SpringerWienNewYork

Texts and Monographs in Symbolic Computation

A Series of the Research Institute for Symbolic Computation, Johannes Kepler University, Linz, Austria

For further volumes:
http://www.springer.com/series/3073

Lorenzo Robbiano · John Abbott
Editors

Approximate Commutative Algebra

SpringerWienNewYork

Editors
Prof. Lorenzo Robbiano
Università di Genova
Dipartimento di Matematica
Via Dodecaneso, 35
16146 Genova
Italy
robbiano@dima.unige.it

Dr. John Abbott
Università di Genova
Dipartimento di Matematica
Via Dodecaneso, 35
16146 Genova
Italy
abbott@dima.unige.it

SpringerWienNewYork is a part of Springer Science + Business Media
springer.at

Printed on acid-free paper
SPIN 12540785

With 11 Figures

Library of Congress Control Number: 2009934122

ISSN 0943-853X
ISBN 978-3-211-99313-2 e-ISBN 978-3-211-99314-9
DOI 10.1007/978-3-211-99314-9
SpringerWienNewYork

Foreword

What exactly is *Approximate Commutative Algebra*? Where precisely can the approximateness arise? Don't think that it just means

$$xy = 0.9999\,yx$$

and be aware there are certainly some important places where approximation and vagueness are definitely not allowed: *e.g.* in the theorems!

The name ApCoA is an acronym for "Approximate Commutative Algebra". It has received some criticism for its self-contradictory nature: algebra is exact, so it cannot be approximate — but it is for this very same reason that we like it! Our explicit goal is precisely that of building a bridge between the approximate data of the real world and the exact structures of commutative algebra. We believe that the nine papers contained in this volume give an excellent insight into this emerging field of research, and will contribute to the building of this important bridge.

The original stimulus for this book was the first ApCoA workshop hosted in February 2006 by the Radon Institute of Computational and Applied Mathematics (RICAM) of the Austrian Academy of Science and the Research Institute for Symbolic Computation (RISC) of the Johannes Kepler University in Linz, Austria. As interest spread and many new ideas and results sprang up, it quickly became clear that a second ApCoA workshop was warranted. This second workshop was part of the RISC Summer 2008 event, and was again co-organized by RICAM. Most of the articles in this book grew out of the presentations given at this second workshop.

Preface

We have gathered together in this volume nine articles offering highly varied points of view as to what *Approximate Commutative Algebra* (ApCoA) comprises. These diverse perspectives furnish an accessible overview of the current state of research in this burgeoning area. We believe that bringing together these surveys creates a single reference point which will be of benefit both to existing practitioners who wish to expand their horizons, and also to new researchers aspiring to enter this exciting and rapidly developing field. The presentations are intended also to appeal to the interested onlooker who wants to stay informed about recent developments in the field.

The contributions to this book come from active university researchers with a keen interest in ApCoA. Some of them have extensive experience in the field, while others are relative newcomers bringing with them new tools and techniques. The survey articles by their very nature can only scratch the surface, but each one comes with its own bibliography for those who desire to delve more deeply into the numerous topics discussed.

To help the reader orient himself, the paragraphs below summarise the scope of each of the contributed articles. Read and enjoy!

Kreuzer, Poulisse, Robbiano
From Oil Fields to Hilbert Schemes

New techniques for dealing with problems of numerical stability in computations involving multivariate polynomials allow a new approach to real world problems. Using a modelling problem for oil field production optimization as a motivation, the paper presents several recent developments involving border bases of polynomial ideals. To get a deeper understanding for the algebra underlying this approximate world, recent advances concerning border basis and Gröbner basis schemes are discussed. For the reader it will be a long, tortuous, sometimes dangerous, yet hopefully fascinating journey from oil fields to Hilbert schemes.

Bates, Hauenstein, Peterson, Sommese
Numerical Decomposition of the Rank-Deficiency Set of a Matrix of Multivariate Polynomials

Let A be a matrix whose entries are algebraic functions defined on a reduced quasi-projective algebraic set X, *e.g.* multivariate polynomials defined on $X := \mathbb{C}^N$. The sets $S_k(A)$, consisting of $x \in X$ where the rank of the matrix function $A(x)$ is at most k, arise in a variety of contexts: for example, in the description of both the singular locus of an algebraic set and its fine structure; in the description of the degeneracy locus of maps between algebraic sets; and in the computation of the irreducible decomposition of the support of coherent algebraic sheaves, *e.g.* supports of finite modules over polynomial rings. The article presents a numerical algorithm to compute the sets $S_k(A)$ efficiently.

Wu, Reid, Golubitsky
Towards Geometric Completion of Differential Systems by Points

Numerical Algebraic Geometry represents the irreducible components of algebraic varieties over $\mathbb{C}$ by certain points on their components. Such *witness points* are efficiently approximated by Numerical Homotopy Continuation methods, as the intersection of random linear varieties with the components. The paper outlines challenges and progress for extending such ideas to systems of differential polynomials, where prolongation (differentiation) of the equations is required to yield existence criteria for their formal (power series) solutions.

Scott, Reid, Wu, Zhi
Geometric Involutive Bases and Applications to Approximate Commutative Algebra

This article serves to give an introduction to some classical results on Involutive Bases for polynomial systems. Further, it surveys recent developments, including a modification of the above: geometric projected involutive bases, for the treatment of approximate systems, and their application to ideal membership testing and Gröbner basis computation.

Zeng
Regularization and Matrix Computation in Numerical Polynomial Algebra

Numerical polynomial algebra emerges as a growing field of study in recent years with a broad spectrum of applications and many robust algorithms. Among the challenges faced when solving polynomial algebra problems with floating-point arithmetic, the most frequently encountered difficulties include the removal of ill-posedness and the handling of large matrices. This survey develops regularization principles that reformulate the algebraic problems for their well-posed approximate solutions, derives matrix computations arising in numerical polynomial algebra, as well as a subspace strategy that substantially improves the computational efficiency by reducing the matrix sizes. These strategies have been successfully applied to numerical polynomial algebra problems such as GCD, factorization, elimination and determination of multiplicity structure.

Shekhtman

Ideal Interpolation: Translation to and from Algebraic Geometry

This paper discusses four themes that surfaced in multivariate interpolation and which seem to have analogues in algebraic geometry. The hope is that mixing these two areas together will benefit both. In Approximation Theory (AT) the limits of Lagrange projectors correspond to components of the Hilbert scheme of points in Algebraic Geometry (AG). Likewise, error formulas in (AT) may correspond to ideal representations in (AG), and so on.

Riccomagno, Wynn

An Introduction to Regression and Errors in Variables from an Algebraic Viewpoint

There is a need to make a closer connection between classical response surface methods and their experimental design aspects, including optimal design, and algebraic statistics, based on computational algebraic geometry of ideals of points. This is a programme which was initiated by Pistone and Wynn (Biometrika, 1996) and is expanding rapidly. Particular attention is paid to the problem of errors in variables which can be taken as a statistical version of the ApCoA research programme.

Stetter

ApCoA = Embedding Commutative Algebra into Analysis: (my view of computational algebra over $\mathbb{C}$)

This paper deals with the philosophical problem of understanding what ApCoA should mean and, most importantly, what it should do. The main position is that ApCoA comprises consideration of problems of Commutative Algebra over the complex or real numbers, admission of some data of limited accuracy, and use of floating-point arithmetic for the computation of numerical results. In the presence of empirical data, *i.e.* with nearly all computational problems arising from real world applications, the analytic viewpoint is indispensable. The spread of the data may include singular or degenerate situations which would be overlooked if the neighbourhood of a specified problem were neglected.

Kaltofen

Exact Certification in Global Polynomial Optimization Via Rationalizing Sums-Of-Squares

Errors in the coefficients due to floating point round-off or through physical measurement can render exact symbolic algorithms unusable. Hybrid symbolic-numeric algorithms compute minimal deformations of those coefficients that yield non-trivial results, *e.g.* polynomial factorizations or sparse interpolants. The question is: *are the computed approximations the globally nearest to the input?* This paper presents a new alternative to numerical optimization, namely the exact validation via symbolic methods of the global minimality of our deformations.

Acknowledgements Most of the papers in this book grew out of the two ApCoA workshops held at Linz and Hagenberg in Austria. The first workshop was conducted during the *Special Semester on Gröbner Bases* (which lasted from 1st February to 31st July 2006) organized by RICAM (Radon Institute for Computational and Applied Mathematics) of the Austrian Academy of Sciences and RISC (Research Institute for Symbolic Computation) of the Johannes Kepler University, Linz, Austria, under the direction of Professor Bruno Buchberger. The special semester consisted mainly of a series of workshops on various aspects of the theory of Gröbner bases and on important applications of the method.

The second workshop formed part of RISC Summer 2008 event held in the unique and special atmosphere of the Castle of Hagenberg (RISC, Austria). Again RICAM helped organize the meeting. The summer event brought together several important conferences relating to all aspects of symbolic computation — a programme into which ApCoA fitted perfectly.

Contents

Chapter 1
From Oil Fields to Hilbert Schemes

Martin Kreuzer, Hennie Poulisse, and Lorenzo Robbiano

Abstract New techniques for dealing with problems of numerical stability in computations involving multivariate polynomials allow a new approach to real world problems. Using a modelling problem for the optimization of oil production as a motivation, we present several recent developments involving border bases of polynomial ideals. After recalling the foundations of border basis theory in the exact case, we present a number of approximate techniques such as the eigenvalue method for polynomial system solving, the AVI algorithm for computing approximate border bases, and the SOI algorithm for computing stable order ideals. To get a deeper understanding for the algebra underlying this *approximate world*, we present recent advances concerning border basis and Gröbner basis schemes. They are open subschemes of Hilbert schemes and parametrize flat families of border bases and Gröbner bases. For the reader it will be a long, tortuous, sometimes dangerous, and hopefully fascinating journey from oil fields to Hilbert schemes.

Key words: oil field, polynomial system solving, eigenvalue method, Buchberger-Möller algorithm, border basis, approximate algorithm, border basis scheme, Gröbner basis scheme, Hilbert scheme

Martin Kreuzer
Fakultät für Informatik und Mathematik, Universität Passau, D-94030 Passau, Germany,
e-mail: kreuzer@uni-passau.de

Hennie Poulisse
Shell Int. Exploration and Production, Exploratory Research, Kessler Park 1, NL-2288 GD Rijswijk, The Netherlands, e-mail: hennie.poulisse@shell.com

Lorenzo Robbiano
Dipartimento di Matematica, Università di Genova, Via Dodecaneso 35, I-16146 Genova, Italy,
e-mail: robbiano@dima.unige.it

L. Robbiano, J. Abbott (eds.), *Approximate Commutative Algebra*, Texts and Monographs in Symbolic Computation, DOI 10.1007/978-3-211-99314-9_1,

Contents

Introduction

Why did the chicken cross the road?
To boldly go where no chicken has gone before.
(James Tiberius Kirk)

A Bridge Between Two Worlds. Oil fields and Hilbert schemes are connected to very different types of ingredients for algorithmic and algebraic manipulation: continuous and discrete data. This apparent dichotomy occurs already in a single polynomial over the real number field. It consists of a discrete part, the support, and a continuous part, the set of its coefficients. The support is well understood and the source of a large amount of literature in classical algebra. On the other hand, if the coefficients are not exact real numbers but *approximate data*, the very notion of a polynomial and all algebraic structures classically derived from it (such as ideals, free resolutions, Hilbert functions, etc.) tend to acquire a blurred meaning.

An easy example is the following. Consider three distinct non-aligned points in the affine plane over the reals. First of all, if the coordinates are not exact, it is not even clear what we mean by "non-aligned"; a better description might be "far from aligned". The vanishing ideal of the three points is generated by three quadratic polynomials. However, if we change some of the coefficients of these polynomials by a small amount, almost surely we get the unit ideal, since the first two conics still intersect in four points, but the third will almost certainly miss all of them.

How can we cope with this situation? And why should we? The first, easy answer is that approximate coefficients are virtually inevitable when we deal with *real world problems*. In this paper we concentrate on a specific problem where vectors with approximate components encode measurements of physical quantities taken in an **oil field**. Based on actual industrial problems in the field of oil production, we want to popularize the idea that good models of many physical phenomena can be constructed using a *bottom-up* process. The heart of this method is to derive mathematical models by interpolating measured values on a finite set of points. This task can be solved if we know the vanishing ideal of the point set and a suitable vector space basis of its coordinate ring.

This leads us to the next question. Given a zero-dimensional ideal I in a polynomial ring over the reals, if we assume that the coefficients of the generating polynomials are inexact, is it still an ideal? What is the best way of describing this situation? The fact that Gröbner bases are not suitable for computations with inexact data has long been well-known to numerical analysts (see [30]). This is due to the rigid structure imposed by term orderings. Other objects, called *border bases*, behave better. They have emerged as good candidates to complement, and in many cases substitute for, Gröbner bases (see [17], [21], [22], [26], [29]). But possibly the most important breakthrough is the recent discovery of a link between border bases and **Hilbert schemes**. We believe that it may provide a solid mathematical foundation for this new emerging field which tries to combine approximate methods from numerical analysis with exact methods from commutative algebra and algebraic geometry.

You got to be careful if you don't know where you're going
because you might not get there.
(Yogi Berra)

Our Itinerary. In the first part of the introduction we have already suggested the existence of an unexpected bridge between oil fields and Hilbert schemes. Let us now be more specific about the content of the paper and indicate how it tries to build that bridge. Section 1 provides an introduction to one of the main problems arising in oil fields, namely the control of the production. Since we assume that our typical reader is not an expert geologist, we provide some background about the physical nature of an oil reservoir, illustrate the main production problem, and describe a new mathematical approach to solve it. We call it "new", since in our opinion it is very different from the standard view on how to use mathematical models in such a context.

Border bases, the main technical tool we use later, are described in Section 2. This material is mainly taken from [21], Section 6.4 and [17]. We describe the definition and the main properties of border bases and compare them to Gröbner bases using suitable examples. Several important results about border bases are described, in particular their characterization via the commutativity of the formal multiplication matrices due to B. Mourrain (see [26]). A brief excursion is taken into the realm of syzygies, their relation to the border web, and their importance in another fundamental characterization of border bases based on the work of H. Stetter (see [30]).

A useful aspect of border basis theory is that we try to specify a "nice" vector space basis of the quotient ring $\mathbb{R}[x_1,\ldots,x_n]/I$. This sort of basis plays a fundamental role in the problem of solving polynomial systems. Notwithstanding the fact that solving polynomial systems is not a main topic in our presentation, we decided to use Section 3 to give a description of a technique which comes from numerical analysis and uses linear algebra methods, in particular eigenvalues and eigenvectors (see [4], [5], and [9]). The importance of a special kind of matrices, called non-derogatory matrices, is illustrated by Example 1.3.9 and also used in [19] in the context of border basis theory.

Sections 4 and 5 are the computational heart of the paper. They describe two somehow complementary algorithmic approaches to the problem of computing the “approximate vanishing ideal” of a finite set of approximate (empirical) points and a basis of the corresponding quotient ring. In particular, the first part of Section 4 deals with the AVI algorithm and is based on the presentation in [14]. The AVI algorithm makes extensive use of the singular value decomposition (SVD) described in Subsection 4.A and of the stable reduced row echelon form explained in Subsection 4.B. Its main outputs are an order ideal of monomials $\mathcal{O}$ and an approximate $\mathcal{O}$-border basis, a concept introduced in Subsection 4.C. The AVI algorithm is then applied in Subsection 4.D to the concrete construction of polynomial models describing the production of a two-zone oil well.

Section 5 deals with the SOI algorithm which treats the following problem: given a finite set of points $\mathbb{X}$ whose coordinates are given with limited precision, find, if there exists one, an order ideal $\mathcal{O}$ such that the residue classes of its elements form a stable basis of the quotient ring $P/\mathscr{I}(\mathbb{X})$ where $P = \mathbb{R}[x_1,\dots,x_n]$ and $\mathscr{I}(\mathbb{X})$ is the vanishing ideal of $\mathbb{X}$. Here stable means that the residue classes of the elements in $\mathcal{O}$ form a basis of the quotient ring for every small perturbation of the set $\mathbb{X}$. This section summarizes the results of [2]. In Subsection 5.B we describe several easy, but illustrative examples and compare the behaviour of the SOI and the AVI algorithm in these cases. The topic studied in Sections 4 and 5 is an active area of research, and several further approaches have been suggested (see for instance [10] and [25]).

Having done all the *dirty* work (oil fields are not places to be dressed formally), it is time to leave the sedimentary rocks and to look at the problems concerning approximate data from a more general perspective. Polynomials with empirical coefficients can be viewed as *families of polynomials*. So, the next question is whether we can describe families of polynomial ideals algebraically. The answer is yes! The possibility of parametrizing families of schemes by one big scheme is a remarkable feature of algebraic geometry. Hilbert schemes are the most widely known instance of this phenomenon, and consequently they have been studied thoroughly. Moreover, the Hilbert scheme of all zero-dimensional ideals in P of colength s can be covered by affine open subschemes which parametrize all subschemes $\mathrm{Spec}(P/I)$ of the affine space $\mathbb{A}^n_K$ with the property that P/I has a fixed vector space basis. It is interesting to note that the construction of such subschemes is performed using border bases (see for instance [15], [16], and [24]). Also Gröbner bases can be used, since they provide tools for constructing suitable stratifications of Hilbert schemes.

Section 6 is devoted to the explanation of these ideas. Its main sources are the two papers [22] and [28]. In Subsection 6.A we start with an informal explanation of two examples (see Examples 1.6.1 and 1.6.2) which are very easy but nevertheless suitable to illustrate the topic. Then we move to Subsection 6.B where we introduce border basis schemes and their associated border basis families. We show the difficulties of generalizing one of the fundamental tools of Gröbner basis theory to the border basis setting, namely the flat deformation to the leading term ideal. Indeed, the problem is only partially solved and still open in general. The final part of the subsection contains Example 1.6.14 where explicit defining equations are given for one particular border basis scheme, and the connection to the approximate border bases of Section 4 is made.

The final Subsection 6.C is devoted to Gröbner basis schemes and summarizes the presentation in [28]. It is shown that Gröbner basis schemes and their associated universal families can be viewed as weighted projective schemes (see Theorem 1.6.19), a fact that constitutes a remarkable difference between Gröbner and border basis schemes. A comparison between the two types of schemes is given by Theorem 1.6.20 and Corollary 1.6.21, and their equality is examined in Proposition 1.6.24. Throughout the section we highlight the connection between border basis schemes, Gröbner basis schemes, and Hilbert schemes.

At that point the journey from oil fields to Hilbert schemes is over. To get you started with this itinerary, let us point out that, unless specifically stated otherwise, our notation follows the two books [20] and [21]. The algorithms we discuss have been implemented in the computer algebra system CoCoA(see [8]) and in the ApCoCoA library (see [3]).

1.1 A Problem Arising in Industrial Mathematics

Are oil fields commutative?
Are they infinite?
What is their characteristic?
Are they stable?
What are their bases?
(from "The Book of Mathematical Geology")

1.1.A. Oil Fields, Gas Fields and Drilling Wells. Research in relation to oil reservoirs faces many times the same kind of difficulty: the true physical state of an intact, working reservoir cannot be observed. Neither in an experiment of thought, for instance a simulation, nor in a physical experiment using a piece of source rock in a laboratory, the reservoir circumstances can be imitated exactly. This means that the physical laws, *i.e.* the relations between the physical quantities, are not known under actual reservoir circumstances.

To shed some additional light upon this problem, let us have a brief look at oil field formation and exploitation. The uppermost crust of the earth in oil and gas-containing areas is composed of sedimentary rock layers. Since the densities of oil and gas are smaller than the density of water, buoyancy forces them to flow upward through small pores in the reservoir rock. When they encounter a *trap*, *e.g.* a dome or an anticline, they are stopped and concentrated according to their density: the gas is on top and forms the free *gas cap*, the oil goes in the middle, and the (salt) water is at the bottom. To complete the trap, a *caprock*, that is a seal which does not allow fluids to flow through it, must overlie the reservoir rock.

Early drillings had some success because many subsurface traps were leaking. Only by the early 1900s it became known that traps could be located by mapping the rock layers and drilling an exploration well to find a new reservoir. If commercial amounts of oil and gas turn out to be present, a long piece of steel pipe (called the *production tubing*) is lowered into the bore hole and connected to the production facilities.

In a gas well, gas flows to the surface by itself. There exist some oil wells, early in the development of an oil field, in which the oil has enough pressure to flow up the surface. Most oil wells, however, do not have enough pressure and a method called *artificial lift* may then be used. This means that gas is injected into the production tubing of the well. The injected gas mixes with the oil and makes it lighter, thereby reducing the back pressure of the reservoir. On the surface the fluids are transported through long pieces of tubing to a large vessel called *separator* where the three physical phases – oil, water and gas – are separated.

During the exploitation of a reservoir, the pressure of the fluid still in the reservoir drops. This decrease of the reservoir pressure over time is depicted by the *decline curve*. The shape of the decline curve and the total volume of fluid that can be produced from a reservoir (which is called the *ultimate recovery*) depend on the *reservoir drive*, the natural energy that pushes the oil or the gas through the subsurface and into the inflow region of the well. The ultimate recovery of gas from a gas reservoir is often about 80% of the gas in the reservoir. Oil reservoirs are far more variable and less efficient: on average, the ultimate recovery is only 30%. This leaves 70% of the oil remaining in the pressure depleted reservoir which cannot be produced anymore.

Thus, on the most abstract level, the problem we want to address is how to increase the ultimate recovery of an oil reservoir.

1.1.B. Production from Multi-Zone Wells. A well may produce from different parts, called *pockets* or *zones*, of an oil reservoir. The total production of such a well consists of contributions from the different zones. The separate contributions can be controlled by valves, called the *down-hole valves*, which determine the production volume flowing into the well tubing at the locations of the different zones. For such a *multi-zone well*, there may be interactions between the zones in the reservoir. Most certainly, the different contributions will interact with each other when they meet in the common production tubing of the multi-zone well. This situation is called *commingled* production.

In this paper we consider a multi-zone well consisting of **two** producing and interacting zones. Like in a single oil well, the common production flows to the bulk separator where the different phases are separated and the production rates of the separated phases are measured. Besides the phase productions, measurements like pressures, temperatures and injected "lift-gas" are collected; down-hole valves positions are also recorded. A typical set of production variables for a such multi-zone well is:

1. the opening of the valve through which the oil from the first zone is entering the multi-zone well; the opening of the valve is measured in percentages: 0% means that the valve is closed; 100% means that the valve is completely open;
2. the opening of the valve through which the oil from the second zone is entering the multi-zone well;

3. the pressure difference over the down-hole valve of the second zone which is a measure for the inflow from the reservoir into the well at the valve position; if the valve is closed we assume this value to be zero;
4. the pressure difference over the down-hole valve of the first zone when the valve in that zone is open; if the valve is closed we assume this value to be zero;
5. the volume of gas produced simultaneously with the oil;
6. the pressure difference between the inflow locations in the production tubing;
7. the pressure difference which drives the oil through the transportation tubing.

One might be tempted to think that the total oil production of a multi-zone well is the sum of the productions of each zone when producing separately. This is in any case the current state of the art, where the total production is regressed against the separate productions, that is the total production is written as a linear combination of the separate productions. The coefficients in this linear sum are called *reconciliation factors*. The oil produced by one of the zones may push back the oil which tries to flow into the well at the other zone. Likewise, the gas which is produced simultaneously with the oil may have stimulating or inhibiting effects on the inflow of the oil with respect to the situation of single zone productions. With reference to the remarks above, this behavior does not sound very linear. Indeed, in Section 4.D we will use our algebraic approach in a two-zone well example to demonstrate that the total production is not a linear combination of the separate productions. We believe that the reason of the (usually) low ultimate recovery of a multi-zone well is due to the fact that the interactions among the different producing zones are unknown.

This leads us to a first concretization of the problem we want to study: find a model for the total production of an oil well which takes the interactions into account and describes the behavior correctly on longer time scales.

1.1.C. Algebraization of the Production Problem. Before plunging into the creation of an algebraic setting for the described production problem, let us spend a few words on why we believe that approximate computational algebra is an appropriate method to deal with it.

The available data correspond to a finite set of points $\mathbb{X}$ in $\mathbb{R}^n$. Their coordinates are *noisy* measurements of physical quantities associated with the well: pressures, oil and gas production, valve positions, etc. These points represent the behavior of the well under various production conditions. The combination of the contribution of the individual zones to the total production is a *sum* which has to be corrected by taking into account the effect of the interactions. As in many other situations (for instance, in statistics), the interactions are related to *products* of the collected data series. Many of the known physical laws and model equations are of a polynomial nature. And even if they are not, some elementary insights into the system (*e.g.* that the result depends exponentially on a certain data series) allow us to prepare the data series appropriately (*e.g.* by computing their logarithms). Consequently, the starting point for us is the polynomial ring $P = \mathbb{R}[x_1, \ldots, x_n]$.

In the following we will deal with the case of a two-zone well. The production situation is depicted schematically in Figure 1.1. The notation ΔP refers to pressure differences.

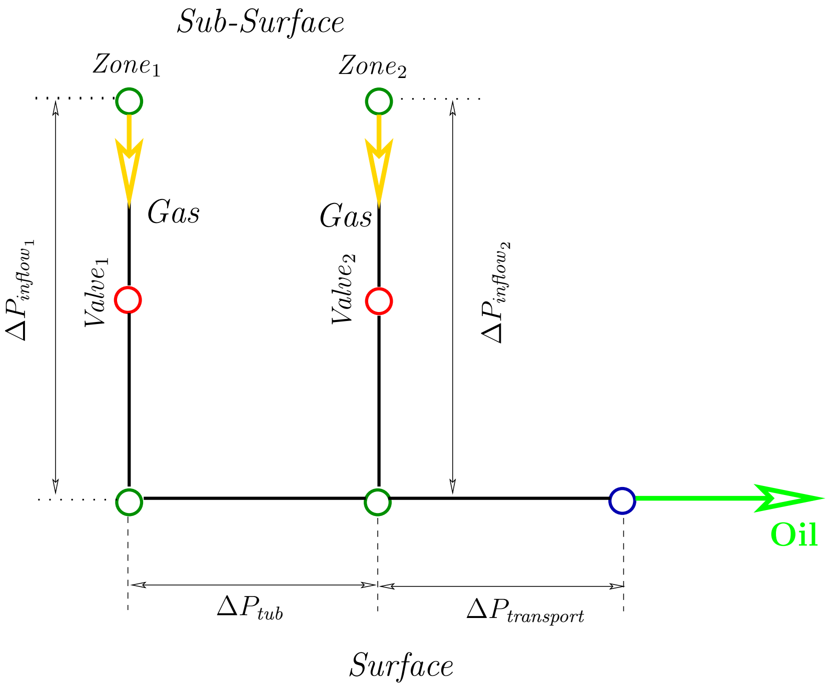

Fig. 1.1 Schematic representation of a two-zone well.

The valves indicated in this figure are used to influence the inflow of the fluids at the two locations into the production tubing of the well. If a valve is closed, there is no inflow from the reservoir at the location of the valve. If the valve is open, the inflow depends on the valve opening and the interactions with the fluids which enter the well through the other inflow opening. In particular, a valve in open position does not imply that there is inflow from the reservoir into the well at its location.

Next we try to formulate the problems associated with this production system more explicitly. Notice that the reservoir is a very special physical system in that it is not possible to check "how it works" using a computer simulation experiment or a physical model laboratory experiment. Traditional modelling techniques assume that equations which describe the flow of the fluids through the reservoir are available. Their limited success is in our view due to the fact that there is no proper representation of the interactions occurring in the production situation. Without these, actions taken to influence the production may have devastating consequences in that the "wrong" effects are stimulated. It is fair to state that the existing low ultimate recovery rates are to a large extent caused by the fact that the interactions in production units have not been acknowledged properly.

As a starting point, let us formulate the production problem in intuitive rather than in precise mathematical terms.

Problem 1. Assume that no *a priori* model is available to describe the production of the two-zone well of Figure 1.1 in terms of measurable physical quantities which determine the production. Find an algebraic model of the production in terms of the determining, measurable physical quantities which specifically models the interactions occurring in this production unit.

Now let us phrase this problem using the polynomial ring $P = \mathbb{R}[x_1,\dots,x_n]$. The first step is to associate the indeterminates x_i with physical quantities in the production problem in the sense that when the indeterminate x_i is evaluated at the points of $\mathbb{X}$, the evaluations are the measurements of the physical quantity associated to x_i. In the sequel we use $n = 5$ and the following associations, where the physical quantities are the ones referenced in Figure 1.1.

$$
\begin{aligned}
x_1 &: \Delta P_{inflow_1}\\
x_2 &: \Delta P_{inflow_2}\\
x_3 &: \textit{Gas production}\\
x_4 &: \Delta P_{tub}\\
x_5 &: \Delta P_{transport}
\end{aligned}
$$

Table 1.1 Physical interpretation of the indeterminates.

Note that we have not listed an indeterminate associated to the oil production. The explanation for this is that the physical quantities listed in the above table may all be interpreted as *driving forces* for the oil production. For the pressure differences ΔP this is clear. But it holds also for the gas production. When a large amount of gas is produced in the deeper parts of the reservoir, it disperses in the fluid mixture, makes it lighter, and in this way stimulates oil production through this lifting process. Thus the physical quantities listed in the above table may all be viewed as the *causing* quantities, or *inputs*, and the oil production is their *effect*, or *output*. So, basically we make the following crucial assumption.

Assumption. *There exists a causal relationship between the production and the driving forces. Using suitable inputs, this causal relationship is of polynomial nature.*

Denoting the production by f, the algebraic translation of the causal relationship assumption is $f \in \mathbb{R}[x_1,\dots,x_5]$ where the indeterminates x_i are labeled as in the above table. That is, the production is not associated with an indeterminate, but with a polynomial, and the production measurements are the evaluations of this polynomial over the set $\mathbb{X}$. Hence the statement of Problem 1 can be reformulated as follows.

Problem 2. Find the polynomial $f \in \mathbb{R}[x_1,\dots,x_5]$, using only the evaluations $\mathbb{X}$ of the quantities x_i and the evaluations of f!

The information registered in the set $\mathbb{X}$ refers to the situation where at most one of the valves is closed. The only possible inflows from the reservoir into the production tubing of the two-zone well are at the location of Zone 1, or of Zone 2, or both. Moreover, in all three situations data have been collected at different valve openings. Furthermore, in order for the data in $\mathbb{X}$ to deserve the qualification *driving forces*, some pre-processing has been applied: with reference to Figure 1.1, if $valve_1$ is closed, it may very well be that the pressure difference ΔP_{inflow_1} is not zero, but it does not have the meaning of a driving force over the valve opening because there is no flow over the valve. Hence in the data set $\mathbb{X}$, we set ΔP_{inflow_1} to zero for this situation. Of course, we do the same for $valve_2$ with respect to ΔP_{inflow_2}. Finally, if the valve associated with the deepest zone $valve_1$ is closed, there is no transport of fluids in the lowest part of the production tubing of the well. That is, for ΔP_{tub} really to have the significance of a driving force, it is set to zero if $valve_1$ is closed.

Notice also that all data are based on measurements, *i.e.* they may contain measurement errors. Consequently, we can only expect that the desired polynomial f vanishes *approximately* at the points of $\mathbb{X}$. In Section 4 we will return to this instance of the production problem and solve it with the methods we are going to present.

1.2 Border Bases

Ideally, inside the border
there is order.
(Three anonymous authors)

1.2.A. Motivation and Definition. The problems considered in the previous section lead us to study zero-dimensional ideals in $P = K[x_1,\dots,x_n]$ where K is a field. The two most common ways to describe such an ideal I are by either providing a special system of generators (for instance, a Gröbner basis) of I or by finding a vector space basis $\mathcal{O}$ of P/I and the matrices of the multiplications by the indeterminates with respect to $\mathcal{O}$. One possibility to follow the second approach is to use $\mathcal{O} = \mathbb{T}^n \setminus \mathrm{LT}_\sigma(I)$, the complement of a leading term ideal of I. By Macaulay's Basis Theorem, such a set $\mathcal{O}$ is a K-basis of P/I. Are there other suitable sets $\mathcal{O}$?

A natural choice is to look for sets of terms. We need to fix how a term b_j in the *border* $\partial\mathcal{O} = (x_1\mathcal{O} \cup \cdots \cup x_n\mathcal{O}) \setminus \mathcal{O}$ of $\mathcal{O}$ is rewritten as a linear combination of the terms in $\mathcal{O}$. Thus, for every $b_j \in \partial\mathcal{O}$, a polynomial of the form

$$g_j = b_j - \sum_{i=1}^{\mu} c_{ij} t_i$$

with $c_{ij} \in K$ and $t_i \in \mathcal{O}$ should be contained in I. Moreover, we would not like that $x_k g_j \in I$. Hence we want $x_k b_j \notin \mathcal{O}$. Therefore the set $\mathbb{T}^n \setminus \mathcal{O}$ should be a monoideal. Consequently, $\mathcal{O}$ should be an *order ideal*, that is it should be closed under forming divisors. Let us formulate precise definitions.

Definition 1.2.1. Let $\mathcal{O}$ be a finite set of terms in $\mathbb{T}^n$.

a) The set $\mathcal{O}$ is called an **order ideal** if $t \in \mathcal{O}$ and $t' \mid t$ implies $t' \in \mathcal{O}$.
b) Let $\mathcal{O}$ be an order ideal. The set $\partial\mathcal{O} = (x_1\mathcal{O} \cup \cdots \cup x_n\mathcal{O}) \setminus \mathcal{O}$ is called the **border** of $\mathcal{O}$.
c) Let $\mathcal{O} = \{t_1, \ldots, t_\mu\}$ be an order ideal and $\partial\mathcal{O} = \{b_1, \ldots, b_\nu\}$ its border. A set of polynomials $\{g_1, \ldots, g_\nu\} \subset I$ of the form

$$g_j = b_j - \sum_{i=1}^{\mu} c_{ij} t_i$$

with $c_{ij} \in K$ and $t_i \in \mathcal{O}$ is called an $\mathcal{O}$**-border prebasis** of I.
d) An $\mathcal{O}$-border prebasis of I is called an $\mathcal{O}$**-border basis** of I if the residue classes of the terms in $\mathcal{O}$ are a K-vector space basis of P/I.

The following example will be used frequently throughout this paper.

Example 1.2.2. In the ring $P = \mathbb{R}[x, y]$, consider the ideal $I = (f_1, f_2)$ where

$$f_1 = \tfrac{1}{4}x^2 + y^2 - 1$$
$$f_2 = x^2 + \tfrac{1}{4}y^2 - 1$$

The zero set of I in $\mathbb{A}^2(\mathbb{R})$ consists of the four points $\mathbb{X} = \{(\pm\sqrt{0.8}, \pm\sqrt{0.8})\}$. This setting is illustrated in Figure 1.2.

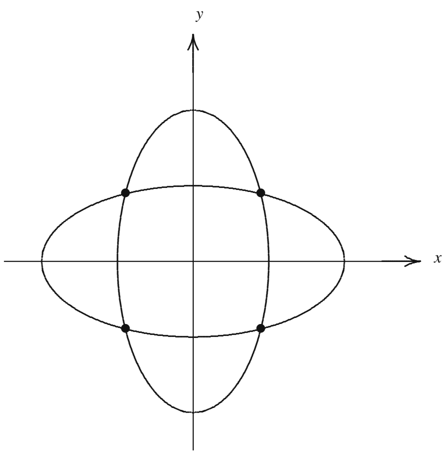

Fig. 1.2 Two ellipses intersecting in four points.

We use $\sigma = \texttt{DegRevLex}$ and compute $\mathrm{LT}_\sigma(I) = (x^2, y^2)$. Thus the order ideal $\mathcal{O} = \{1, x, y, xy\}$ represents a basis of P/I. Its border is $\partial\mathcal{O} = \{x^2, x^2y, xy^2, y^2\}$. The following figure illustrates the order ideal $\mathcal{O}$ and its border.

An $\mathcal{O}$-border basis of I is given by $G = \{g_1, g_2, g_3, g_4\}$ where