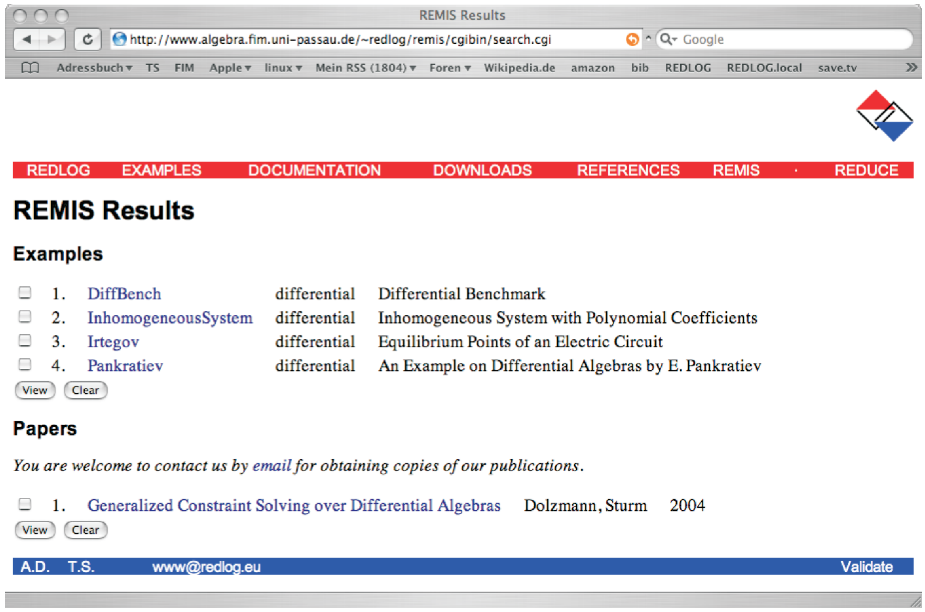



Figure 1: REMIS result page for the search text *differential*

- bibliographical information
- links to other versions (preprint, final, etc.) in the database
- an external link to the most canonical online resource providing the paper (DOI if existent)
- PDF full text to the extent permitted by copyright agreements.

The database design is quite flexible such that arbitrary extra information can be, and in fact is, stored for certain papers and examples whenever this appears reasonable.

REMIS features a keyword search as well as full text search. Figure 1 gives an impression of the result page for the search text *differential*. The names of the examples and of the paper are links to the detailed information pages.

Finally, the REDLOG website offers a comprehensive list of references, which is consequently automatically generated from the REMIS database. It is organized into four different categories of papers: system papers, applications, third-party applications, and theoretical foundations.

# 5 Conclusions

In the first part, we have given an overview of the computer logic system REDLOG, the currently available domains, and its applicability to questions related

to differential equation. In the second part, we have discussed the services available online on the REDLOG website for the user community. Our article is intended to encourage and provide a starting point to researchers concerned with differential equations to consider the integration of quantifier elimination techniques into their set of mathematical tools. Then REDLOG could become a valuable software tool for them as well.

**Acknowledgments.** The majority of the work discussed throughout this paper is joint work with Andreas Dolzmann.

# References

[1] Hirokazu Anai and Volker Weispfenning. Reach set computations using real quantifier elimination. In *Hybrid Systems, Computation and Control (HSCC 2001)*, volume 2034 of *Lecture Notes in Computer Science (LNCS)*. Springer-Verlag, Berlin Heidelberg, 2001.

[2] Lenore Blum. *Generalized Algebraic Structures: A Model Theoretical Approach*. Ph.D. thesis, MIT, Cambridge, MA, 1968.

[3] Lenore Blum. Differentially closed fields: A model theoretic tour. In H. Bass, P. J. Cassidy, and J. Kovacic, editors, *Contributions to Algebra. A Collection of Papers Dedicated to Ellis Kolchin*, pages 37–61. Academic Press, New York, 1977.

[4] Christopher W. Brown, M'hammed El Kahoui, Dominik Novotni, and Andreas Weber. Algorithmic methods for investigating equilibria in epidemic modeling. *Journal of Symbolic Computation*, 41(11):1157–1173, November 2006. Special Issue on the Occasion of Volker Weispfenning's 60th Birthday.

[5] Andreas Dolzmann. Solving geometric problems with real quantifier elimination. In Xiao-Shan Gao, Dongming Wang, and Lu Yang, editors, *Automated Deduction in Geometry*, volume 1669 of *Lecture Notes in Artificial Intelligence (Subseries of LNCS)*, pages 14–29. Springer-Verlag, Berlin Heidelberg, 1999.

[6] Andreas Dolzmann. *Algorithmic Strategies for Applicable Real Quantifier Elimination*. Doctoral dissertation, Department of Mathematics and Computer Science. University of Passau, Germany, D-94030 Passau, Germany, July 2000.

[7] Andreas Dolzmann and Thomas Sturm. Redlog user manual. Technical Report MIP-9616, FMI, Universität Passau, D-94030 Passau, Germany, October 1996. Edition 1.0 for Version 1.0.

[8] Andreas Dolzmann and Thomas Sturm. Guarded expressions in practice. In Wolfgang W. Küchlin, editor, *Proceedings of the 1997 International Symposium on Symbolic and Algebraic Computation (ISSAC 97)*, pages 376–383, Maui, HI, July 1997. ACM, ACM Press, New York, 1997.

[9] Andreas Dolzmann and Thomas Sturm. Redlog: Computer algebra meets computer logic. *ACM SIGSAM Bulletin*, 31(2):2–9, June 1997.

[10] Andreas Dolzmann and Thomas Sturm. Redlog user manual. Technical Report MIP-9905, FMI, Universität Passau, D-94030 Passau, Germany, April 1999. Edition 2.0 for Version 2.0.

[11] Andreas Dolzmann and Thomas Sturm. Generalized constraint solving over differential algebras. In V. G. Ganzha, E. W. Mayr, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing. Proceedings of the CASC 2004*, pages 111–125. Institut für Informatik, Technische Universität München, München, Germany, 2004.

[12] Andreas Dolzmann, Thomas Sturm, and Volker Weispfenning. A new approach for automatic theorem proving in real geometry. *Journal of Automated Reasoning*, 21(3):357–380, 1998.

[13] Andreas Dolzmann, Thomas Sturm, and Volker Weispfenning. Real quantifier elimination in practice. In B. H. Matzat, G.-M. Greuel, and G. Hiss, editors, *Algorithmic Algebra and Number Theory*, pages 221–247. Springer, Berlin, 1998.

[14] Andreas Dolzmann, Thomas Sturm, and Volker Weispfenning. Automatic theorem proving in geometry. In Johannes Grabmeier, Erich Kaltofen, and Volker Weispfenning, editors, *Computer Algebra Handbook*, chapter 3.2.3, pages 201–207. Springer, Berlin, 2003.

[15] Andreas Dolzmann and Volker Weispfenning. Multiple object semilinear motion planning. *Journal of Symbolic Computation*, 42(3):324–337, March 2007.

[16] Armin Größlinger, Martin Griebl, and Christian Lengauer. Quantifier elimination in automatic loop parallelization. *Journal of Symbolic Computation*, 41(11):1206–1221, November 2006.

[17] David Hilbert. Über die vollen Invarianzsysteme. *Mathematische Annalen*, 42:313–373, 1893.

[18] Hoon Hong, Richard Liska, and Stanly Steinberg. Testing stability by quantifier elimination. *Journal of Symbolic Computation*, 24(2):161–187, August 1997. Special issue on applications of quantifier elimination.

[19] Nikolaos Ioakimidis. Finite differences/elements in classical beam problems: Derivation of feasibility conditions under parametric inequality constraints with the help of reduce and redlog. *Computational Mechanics*, 27(2):145–153, February 2001.

[20] Nikolaos I. Ioakimidis. Automatic derivation of positivity conditions inside boundary elements with the help of the REDLOG computer logic package. *Engineering Analysis with Boundary Elements*, 23(10):847–856, December 1999.

[21] Nikolaos I. Ioakimidis. REDLOG-aided derivation of feasibility conditions in applied mechanics and engineering problems under simple inequality constraints. *Journal of Mechanical Engineering (Strojnícky Časopis)*, 50(1):58–69, 1999.

[22] Mats Jirstrand. Nonlinear control system design by quantifier elimination. *Journal of Symbolic Computation*, 24(2):137–152, August 1997. Special issue on applications of quantifier elimination.

[23] M'hammed Kahoui and Andreas Weber. Deciding hopf bifurcations by quantifier elimination in a software component architecture. *Journal of Symbolic Computation*, 30(2):161–179, August 2000.

[24] M'hammed Kahoui and Andreas Weber. Symbolic equilibrium point analysis in parameterized polynomial vector fields. In Victor G. Ganzha, Ernst W. Mayr, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing. Proceedings of the CASC 2002*, pages 71–83. Institut für Informatik, Technische Universität München, München, Germany, 2002.

[25] Gerardo Lafferriere, George J. Pappas, and Sergio Yovine. Symbolic reachability computation for families of lineaar vector fields. *Journal of Symbolic Computation*, 32(3):231–253, 2001.

[26] Aless Lasaruk and Thomas Sturm. Weak quantifier elimination for the full linear theory of the integers. *Applicable Algebra in Engineering Communication and Computing*, 2007. DOI 10.1007/s00200-007-0053-x

[27] Aless Lasaruk and Thomas Sturm. Weak integer quantifier elimination beyond the linear case. In V. G. Ganzha, E. W. Mayr, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing: 10th International Workshop, Proceedings of the CASC 2007*, pages 275–294, volume 4770 of Lecture Notes in Computer Science (LNCS), Springer-Verlag Berlin Heidelberg, 2007.

[28] Rüdiger Loos and Volker Weispfenning. Applying linear quantifier elimination. *The Computer Journal*, 36(5):450–462, 1993. Special issue on computational quantifier elimination.

[29] J. Marti, A. C. Hearn, M. L. Griss, and C. Griss. Standard Lisp report. *SIGSAM Bulletin*, 14(1):23–41, Feb 1980.

[30] Joseph F. Ritt. *Differential Algebra*, volume 33 of *AMS Colloquium Publications*. American Mathematical Society, New York, NY, 1950.

[31] Abraham Robinson. On the concept of a differentially closed field. In H. J. Kreisler, S. Körner, W. A. Luxemburg, and A. D. Young, editors, *Abraham Robinson. Selected Papers*, volume 1, pages 440–455. Yale University Press, New Haven and London, 1979. Originally in Bull. Res. Council Israel 8F (1959), 113–128. M.R. 23 #2323.

[32] Abraham Robinson. Relative model-completeness and the elimination of quantifiers. In H. J. Kreisler, S. Körner, W. A. Luxemburg, and A. D. Young, editors, *Abraham Robinson. Selected Papers*, volume 1, pages 146–159. Yale University Press, New Haven and London, 1979. Originally in Dialectica 12 (1958) 394–407. M.R. 21 #1265.

[33] Gerald E. Sacks. *Saturated Model Theory.* Mathematics Lecture Note Series. W. A. Benjamin, Inc., Reading, MA, 1972.

[34] Abraham Seidenberg. An elimination theory for differential algebra. *University of California Publications in Mathematics. New Series*, 3:31–66, 1956.

[35] Andreas M. Seidl and Thomas Sturm. Boolean quantification in a first-order context. In V. G. Ganzha, E. W. Mayr, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing. Proceedings of the CASC 2003*, pages 329–345. Institut für Informatik, Technische Universität München, München, Germany, 2003.

[36] Werner M. Seiler and Andreas Weber. Deciding ellipticity by quantifier elimination. In V. G. Ganzha, E. W. Mayr, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing. Proceedings of the CASC 2003*, pages 347–355. Institut für Informatik, Technische Universität München, München, Germany, 2003.

[37] Gregor Snelting. Quantifier elimination and information flow control for software security. In Andreas Dolzmann, Andreas Seidl, and Thomas Sturm, editors, *Algorithmic Algebra and Logic. Proceedings of the A3L 2005*, pages 237–242. BoD, Germany, Norderstedt, 2005.

[38] Gregor Snelting, Torsten Robschink, and Jens Krinke. Efficient path conditions in dependence graphs for software safety analysis. *ACM Transactions on Software Engineering and Methodology*, 15(4):410–457, October 2006.

[39] Christian Straßer. Quantifier elimination for queues. In Jan Draisma and Hanspeter Kraft, editors, *Rhine Workshop on Computer Algebra. Proceedings of the RWCA 2006*, pages 239–248. Universität Basel, Basel, 2006.

[40] Thomas Sturm. *Real Quantifier Elimination in Geometry.* Doctoral dissertation, Department of Mathematics and Computer Science. University of Passau, Germany, D-94030 Passau, Germany, December 1999.

[41] Thomas Sturm. Reasoning over networks by symbolic methods. *Applicable Algebra in Engineering, Communication and Computing*, 10(1):79–96, September 1999.

[42] Thomas Sturm. An algebraic approach to offsetting and blending of solids. In V. G. Ganzha, E. W. Mayr, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing. Proceedings of the CASC 2000*, pages 367–382. Springer, Berlin, 2000.

[43] Thomas Sturm. Linear problems in valued fields. *Journal of Symbolic Computation*, 30(2):207–219, August 2000.

[44] Thomas Sturm and Volker Weispfenning. Rounding and blending of solids by a real elimination method. In Achim Sydow, editor, *Proceedings of the 15th IMACS World Congress on Scientific Computation, Modelling, and Applied Mathematics (IMACS 97)*, volume 2, pages 727–732, Berlin, August 1997. IMACS, Wissenschaft & Technik Verlag, Berlin, 1997.

[45] Thomas Sturm and Volker Weispfenning. Computational geometry problems in Redlog. In Dongming Wang, editor, *Automated Deduction in Geometry*, volume 1360 of *Lecture Notes in Artificial Intelligence (Subseries of LNCS)*, pages 58–86. Springer-Verlag, Berlin Heidelberg, 1998.

[46] Thomas Sturm and Volker Weispfenning. Quantifier elimination in term algebras. The case of finite languages. In Victor G. Ganzha, Ernst W. Mayr, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing. Proceedings of the CASC 2002*, pages 285–300. Institut für Informatik, Technische Universität München, München, Germany, 2002.

[47] Andreas Weber. Quantifier elimination on real closed fields and differential equations. In Frieder Haug, Benedikt Löwe, and Torsten Schatz, editors, *Algebra, Logic, Set Theory – Festschrift für Ulrich Felgner zum 65. Geburtstag*, volume 4 of *Studies in Logic*, pages 291–311. College Publications, London, January 2007.

[48] Volker Weispfenning. Semilinear motion planning among moving objects in REDLOG. In V. G. Ganzha and E. W. Mayr, editors, *Computer Algebra in Scientific Computing. Proceedings of the CASC 2001*, pages 541–553. Springer, Berlin, 2001.

[49] Volker Weispfenning. Semilinear motion planning in REDLOG. *Applicable Algebra in Engineering, Communication and Computing*, 12:455–475, June 2001.

# ACTA ACADEMIAE ABOENSIS, SER. B

Mathematica et physica

Mathematics, science, engineering

## Vol. 63 (2003)

1 **Gustafsson, Jan**: The effect of pH and electrolyte concentration on the dispersion properties of titanium dioxide. 54 pp.

## Vol. 64 (2004)

1 **Mylläri, Tatiana**: Studies in the theory of random forests. 88 pp.

## Vol. 65 (2005)

1 **Meriluoto, Jussi and Codd, Geoffrey A.** (editors): TOXIC : cyanobacterial monitoring and cyanotoxin analysis. 149 pp.
2 **Glader, Christer**: Constructive methods for rational interpolation and uniform approximation on the unit disc. 99 pp.
3 **Carlsson, Niclas**: Markov chains on metric spaces : invariant measures and asymptotic behaviour. 28 pp.
4 **Saarela, K-E., Harju, L., Lill, J-O., Heselius, S-J., Rajander, J., Lindroos, A.**: Quantitative elemental analysis of dry-ashed bark and wood samples of birch, spruce and pine from south-western Finland using PIXE. 27 pp.

## Vol. 66 (2006)

1 **Koissi-Kouassi, Marie-Claire**: Modeling mortality rates : improvements, uncertainties and effects on life annuities. 39 pp.
2 **Palmberg, Niklas**: Analytic function spaces : properties of operators and duality. 38 pp.

## Vol. 67 (2007)

1 **Torvela, Taija**: The Sottunga-Jurmo shear zone : structure and deformation history of a crustal-scale ductile shear zone in SW Finland. 187 pp.
2 **Mylläri, A., Edneral, V. and Ourusoff, N.** (editors): CADE – 2007 : computer algebra and differential equations. 191 pp.

Åbo Akademis förlag – Åbo Akademi University Press
Biskopsgatan 13, FI-20500 ÅBO, Finland
Tel. +358 2 215 3292, fax +358 2 215 4490
Distribution: Oy Tibo-Trading Ab
PB 33, FI-21601 PARGAS, Finland
Tel. +358 2 454 9200, fax + 358 2 448 9220