SCUOLA NORMALE SUPERIORE

Tesi di Perfezionamento in Matematica per le Tecnologie Industriali

Applications of Algebra in the Oil Industry

CANDIDATO: Dott.ssa Maria-Laura Torrente
RELATORE: Prof. Lorenzo Robbiano

Anno accademico 2008-09

Contents

List of Figures

List of Tables

Introduction

Often numerical data in scientific computing arise from real-world measurements, and so are perturbed by noise, uncertainty and approximation. Recently some attempts have been made to describe them using multivariate polynomial models, which, in the numerical world, mix two apparently inconsistent types of data: the continuous and the discrete. This apparent inconsistency is already clear when we deal with a single polynomial over the reals. It consists of a discrete part, the support, and a continuous part, the set of its coefficients. The support is a well-known concept in classical algebra and is defined as the set of monomials having a non-zero coefficient. The set of coefficients inherits the continuity from the space to which it belongs, that is the field of real numbers; if the polynomial coefficients are not exact, but derive from the numerical world, the very notion of polynomial acquires a blurred meaning. In the same vein, any other structure based on polynomials, such as an ideal, which involves real data loses its rigorous algebraic nature. From a computational point of view, it cannot be handled using exact methods, since the sophisticated tools provided by Computer Algebra can no longer be applied, while the general-purpose techniques provided by Numerical Analysis generally return unsatisfactory "local" results. Since the late 1990s a new field of investigation is emerging. It has been given different names, the most important ones being *Numerical Polynomial Algebra*, introduced by Hans J. Stetter, *Approximate Commutative Algebra* (ApCoA), introduced by L. Robbiano, and *Numerical Algebraic Geometry*, introduced by A. Sommese.

The motivation for our work comes from a problem of modeling oil production. It arose within the *Algebraic Oil Project*, an international cooperation between Shell International Exploration & Production (Rijswijk, The Netherlands), the Department of Mathematics of the University of Genova (Italy), and the Faculty of Informatics and Mathematics of the University of Passau (Germany). The optimization of oil production is obviously one of the main problems in oil industry: increasing the ultimate recovery of an oil or gas reservoir is, up to now, the most challenging problem in the extraction operations. An oil reservoir is a very special physical system which is nearly impossible to study using a computer simulation or physical laboratory experiments. Traditional modeling techniques assume that equations which describe the flow of the fluids through the reservoir are available. However their limited success suggests that they do not provide a good representation of the interactions occuring dur-

ing the production phase. The current low ultimate recovery rates may well be caused by the fact that the interactions between the production units are still unknown, so that actions taken to increase the extraction could end up inhibiting rather than stimulating the production. Our aim is to find a model for the total production of a group of wells or a collection of zones which describes the production behaviour correctly over longer time scales; the idea is to take into account the interactions between the zones and so to provide a decomposition of the total production as a combination of the separate contributions of the individual wells.

Rather than starting from the physical knowledge of the phenomenon, we want to strengthen the idea that good models for many industrial problems can be constructed using a *bottom-up* process, in which the mathematical model is derived by interpolating the measured values at a finite set of points. In the exact case the problem of interpolation can be solved using the vanishing ideal of the set of points, that is the ideal comprising all polynomials which vanish at the given points, and a suitable vector space basis of its coordinate ring. Notice that the vanishing ideal enables us to find all the equivalent models, since it embodies all the relationships among the points. Classically, it is computed using the Buchberger-Möller Algorithm [BM82], a low complexity method which returns a Gröbner basis of the ideal. Unfortunately, as already stated, the data coming from real-world experiments are always affected by noise, they are imprecise, or known with limited accuracy. Our aim is to solve the interpolation problem in the approximate case.

We formalize the concept of approximate datum by using the notion of *empirical point* already introduced by Stetter [Ste04]. We view the measurements coming from a finite number of real-life experiments as empirical points of an affine space: we assume that each experimental datum consists of the measurements of n different physical quantities, so that it can be encoded as a point of K^n, where K is a field usually equal to $\mathbb{R}$; the whole empirical data set can thus be organized as a finite set of points of K^n. Note that since each point corresponds to a single test, and each coordinate to a single measurement, the error affecting each component of the point mostly derives from the limits of accuracy of the measuring instruments. For this reason, we suppose we know the tolerances on the empirical data, that is the absolute error on each data coordinate. Naturally, the tolerance in one coordinate may differ from that of another coordinate. But we do require that the tolerance in a given coordinate be the same for all the data points (*e.g.* identical instruments were used to obtain the measurements in that coordinate). More formally, we require that there is a common tolerance vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ valid for all the input points. Given the tolerance ε and a data point $p \in K^n$, we view the pair (p, ε) as an empirical point representing a "cloud" of data which differ from p by less than the tolerance. Any point lying inside this cloud can be considered equivalent to the measurement p from a numerical point of view, and is called an *admissible perturbation* of the point p.

Often in the experimental tests a way to bound the inaccuracy affecting the data consists of registering many times the value of the same datum. This

leads to deal with large body of experimental data, often characterized by a high level of redundancy. To reduce the volume of data we find a good way of "thinning out" such large sets before using the symbolic-numeric approach. The preprocessing technique is based on the idea of reducing "redundancy" in the original data: we regard subsets of original points which lie close to each other as repeated measurements, and replace them by a single representative value. If the intersection of different empirical points (intuitively represented as clouds) is "sufficiently large", we can replace them by a single empirical point carrying essentially the same empirical information. Based on this idea of clustering together empirical points we design three algorithms which thin out a large set of redundant data to produce a smaller set of "equivalent" empirical points. Naturally, the degree of the reduction depends on how much redundancy is present in the original data.

Having introduced a good mathematical formalization for an empirical data set, and having defined a useful technique for reducing the redundancy, we face the problem of approximate interpolation (that is interpolation in the approximate case). As in the exact case, the main task consists of computing the "vanishing ideal" of a finite set of empirical points, where we put the expression vanishing ideal in inverted commas as we must clarify what we mean by it. Indeed, there exist different approaches to this problem: among them we mention the works of H. M. Möller et al. [MS00], T. Sauer [Sau07], D. Heldt et al. [HKPP06], M. Kreuzer et al. [KPR08], and C. Fassino [Fas08].

Our point of view is the following: in order to emphasize the numerical equivalence among all the feasible and small perturbations of the original set of points, we provide a common characterization of their vanishing ideals, and, when possible, we compute a numerically stable representation of them. The idea is the following. We let $P = K[x_1, \ldots, x_n]$ be the polynomial ring in the indeterminates $x_1, \ldots, x_n$ over K, $\mathbb{X}$ a finite set of distinct points of K^n, and $\mathcal{I}(\mathbb{X}) \subseteq P$ its vanishing ideal; we let ε be the common tolerance on the points of $\mathbb{X}$. If $\widetilde{\mathbb{X}}$ is another set of points of K^n, each differing by less than the uncertainty from the corresponding element of $\mathbb{X}$, then $\widetilde{\mathbb{X}}$ is called an *admissible perturbation* of $\mathbb{X}$ and (intuitively) the two sets can be considered as equivalent. Nevertheless, given two distinct admissible perturbations $\widetilde{\mathbb{X}}_1$ and $\widetilde{\mathbb{X}}_2$ of $\mathbb{X}$, it can happen that their vanishing ideals $\mathcal{I}(\widetilde{\mathbb{X}}_1)$ and $\mathcal{I}(\widetilde{\mathbb{X}}_2)$ have very different bases: this is a well known phenomenon when using Gröbner basis theory (see [KKR05] and [KR00]) which turns out to be unsuitable as a numerical tool. To overcome this drawback and to highlight the numerical equivalence among all the admissible perturbations $\widetilde{\mathbb{X}}$ of $\mathbb{X}$, we focus our attention on determining an order ideal $\mathcal{O}$ such that the residue classes of its elements form a vector space basis of the quotient ring $P/\mathcal{I}(\widetilde{\mathbb{X}})$, for any perturbation $\widetilde{\mathbb{X}}$. Such an $\mathcal{O}$ is called a *stable quotient basis* for the empirical set $(\mathbb{X}, \varepsilon)$. Now, suppose that $\#\mathcal{O} = \#\mathbb{X}$, and let $M_{\mathcal{O}}(\mathbb{X})$ be the matrix whose rows are the images of the terms $t \in \mathcal{O}$ under the evaluation map at $\mathbb{X}$. A necessary and sufficient condition for $\mathcal{O}$ to be a (monomial) basis of the vector space $P/\mathcal{I}(\mathbb{X})$ is that $M_{\mathcal{O}}(\mathbb{X})$ is a non-singular matrix, that is its determinant is not zero. Since the determinant is a continu-

ous function in the matrix entries, this defines an open condition, and explains the choice of the adjective stable. Notice that stable quotient bases provide a common characterization of the ideals $\mathcal{I}(\mathbb{X})$ and $\mathcal{I}(\widetilde{\mathbb{X}})$: they highlight the geometrical properties of the empirical set $(\mathbb{X}, \varepsilon)$ and, via the border basis theory, guarantee the existence of a structurally stable representation of $\mathcal{I}(\mathbb{X})$.

Border bases appeared for the first time in connection with problems arising in numerical analysis during the 1980s, thanks to the work of Hans J. Stetter (see [AS88] and [Ste04]); then, during the 1990s, the importance of these results for computer algebra was pointed out by H. Michael Möller (see [MS95] and [Möl93]). In 1999 the first algebraic properties of border bases were presented by B. Mourrain (see [Mou99]). In 2005, A. Kehrein, M. Kreuzer and L. Robbiano wrote a survey devoted to laying the algebraic foundations of the border basis theory for zero-dimensional ideals (see [KKR05] and [KR05]). Recently, M. Kreuzer and L. Robbiano (see [KR08]), and later L. Robbiano (see [Rob08]) examined a natural link between border bases and Hilbert schemes which provides a further improvement to the solid mathematical foundations of the border basis theory.

Our decision to use border bases for describing the vanishing ideal $\mathcal{I}(\mathbb{X})$ is based on two main reasons: firstly, the works mentioned above certify border bases as a good tool for dealing with numerical problems; secondly, border bases are easy to compute since, once a basis $\mathcal{O}$ of the quotient ring $P/\mathcal{I}(\mathbb{X})$ is fixed, the corresponding border basis can be obtained by simple combinatoric and linear algebra computations. The border basis $\mathcal{B}$ of $\mathcal{I}(\mathbb{X})$ built upon a stable quotient basis $\mathcal{O}$ is also called *stable*. Indeed $\mathcal{B}$ exhibits good numerical behaviour: any other $\mathcal{O}$-border basis $\widetilde{\mathcal{B}}$ of $\mathcal{I}(\widetilde{\mathbb{X}})$ can be obtained by means of a small and continuous variation of the coefficients of the polynomials of $\mathcal{B}$, while the supports remain unchanged.

Notice that stable quotient bases (and consequently stable border bases) do not always exist (see Example 5.1.5, Chapter 5): in such cases we exploit the wider notion of *stable order ideal*. Though stable order ideals do not define a monomial basis of the vector space $P/\mathcal{I}(\mathbb{X})$, they nevertheless provide $P/\mathcal{I}(\widetilde{\mathbb{X}})$ with a common structure for any admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}$, and give information on the geometrical configuration of the original points.

The method we present for computing a stable order ideal of an empirical point set is essentially based on the Buchberger-Möller (BM) Algorithm; we generalize it to the numerical frame by introducing a new notion of *numerical linear dependence* of vectors. A practical implementation of this theoretical method is given by the Stable Order Ideal (SOI) Algorithm which is included in CoCoALib [CoC]. The practical approach is based on a first order error analysis of the problem, since, from our point of view, the interest is focussed on small perturbations of the original set of points. In order to investigate the stability of the order ideal, the SOI Algorithm uses a parametric description of the admissible perturbations of the points and uses some results on the first order approximation of rational functions, so that the check of numerical linear dependence of vectors can be greatly simplified. If the cardinality of the stable

order ideal $\mathcal{O}$ equals the cardinality of the set of input points, then $\mathcal{O}$ is a stable quotient basis for the ideal of points, the corresponding $\mathcal{O}$-border basis $\mathcal{B}$ of $\mathcal{I}(\mathbb{X})$ exists and is stable w.r.t. the input tolerance ε. To determine $\mathcal{B}$ it suffices to find the border of $\mathcal{O}$, a simple combinatorical computation, and then solve a full rank linear system for each element of its border $\partial\mathcal{O}$.

In conclusion, the thesis originated from a very specific problem in industrial mathematics, namely the optimization of oil field production. The approach taken here is new in the sense that we discard the idea of using simulation; instead, we try to recover good relations among the physical quantities involved starting from experimental measurements. We prefer a bottom-up procedure over the classical top-down methods. This choice led us to use tools which are more typical of commutative algebra and algebraic geometry, therefore we had to somehow navigate uncharted waters. However, some successes in the applications of these methods are currently encouraging us to continue this line of research.

We divide the thesis into three main parts: in the first part, which consists of Chapters 1 and 2, we recall the background material; in the second part, which is made up of Chapters 3, 4, and 5, we present a frame to deal with approximate data and a new method to tackle the problem of interpolation in the approximate case; finally, in the third part, that is in Chapter 6, we apply this new method to an actual industrial problem arising in oil fields, namely the control of the production of oil.

Based on the specialised texts [KR00], [KR05], and [CLO92], in Chapter 1 we recall the general notation along with some definitions and results of algebraic-geometry theory useful for the topics treated in the thesis. In Chapter 2 we introduce and characterize the concepts of affine point set, the most important mathematical object of this thesis, and its vanishing ideal; we also describe the Buchberger-Möller Algorithm [BM82], a classical and efficient method for computing the vanishing ideal of a finite set of points. We generalize this result in two different ways: in one case we simply replace the operation of row reduction used in it by the least squares method; in the other one we perform a similar variation but with the aim of computing a border basis of the ideal of points. In Chapter 3 we define a formal framework for dealing with indetermination in $\mathbb{R}^n$. In particular, we introduce the basic definition of empirical point and discuss the analogies with the definition given by Stetter [Ste04]; we give the definitions of empirical vector and empirical evaluation vector; we introduce the notion of numerical linear dependence, and adapt it to the empirical evaluation case. Based on [AFT07], in Chapter 4 we describe a method to reduce the "redundancy" in the empirical data set; we present two algorithms, the Agglomerative Algorithm and the Divisive Algorithm, both included in CoCoALib (a GPL C++ library for doing Computation in Commutative Algebra, see [CoC]), which thin out sets of empirical points while preserving their overall geometrical structure. Some numerical examples to illustrate the behaviour of the algorithms on different geometrical configurations of points are also presented. Based on [AFT08], in Chapter 5 we present a symbolic-numeric method to characterize the vanishing

ideal of an empirical set of points. In particular, we introduce the notions of stable order ideal, stable quotient basis and stable border basis of a finite set of empirical points. Then, we describe a theoretical method and a practical algorithm, the SOI Algorithm, available in CoCoALib, for their computation. A section with numerical examples to show the effectiveness of the presented results is also included. In Chapter 6 we address one of the most important problems arising in the oil fields, namely the control of the oil production. We treat it in the case of a multi-zone well by making the crucial assumption of the existence of a causal relationship between the production of oil and a set of variables having a special physical meaning. The results refer to the case of a two-zone well: the redundancy of the set of numerical data coming from it is reduced by applying the techniques of Chapter 4; the new set of empirical points is then used for the SOI computations. Starting from these results, we compute different polynomials for the production, we test their reliability and compare their prediction skills.

Acknowledgements

PhD studies are an important stepping stone for any potential future researcher; in this period a dynamic and stimulating environment plays a crucial role in the process of intellectual growth. I'm indebted to several people and institutions which created such an environment for me throughout my PhD period (since January 2005).

First and foremost I would like to thank my advisor, Prof. Lorenzo Robbiano. His farsightedness set me on the right track with the Algebraic Oil Project, to which I gave my contribution through this thesis. His continuous support and his clear thoughts on the subject have been of inestimable value; his guidance and knowledge essential for the fruitful elaboration of this work.

I thank Claudia Fassino and John Abbott, who have been at the same time collaborators, teachers and friends – a rare and precious mixture. This thesis would not have been achieved without their aid and collaboration; their unceasing support, their careful reading and correcting of my writings, their continual suggestions on the topic have been absolutely fundamental. Our long discussions in a friendly atmosphere played an important role in the pleasant side of doing research.

I would like to thank Hennie Poulisse, promoter and leader of the Algebraic Oil Project in Rijswijk, for his intellectual brilliance, his mathematical intuition, and his optimistic attitude, from which I learnt a lot. My special thanks to Mihaela Popoviciu, for her precious remarks, for our invaluable discussions, and last but not least, for our nice friendship. I thank both Hennie and Mihaela for their warm hospitality during my internship at Shell Exploration & Production, in Rijswijk.

I gratefully acknowledge the support of the Algebraic Oil Project provided through the Stichting Shell Research Foundation in the form of two internships. I am also grateful to SIEP B.V. for their permission to use the genuine indus-

trial data in Section 6. My warmest thanks go to Scuola Normale Superiore, in particular to Prof. Fulvio Ricci, for giving me the opportunity to tackle a problem arising in a new and developing area of mathematics.

I would like to express my sincere gratitude to the other members of the CoCoA Team, Anna Bigatti and Massimo Caboara, for their support and their faith in my abilities. My sincere gratitude goes to all the friends I met during my PhD period, and in particular to my PhD colleagues and friends at Scuola Normale Superiore in Pisa: Giovanni, Lorenzo, Lucia, Rosario, Salvatore, Valeria. They each helped make my time in the PhD program more fun and interesting.

Finally, I would like to thank all my friends, my parents, my sister Elisabetta and my husband Manuel for instilling in me confidence and a drive for pursuing my PhD; I would never finished without their constant support and encouragement.

Chapter 1

Preliminaries on polynomial algebra

In this chapter we introduce the tools of algebraic-geometry theory that we use in the thesis. Almost all the definitions and results we present are taken from the specialised texts [KR00], [KR05], and [CLO92] to which the interested reader is referred for a deeper understanding of the topics.

1.1 Algebraic foundations

Throughout this thesis we will mainly work with multivariate polynomials and rational functions. This section is devoted to introducing their definitions and properties (see Section 1.1.1 and 1.1.2). We start by recalling the definition of the basic algebraic structures.

Definition 1.1.1. A **monoid** $(S, \cdot)$ (or simply S if no ambiguity can arise) is defined to be a set S together with an operation $\cdot : S \times S \to S$ which is associative and for which there exists an identity element, *i.e.* an element $1_S \in S$ such that $1_S \cdot s = s \cdot 1_S = s$ for all $s \in S$. A monoid is called **commutative** if $s \cdot s' = s' \cdot s$ for all $s, s' \in S$. Furthermore, if in S every element is invertible, *i.e.* for all $s \in S$ there exists an element $s' \in S$ which satisfies $s \cdot s' = s' \cdot s = 1_S$, then S is called **group**.

By a **ring** $(R, +, \cdot)$ (or simply R) we shall always mean a **commutative** ring with identity element, *i.e.* a set R together with two associative operations $+, \cdot : R \times R \to R$ such that $(R, +)$ is a commutative group with identity element 0_R, such that $(R, \cdot)$ is a commutative monoid with identity element 1_R, and such that the distributive laws are satisfied. If $1_R = 0_R$ then R is called the **trivial ring** as it contains the single element 0_R. A **field** is a ring K such that $(K \setminus \{0_K\}, \cdot)$ is a group.

For the rest of this chapter we let R be a ring. Some elements of a ring have special properties. For instance, if $r \in R$ and $rr' = 0$ implies $r' = 0$ for all

$r' \in R$, then r is called a **non zero-divisor**. A ring whose non-zero elements are non zero-divisors is called an **integral domain**. For instance, every field is an integral domain.

We recall the concept of a ring homomorphism.

Definition 1.1.2. Let R, S be rings. A map $\varphi : R \to S$ is called a **ring homomorphism** if $\varphi(1_R) = 1_S$ and for all elements $r, r' \in R$ we have $\varphi(r+r') = \varphi(r) + \varphi(r')$ and $\varphi(r \cdot r') = \varphi(r) \cdot \varphi(r')$, *i.e.* if φ preserves the ring operations.

Sometimes a field and a group are tied together by an operation of the field on the group to produce the very well known algebraic structure of **vector space**. In this case the elements of the fields are usually called **scalars**, the elements of the group are called **vectors**, and the operation is called **scalar multiplication**.

The following definition introduces the very important notion of ideal.

Definition 1.1.3. Let R be a ring; a subset $I \subseteq R$ is called an **ideal** of R if it is an additive subgroup of R and $R \cdot I = \{r \cdot i \mid r \in R,\ i \in I\} \subseteq I$.

Note that, given any ideal I in a ring R, we can form the residue class ring R/I in the obvious way; the canonical map $R \to R/I$ is a ring homomorphism. Now, consider the following definition.

Definition 1.1.4. Let I be an ideal of R.

(a) A set $\{m_\lambda \mid \lambda \in \Lambda\}$ of elements of I is called a **system of generators** of I if every $m \in I$ has a representation $m = r_1 m_{\lambda_1} + \ldots + r_n m_{\lambda_n}$ for some $n \in \mathbb{N}$, $r_1, \ldots, r_n \in R$ and $\lambda_1, \ldots, \lambda_n \in \Lambda$. In this case we write $I = \langle m_\lambda \mid \lambda \in \Lambda \rangle$.

(b) The ideal I is called **finitely generated** if it has a finite system of generators. If I is generated by a single element, it is called a **principal ideal**.

1.1.1 Polynomial rings

Let R be a ring; as in [KR00] we define multivariate polynomial rings over R recursively, starting from the notion of univariate polynomial ring. For this reason we recall here its definition. We consider the set $R^{(\mathbb{N})}$ of all sequences $(r_0, r_1, \ldots)$ of elements $r_0, r_1, \ldots \in R$ such that $r_i \neq 0$ for only finitely many indices $i \geq 0$. We let $e_i = (0, \ldots, 0, 1, 0, 0, \ldots)$ be the element of $R^{(\mathbb{N})}$ having 1 at position $i+1$. Obviously, every element of $R^{(\mathbb{N})}$ has a unique representation $(r_0, r_1, \ldots) = \sum_{i \in \mathbb{N}} r_i e_i$. Given two elements $\sum_{i \in \mathbb{N}} r_i e_i$ and $\sum_{i \in \mathbb{N}} s_i e_i$, we define

$$\left(\sum_{i \in \mathbb{N}} r_i e_i \right) \cdot \left(\sum_{i \in \mathbb{N}} s_i e_i \right) = \sum_{i \in \mathbb{N}} \left(\sum_{j=0}^{i} r_j s_{i-j} \right) e_i$$

It is easy to check that the set $R^{(\mathbb{N})}$, together with componentwise addition and the multiplication defined above, is a commutative ring with identity e_0; further, $e_i = (e_1)^i$ for all $i \in \mathbb{N}$.

Definition 1.1.5. Let R be a ring; let $R^{(\mathbb{N})}$ be equipped with the ring structure defined above.

(a) Let x be a symbol and $x = e_1$; the ring $R^{(\mathbb{N})}$ is called the **polynomial ring in the indeterminate** x **over** R and is denoted by $R[x]$. It is a commutative ring and every element of $R[x]$ has a unique representation $\sum_{i\in\mathbb{N}} r_i x^i$ with $r_i \in R$ and $r_i \neq 0$ for only finitely many indices $i \in \mathbb{N}$.

(b) Let $n \geq 2$ and $x_1, \ldots, x_n$ be symbols; we recursively define $R[x_1, \ldots, x_n] = (R[x_1, \ldots, x_{n-1}])[x_n]$ and call it the **polynomial ring in** n **indeterminates over** R.

(c) The elements of a polynomial ring are called **polynomials**. Polynomials in one indeterminate are often called **univariate polynomials**, while polynomials in several indeterminates are called **multivariate polynomials**.

Some properties of a ring are inherited by polynomial rings over it, as shown in the following proposition.

Proposition 1.1.6. *Let R be an integral domain.*

(a) The units in $R[x_1, \ldots, x_n]$ are the units in R.

(b) The polynomial ring $R[x_1, \ldots, x_n]$ is an integral domain.

Proof. See Proposition 1.1.9 in [KR00]. □

Among the different properties of a polynomial ring we recall the Universal Property, which claims that the ring homomorphisms starting from it are uniquely defined by the images of the indeterminates, and those images may be chosen freely.

Proposition 1.1.7. (Universal Property of the Polynomial Ring) *Let R, S be rings and let $\varphi : R \to S$ be a ring homomorphism; let $n \geq 1$, and $s_1, \ldots, s_n$ be elements in S. Then there exists a unique ring homomorphism $\psi : R[x_1, \ldots, x_n] \to S$ such that $\psi|_R = \varphi$ and $\psi(x_i) = s_i$ for $i = 1, \ldots, n$.*

Proof. See Proposition 1.1.12 in [KR00]. □

A ring homomorphism ψ defined as above is also called an **evaluation homomorphism**; we write the image $\psi(f)$ of f as $f(s_1, \ldots, s_n)$ and call it the **evaluation of** f **at** $(s_1, \ldots, s_n)$.

We shall use the following compact and unique representation of a multivariate polynomial $f \in R[x_1, \ldots, x_n]$:

$$f = \sum_{\alpha\in\mathbb{N}^n} c_\alpha x^\alpha \tag{1.1}$$

where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ and $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, and where only finitely many elements $c_\alpha \in R$ are different from zero. We give the following definitions.

Definition 1.1.8. Let $n \geq 1$.

(a) A polynomial $f \in R[x_1, \ldots, x_n]$ of the form $f = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ such that $(\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ is called a **term** (or **power product**). The set of all terms of $R[x_1, \ldots, x_n]$ is denoted by $\mathbb{T}^n$.

(b) For a term $t = x_1^{\alpha_1} \cdots x_n^{\alpha_n} \in \mathbb{T}^n$, the number $\deg(t) = \alpha_1 + \ldots + \alpha_n$ is called the **degree** of t.

(c) The map $\log : \mathbb{T}^n \to \mathbb{N}^n$ defined by $x_1^{\alpha_1} \cdots x_n^{\alpha_n} \mapsto (\alpha_1, \ldots, \alpha_n)$ is called the **logarithm**.

Note that the set $\mathbb{T}^n$ is a commutative monoid whose identity element is $1 = x_1^0 \cdots x_n^0$. It is easy to prove that the map $\log : \mathbb{T}^n \to \mathbb{N}^n$ is an isomorphism of monoids, and the monoid $\mathbb{T}^n$ does not depend on the ring of coefficients R.

Definition 1.1.9. Let $n \geq 1$ and let $f = \sum_{\alpha \in \mathbb{N}^n} c_\alpha x^\alpha \in R[x_1, \ldots, x_n]$ be a polynomial.

(a) For every $\alpha \in \mathbb{N}^n$ the element $c_\alpha \in R$ is called the **coefficient** of the term x^α in f.

(b) The set $\{x^\alpha \in \mathbb{T}^n \mid c_\alpha \neq 0\}$ is called the **support** of f and denoted by $\mathrm{Supp}(f)$. In particular $\mathrm{Supp}(0) = \emptyset$.

(c) If $f \neq 0$, the number $\max\{\deg(x^\alpha) \mid x^\alpha \in \mathrm{Supp}(f)\}$ is called the **degree** of f and denoted by $\deg(f)$. The degree of the zero polynomial is not defined.

For example, let $f \in \mathbb{Q}[x_1, x_2, x_3]$ be the polynomial

$$f \;=\; x_1x_2x_3 - \frac{3}{5}x_1^3 + 5x_2^3 - 4x_3^3 + x_1x_2 - 5x_1 - 7x_2 + 15$$

The support of f is

$$\mathrm{Supp}(f) \;=\; \{x_1^3,\ x_2^3,\ x_3^3,\ x_1x_2,\ x_1,\ x_2,\ 1\}$$

and consists of 7 terms. The terms in $\mathrm{Supp}(f)$ have been ordered by decreasing degree; the sequence of degrees is $3, 3, 3, 2, 1, 1, 0$. However, this is not enough to order them completely since there are several terms with the same degree. Complete orderings on $\mathbb{T}^n$ will be introduced in Section 1.2.1.

Consider the following definition.

Definition 1.1.10. Let $n, k \geq 1$, and let $f_1, \ldots, f_k \in R[x_1, \ldots, x_n]$; then we define the set

$$\langle f_1, \ldots, f_k \rangle = \left\{ \sum_{i=1}^{k} h_i f_i \mid h_1, \ldots, h_k \in R[x_1, \ldots, x_n] \right\}$$

The crucial fact is that $\langle f_1, \ldots, f_k \rangle$ is an ideal in $R[x_1, \ldots, x_n]$ (a proof is given in Lemma 3, Chapter 1.4, in [CLO92]) which is called the **ideal generated by** $f_1, \ldots, f_k$.

We recall Definition 1.1.4 and specify it further in the case of polynomial ideals. We say that a polynomial ideal $I \subseteq R[x_1, \ldots, x_n]$ is **finitely generated** if there exist $f_1, \ldots, f_k \in R[x_1, \ldots, x_n]$ such that $I = \langle f_1, \ldots, f_k \rangle$; then the set $\{f_1, \ldots, f_k\}$ is called a **system of generators** or a **basis** of I.

In this respect, there is a nice analogy between polynomial algebra and linear algebra. Let $P = K[x_1, \ldots, x_n]$ be a polynomial ring over the field K; the definition of ideal of P is similar to the definition of vector subspace over K; further, the ideal generated by a set of polynomials $f_1, \ldots, f_k \in P$ is similar to the span of a finite number of vectors $v_1, \ldots, v_k$, since the linear combinations are built up using field coefficients for the span, and polynomial coefficients for the ideal generated. In contrast to vector spaces, a basis of a polynomial ideal does not need to meet the condition of linear independence. This fact causes difficulties in extending the univariate division algorithm to the multivariate case (see Section 1.2.2) and, consequently, implies that a polynomial in an ideal could be expressed as a P-linear combination of the basis elements in different ways. Nevertheless, for polynomial rings over fields, the following fundamental result holds.

Theorem 1.1.11. (Hilbert's Basis Theorem)
Let $n \geq 1$, and let K be a field. Every ideal $I \subseteq K[x_1, \ldots, x_n]$ is finitely generated, that is $I = \langle f_1, \ldots, f_k \rangle$ for some $f_1, \ldots, f_k \in I$.

Proof. See Theorem 4, Section 5, Chapter 2 in [CLO92]. □

Hilbert's Basis Theorem guarantees that any ideal $I \subseteq K[x_1, \ldots, x_n]$ has a finite system of generators. Nevertheless I may have many different bases: in Section 1.2.3 we will define a very useful type of basis, namely a *Gröbner basis*, which rapidly became, since its invention, a fundamental tool for modern algebra, both for its theoretical and practical consequences. Then in Section 1.3.4 and only for zero-dimensional ideals we will introduce the notion of *border basis*, a useful generalization of Gröbner bases to tackle problems arising in numerical analysis.

1.1.2 Fields of rational functions

It is well-known that the ring of integers $\mathbb{Z}$ can be embedded in any field of characteristic 0, the "smallest" being the field of rational numbers $\mathbb{Q}$, since $\mathbb{Q}$ is formed by the fractions $\frac{m}{n}$, where $m \in \mathbb{Z}$, $n \in \mathbb{Z} \setminus \{0\}$. This construction can be generalized to any integral domain R.

Proposition 1.1.12. *Let R be an integral domain. We consider the set of pairs $\{(r, s) \mid r, s \in R \text{ and } s \neq 0\}$. For two such pairs $(r, s), (r', s')$, we let $(r, s) \sim (r', s')$ if and only if $rs' - r's = 0$.*

(a) The relation $\sim$ is an equivalence relation.

(b) Let us denote the set of all equivalence classes by $Q(R)$ and the equivalence class of a pair (r, s) by $\frac{r}{s}$. Then the rules

$$\frac{r}{s} + \frac{r'}{s'} = \frac{s'r + sr'}{ss'} \quad \textit{and} \quad \frac{r}{s} \cdot \frac{r'}{s'} = \frac{rr'}{ss'}$$

for all $r, r' \in R$, and for all $s, s' \in R \setminus \{0\}$ make $Q(R)$ into a field.

(c) The map $R \to Q(R)$ defined by $r \to \frac{r}{1}$ is a ring homomorphism.

Proof. See Proposition 3.5.2 in [KR00]. □

Definition 1.1.13. Let R be an integral domain, and let $Q(R)$ be the set defined in Proposition 1.1.12; then $Q(R)$ is called the **quotient field**, or **field of fractions** of R.

Let K be an arbitrary field, let $x_1, \ldots, x_n$ be indeterminates, and $P = K[x_1, \ldots, x_n]$ be a polynomial ring. A well known example of the previous construction is given by the **field of rational functions**

$$K(x_1, \ldots, x_n) = Q(P) = \left\{ \frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)} \ : \ f, g \in K[x_1, \ldots, x_n], g \neq 0 \right\}$$

which is the field of fractions of the polynomial ring P.

1.2 Introduction to Gröbner bases

1.2.1 Term orderings and leading terms

In Section 1.1.1 we hinted at the problem of how to arrange the elements of the support of a polynomial: here, we consider in detail this problem.

Using the recursive definition of multivariate polynomials, we see that the way of writing the terms in the support depends on the univariate case, and thus on how $\mathbb{T}^1$ is ordered. There is no unique way to do it. For instance, if we look at the univariate polynomial $f(x) = 1 + 3x - 2x^3$ we see that there are 6 different representations of f which are related to the 6 different ways of ordering the three elements in $\mathrm{Supp}(f) = \{1, x, x^2\}$, namely $1 + 3x - 2x^3$, $1 - 2x^3 + 3x$, $3x + 1 - 2x^3$, $3x - 2x^3 + 1$, $-2x^3 + 1 + 3x$ and $-2x^3 + 3x + 1$. Indeed, there is a technical reason which validates only the first and the last representations. Suppose we want to multiply $f(x)$ by x^2, say. After termwise multiplication, the rule continues to hold and we do not have to reorder the result, which on the contrary does not happen in the other cases. This leads to an extra property that an ordering of terms should have, the property of being compatible with multiplication. In a more technical setting we require that the total ordering on $\mathbb{T}^1$ makes it into an ordered monoid. For instance, the specification $1 < x$ implies $x < x^2 < x^3$, and so on; for $\mathbb{T}^1$ only two possible

orderings are left, the one described by $1 < x$ and the one described by $x < 1$. This is enough for the univariate polynomials and also for the multivariate ones, if a recursive representation is used.

Nevertheless the problem of how to order terms in $\mathbb{T}^n$ (when $n > 1$) cannot be solved using the previous arguments, since, for instance, it is not well defined if the polynomial $f(x_1, x_2, x_3) = x_1x_3 + 3x_2^2$ should be written as $x_1x_3 + 3x_2^2$ or rather as $3x_2^2 + x_1x_3$. Thus total orderings on $\mathbb{T}^n$ are introduced and discussed in this section.

Definition 1.2.1. Let $(\Gamma, \circ)$ be a commutative monoid. A **relation** σ on Γ is a subset of $\Gamma \times \Gamma$. If a pair (γ_1, γ_2) is in that subset, we shall write $\gamma_1 \geq_\sigma \gamma_2$. A relation σ on Γ is called **complete** if any two elements $\gamma_1, \gamma_2 \in \Gamma$ are comparable, *i.e.* if we have $\gamma_1 \geq_\sigma \gamma_2$ or $\gamma_2 \geq_\sigma \gamma_1$. A complete relation σ on Γ is called a **monoid ordering** if the following conditions are satisfied for all $\gamma_1, \gamma_2, \gamma_3 \in \Gamma$.

(a) $\gamma_1 \geq_\sigma \gamma_1$

(b) $\gamma_1 \geq_\sigma \gamma_2$ and $\gamma_2 \geq_\sigma \gamma_1$ imply $\gamma_1 = \gamma_2$

(c) $\gamma_1 \geq_\sigma \gamma_2$ and $\gamma_2 \geq_\sigma \gamma_3$ imply $\gamma_1 \geq_\sigma \gamma_3$

(d) $\gamma_1 \geq_\sigma \gamma_2$ implies $\gamma_1 \circ \gamma_3 \geq_\sigma \gamma_2 \circ \gamma_3$

If, in addition, we have

(e) $\gamma_1 \geq_\sigma 1_\Gamma$ for all $\gamma \in \Gamma$

then σ is called a **term ordering** on Γ.

If σ is a relation on Γ, and if $\gamma_1, \gamma_2 \in \Gamma$ are such that $\gamma_1 \geq_\sigma \gamma_2$, we also write $\gamma_2 \leq_\sigma \gamma_1$.

Let $n \geq 1$, and $\mathbb{T}^n$ be the set of all terms in the indeterminates $x_1, \ldots, x_n$. Under the isomorphism of monoids $\log : \mathbb{T}^n \to \mathbb{N}^n$, term orderings on $\mathbb{T}^n$ correspond $1-1$ to term orderings on $\mathbb{N}^n$. We recall some of the most important term orderings on $\mathbb{T}^n$.

Definition 1.2.2. Let $t_1, t_2 \in \mathbb{T}^n$. We say that

(a) $t_1 \geq_{\texttt{Lex}} t_2$ if and only if the first non-zero component of $\log(t_1) - \log(t_2)$ is positive or $t_1 = t_2$. The relation $\geq_{\texttt{Lex}}$ defines a term ordering on $\mathbb{T}^n$ which is called the **lexicographic term ordering** and is denoted by `Lex`.

(b) $t_1 \geq_{\texttt{DegLex}} t_2$ if we have $\deg(t_1) > \deg(t_2)$, or if $\deg(t_1) = \deg(t_2)$ and $t_1 \geq_{\texttt{Lex}} t_2$. The relation $\geq_{\texttt{DegLex}}$ defines a term ordering on $\mathbb{T}^n$ which is called the **degree-lexicographic term ordering** and is denoted by `DegLex`.

(c) $t_1 \geq_{\texttt{DegRevLex}} t_2$ if we have $\deg(t_1) > \deg(t_2)$, or if $\deg(t_1) = \deg(t_2)$ and the last non-zero component of $\log(t_1) - \log(t_2)$ is negative, or $t_1 = t_2$. The relation $\geq_{\texttt{DegRevLex}}$ defines a term ordering on $\mathbb{T}^n$ which is called the **degree-reverse-lexicographic term ordering** and is denoted by `DegRevLex`.

Example 1.2.3. Let $x_1^2x_3^2$, $x_1x_2^2x_3$, and x_2^5 be power products in $\mathbb{T}^3$.

Using `Lex`, the indeterminates $x_1, \ldots, x_n$ are ordered decreasingly, *i.e.* we have $x_1 >_{\texttt{Lex}} \ldots >_{\texttt{Lex}} x_n$. For instance, we have $x_1^2x_3^2 >_{\texttt{Lex}} x_2^5$ and $x_1^2x_3^2 >_{\texttt{Lex}} x_1x_2^2x_3$, since both $(2,0,2) - (0,5,0) = (2,-5,2)$ and $(2,0,2) - (1,2,1) = (1,-2,1)$ have a positive first component.

Using `DegLex`, we see that the indeterminates are again ordered decreasingly: $x_1 >_{\texttt{DegLex}} \ldots >_{\texttt{DegLex}} x_n$. But we have a different ordering on the terms: $x_2^5 >_{\texttt{DegLex}} x_1^2x_3^2$, since $\deg(x_2^5) = 5 > 4 = \deg(x_1^2x_3^2)$, and we have $x_1^2x_3^2 >_{\texttt{DegLex}} x_1x_2^2x_3^2$, since $\deg(x_1^2x_3^2) = 4 = \deg(x_1x_2^2x_3)$ and $(2,0,2) - (1,2,1) = (1,-2,1)$ has a positive first component.

Once again, using `DegRevLex`, the indeterminates $x_1, \ldots, x_n$ are ordered decreasingly. In this case we have $x_2^5 >_{\texttt{DegRevLex}} x_1^2x_3^2$, since $\deg(x_2^5) = 5 > 4 = \deg(x_1^2x_3^2)$, and we have $x_1x_2^2x_3 >_{\texttt{DegRevLex}} x_1^2x_3^2$, since both terms have degree 4 and $(1,2,1) - (2,0,2) = (-1,2,-1)$ has a negative last component.

Definition 1.2.4. A monoid ordering σ on $\mathbb{T}^n$ is called **degree compatible** if $t_1 \geq_\sigma t_2$ for $t_1, t_2 \in \mathbb{T}^n$ implies $\deg(t_1) \geq \deg(t_2)$.

Note that the term orderings `DegLex` and `DegRevLex` are degree compatible.

Once a term ordering on $\mathbb{T}^n$ is chosen, each non-zero polynomial f can be represented in a unique way as a sum according to the sorted list of terms in $\mathrm{Supp}(f)$. The hierarchy created in $\mathrm{Supp}(f)$ implies the existence of a power product which is "bigger" and "more important" than the other terms. The last part of this section is devoted to introducing and explaining this notion and related concepts, since one of the main ideas of Computational Commutative Algebra is to study or detect properties of ideals using information coming from these mathematical objects. In what follows, we let R be a ring, $n \geq 1$, $P = R[x_1, \ldots, x_n]$ be a polynomial ring, and σ be a term ordering on $\mathbb{T}^n$. For each polynomial $f \in P \setminus \{0\}$ we shall use the following unique representation.

$$f = \sum_{i=1}^{k} c_i t_i \tag{1.2}$$

where $c_i \in R \setminus \{0\}$, and where $t_1, \ldots, t_s \in \mathbb{T}^n$ are such that $t_1 >_\sigma \ldots >_\sigma t_s$.

Definition 1.2.5. For a non-zero element $f \in P$, let $f = \sum_{i=1}^{k} c_i t_i$ be the representation according to (1.2).

(a) The term $\mathrm{LT}_\sigma(f) = t_1 \in \mathbb{T}^n$ is called the **leading term** of f w.r.t. σ.

(b) The element $\mathrm{LC}_\sigma(f) = c_1 \in R \setminus \{0\}$ is called the **leading coefficient** of f w.r.t. σ. If $\mathrm{LC}_\sigma(f) = 1$, we say that f is **monic**.

(c) The element $\mathrm{LM}_\sigma(f) = \mathrm{LC}_\sigma(f) \cdot \mathrm{LT}_\sigma(f) = c_1t_1$ is called the **leading monomial** of f w.r.t. σ.

Note that for the zero polynomial f these concepts are not defined.

Definition 1.2.6. Let $I \subseteq P$ be an ideal.

(a) The ideal $\mathrm{LT}_\sigma(I) = \langle \mathrm{LT}_\sigma(f) \mid f \in I \setminus \{0\} \rangle$ is called the **leading term ideal** of I with respect to σ.

(b) The set $\{\mathrm{LT}_\sigma(f) \mid f \in I \setminus \{0\}\} \subseteq \mathbb{T}^n$ will be denoted by $\mathrm{LT}_\sigma\{I\}$.

Note that, for $I = \langle 0 \rangle$, we get $\mathrm{LT}_\sigma(I) = \langle 0 \rangle$ and $\mathrm{LT}_\sigma\{I\} = \emptyset$, using the above definitions. If $f_1, \ldots, f_k \in P \setminus \{0\}$, and if $I = \langle f_1, \ldots, f_k \rangle$ is the ideal generated by them, we have $\langle \mathrm{LT}_\sigma(f_1), \ldots, \mathrm{LT}_\sigma(f_k) \rangle \subseteq \mathrm{LT}_\sigma(I)$. The following example shows that this can be a proper inclusion.

Example 1.2.7. Let K be a field, and I be the ideal in $K[x, y]$ generated by $f_1 = x - 1$ and $f_2 = xy - 3$; let $\sigma = \texttt{DegLex}$. Then $f = y \cdot f_1 - 1 \cdot f_2 = -y + 3 \in I$ implies $\mathrm{LT}_\sigma(f) = y \in \mathrm{LT}_\sigma(I)$, but y does not belong to the ideal generated by $\mathrm{LT}_\sigma(f_1) = x$ and $\mathrm{LT}_\sigma(f_2) = xy$.

Nevertheless, there are sets of polynomials of I whose leading terms generate $\mathrm{LT}_\sigma(I)$ as the next proposition shows.

Proposition 1.2.8. *Let $I \subseteq P$ be an ideal. Then there exist $f_1, \ldots, f_k \in I$ such that we have $\mathrm{LT}_\sigma(I) = \langle \mathrm{LT}_\sigma(f_1), \ldots, \mathrm{LT}_\sigma(f_k) \rangle$.*

Proof. See Proposition 1.5.6. in [KR00]. □

In Example 1.2.7 we saw that for $f = y \cdot f_1 - 1 \cdot f_2 = -y + 3$ and $\sigma = \texttt{DegLex}$ the leading monomials of the two summands cancel out, so that y, the leading term of the result, is smaller than the leading terms of the summands. This example shows that some generators have a special behaviour with respect to the leading terms of the elements they generate. More precisely, we have that $y = \mathrm{LT}_\sigma(f) \notin \langle \mathrm{LT}_\sigma(f_1), \mathrm{LT}_\sigma(f_1) \rangle = \langle x, xy \rangle$. However, according to Proposition 1.2.8, there exists another set of generators of $\langle f_1, f_2 \rangle$ whose leading terms generate the leading term ideal. This is a simple example showing that *not all systems of generators of an ideal are equal*: some systems of generators are "more" special than others, as will be evident in Section 1.2.3 once the very important notion of Gröbner basis will be introduced.

In order to point out the importance of the set $\mathrm{LT}_\sigma\{I\}$ we consider the following example.

Example 1.2.9. Let $I \subseteq \mathbb{Q}[x, y, z]$ be the ideal generated by $f_1 = 2x - y$ and $f_2 = y^2 + 2$, and let $\sigma = \texttt{DegLex}$. We assert that $\mathrm{LT}_\sigma(I) = \langle \mathrm{LT}_\sigma(f_1), \mathrm{LT}_\sigma(f_2) \rangle = \langle x, y^2 \rangle$; it follows that the set $\mathrm{LT}_\sigma\{I\}$ contains all the terms which are multiples of x or y^2. Consider the residue class ring $\mathbb{Q}[x, y, z]/I$: it is clear that it can be viewed as a $\mathbb{Q}$-vector space. Note that a basis of $\mathbb{Q}[x, y, z]/I$ is given by the residue classes of the infinite set $\{1, y, z, yz, z^2, yz^2, z^3, \ldots\}$ which is equal to the set $\mathbb{T}^3 \setminus \mathrm{LT}_\sigma\{I\}$.

The conclusion of Example 1.2.9 is indeed an example of a fundamental result in Computational Commutative Algebra, known as Macaulay's Basis Theorem.

Theorem 1.2.10. (Macaulay's Basis Theorem)
Let K be a field, let $P = K[x_1, \ldots, x_n]$ be a polynomial ring over K, let $I \subseteq P$ be an ideal, and let σ be a term ordering on $\mathbb{T}^n$. We denote the set of all terms in $\mathbb{T}^n \setminus \mathrm{LT}_\sigma\{I\}$ by B. Then the residue classes of the elements of B form a basis of the K-vector space P/I.

Proof. See Theorem 1.5.7 in [KR00]. □

Macaulay's Basis Theorem gives us a first idea of how to compute effectively in P/I. If, for some term ordering σ, $\mathrm{LT}_\sigma(I)$ is known, the theorem guarantees that it is possible to represent every element of P/I uniquely as a finite linear combination of the residue classes of the elements of $\mathbb{T}^n \setminus \mathrm{LT}_\sigma\{I\}$.

1.2.2 Division algorithm and rewrite rules

Macaulay's Basis Theorem is the first step towards being able to compute in residue class rings P/I, where $P = K[x_1, \ldots, x_n]$ is a polynomial ring over a field K, σ is a term ordering on $\mathbb{T}^n$, and $I \subseteq P$ is a non-zero ideal. One problem that has still to be addressed is the lack of an effective procedure for writing each residue class as a linear combination of the residue classes of the terms contained in $\mathbb{T}^n \setminus \mathrm{LT}_\sigma\{I\}$. For univariate polynomials the answer is given by the division with remainder algorithm, as shown in the following example.

Example 1.2.11. In the case $K = \mathbb{Q}$, $n = 1$, and $P = K[x_1] = K[x]$, consider the polynomials $f = 3x^4 + 2x^2 + x + 1$ and $g = x^2 + 1$, and let $I = \langle g \rangle \subseteq P$. It is easy to prove that $f \equiv x + 2 \mod I$. If we apply to f and g the division with remainder procedure we obtain:

$$\begin{array}{llll|ll}
3x^4 & +2x^2 & +x & +1 & x^2 & +1 \\
\cline{5-6}
3x^4 & +3x^2 & & & 3x^2 & \\
\cline{1-2}
 & -x^2 & +x & +1 & & \\
 & -x^2 & & -1 & & -1 \\
\cline{2-4}
 & & +x & +2 & 3x^2 & -1
\end{array}$$

In other words we have $f = qg + r$ where $q = 3x^2 - 1$ and $r = x + 2$, and where the characteristic property of the remainder r is $\deg(r) < \deg(g)$.

The univariate polynomial ring $K[x]$ over a field K is a **principal ideal domain**, that is it has the property that all its ideals are principal. In the univariate case the situation is as follows. Consider a non-zero ideal $I \subseteq P$, where $P = K[x]$ is a polynomial ring over the field K, and let $g = a_d x^d + a_{d-1}x^{d-1} + \ldots + a_0$ be a generator of I such that $a_d \neq 0$, *i.e.* such that $\deg(f) = d$. By using the division with remainder algorithm, for any given polynomial f we get a representation $f = qg + r$, where r is either zero or a polynomial of degree less than d. This implies that every element in the ring $K[x]/(g)$ can be uniquely represented as a linear combination of the residue classes $1, \bar{x}, \ldots, \bar{x}^{d-1}$, in which the coefficients are exactly those of r.

When we deal with polynomials in two indeterminates, we can try to imitate the procedure explained in Example 1.2.11 and proceed as follows.

Example 1.2.12. Let $f = xy + x - y$, $g_1 = x - 1$, and $g_2 = xy - 3$ be three polynomials in $\mathbb{Q}[x, y]$. In order to imitate the division with remainder algorithm we look for polynomials q_1, q_2, and r such that $f = q_1 g_1 + q_2 g_2 + r$. With that goal, we eliminate $\mathrm{LT}_{\mathtt{DegLex}}(f)$ step by step as follows:

$$\begin{array}{llll|l} xy & +x & -y & & x \quad -1 \\ xy & & -y & & \hline y \\ \hline & +x & & & \\ & +x & -1 & & \\ \hline & & +1 & & y \quad +1 \end{array}$$

Note that $\mathrm{LT}_{\mathtt{DegLex}}(1) = 1$ is not divisible by $\mathrm{LT}_{\mathtt{DegLex}}(g_1)$ or $\mathrm{LT}_{\mathtt{DegLex}}(g_2)$ so that it has to be added to the remainder. We obtain a representation $f = q_1 g_1 + q_2 g_2 + r$ such that $q_1 = y + 1$, $q_2 = 0$, and $r = 1$. Note that we have $\deg(r) = 0 < 1 = \deg(g_1)$.

Example 1.2.13. The result of the procedure described in the previous example depends very much on the order of the elements g_1, g_2. For instance, if we let $g_1' = g_2$ and $g_2' = g_1$, and look again for the polynomials q_1', q_2', and r' such that $f = q_1' g_1' + q_2' g_2' + r'$, we get a different result:

$$\begin{array}{llll|l} xy & +x & -y & & xy \quad -3 \\ xy & & & -3 & \hline & & & & +1 \\ \hline & +x & -y & +3 & & & & & +1 \end{array} \qquad \begin{array}{llll|l} x & -y & +3 & & x \quad -1 \\ x & & -1 & & \hline & & & & +1 \\ \hline & -y & +4 & & & & & & +1 \end{array}$$

We find a representation $f = q_2' g_1 + q_1' g_2 + r'$ such that $q_1' = 1$, $q_2' = 1$, and $r' = -y + 4 \neq r$.

Let K be a field, $n \geq 1$, $P = K[x_1, \ldots, x_n]$ be a polynomial ring, and σ be a term ordering on $\mathbb{T}^n$. The procedure described in Example 1.2.12 can be extended to the general case to obtain the following algorithm.

Algorithm 1.2.14. (The Division Algorithm)
Let $s \geq 1$, and let $f, g_1, \ldots, g_s \in P \setminus \{0\}$. Consider the following sequence of instructions.

1. Let $q_1 = \ldots = q_s = 0$, $r = 0$, and $p = f$.
2. Find the smallest $i \in \{1, \ldots, s\}$ such that $\mathrm{LT}_\sigma(p)$ is a multiple of $\mathrm{LT}_\sigma(g_i)$ If such an i exists, replace q_i by $q_i + \frac{\mathrm{LM}_\sigma(p)}{\mathrm{LM}_\sigma(g_i)}$ and p by $p - \frac{\mathrm{LM}_\sigma(p)}{\mathrm{LM}_\sigma(g_i)} \cdot g_i$ Otherwise, replace r by $r + \mathrm{LM}_\sigma(p)$ and p by $p - \mathrm{LM}_\sigma(p)$.
3. Repeat step 2 until $p = 0$. Return the tuple $(q_1, \ldots, q_s) \in P^s$, the polynomial $r \in P$, and stop.

Theorem 1.2.15. *Algorithm 1.2.14 returns a tuple $(q_1, \ldots, q_s) \in P^s$ and a polynomial $r \in P$ such that*

$$f = q_1 g_1 + \ldots + q_s g_s + r$$

and either $r = 0$ or r is a K-linear combination of monomials none of which is divisible by any of $\mathrm{LT}_\sigma(g_1), \ldots, \mathrm{LT}_\sigma(g_s)$. Furthermore, if $q_i g_i \neq 0$, then we have

$$\deg(f) \geq \deg(q_i g_i)$$

Proof. See Theorem 3, Section 3, Chapter 2 in [CLO92]. □

The following definition will be of fundamental importance when we discuss normal forms (see Section 1.2.3).

Definition 1.2.16. Let $s \geq 1$, let $f, g_1, \ldots, g_s \in P \setminus \{0\}$, and let $\mathcal{G}$ be the tuple $(g_1, \ldots, g_s)$. We apply the Division Algorithm and obtain a representation $f = q_1 g_1 + \ldots + q_s g_s + r$ with $q_1, \ldots, q_s, r \in P$. Then the polynomial r is called the **normal remainder** of f with respect to $\mathcal{G}$ and is denoted by $\mathrm{NR}_{\sigma,\mathcal{G}}(f)$, or simply by $\mathrm{NR}_{\mathcal{G}}(f)$ if no confusion can arise. For $f = 0$, we let $\mathrm{NR}_{\mathcal{G}}(f) = 0$.

If we carefully look at the Division Algorithm, we see that the event which triggers step 2 is the detection of a term in the support of p which is a multiple of one of the leading terms $\mathrm{LT}_\sigma(g_1), \ldots, \mathrm{LT}_\sigma(g_s)$. Once such a term has been found, we use the corresponding polynomial g_i as a **rewrite rule** for replacing its leading term by its "tail", with the obvious adjustment if g_i is not monic. Note that in the Division Algorithm the rewrite rules have a well defined hierarchy, *i.e.* the application of the first rewrite rule is preferred to the second one, and so on. If instead we were allowed to use at each step any applicable rewrite rule, we could obtain a different result (as shown in Example 1.2.12). Nevertheless in Section 1.2.3 we will prove the existence of sets of rewrite rules with the property that any possible combination of the applicable rewrite rules will eventually lead to the same result. We rephrase the above ideas in a more technical way as follows.

Definition 1.2.17. Let $g_1, \ldots, g_s \in P \setminus \{0\}$ and $G = \{g_1, \ldots, g_s\}$.

(a) Let $f_1, f_2 \in P$, and suppose there exists a constant $c \in K$, a term $t \in \mathbb{T}^n$, and an index $i \in \{1, \ldots, s\}$ such that $f_2 = f_1 - c\, t\, g_i$ and $t \cdot \mathrm{LT}_\sigma(g_i) \notin \mathrm{Supp}(f_2)$. Then we say that f_1 **reduces to** f_2 **in one step** using the **rewrite rule** defined by g_i, and we write $f_1 \xrightarrow{g_i} f_2$. The passage from f_1 to f_2 is also called a **reduction step**.

(b) The transitive closure of the relations $\xrightarrow{g_1}, \ldots, \xrightarrow{g_s}$ is called the **rewrite relation** defined by G and is denoted by $\xrightarrow{G}$.

(c) An element $f_1 \in P$ with the property that there is no $i \in \{1, \ldots, s\}$ and no $f_2 \in P \setminus \{f_1\}$ such that $f_1 \xrightarrow{g_i} f_2$ is called **irreducible** with respect to $\xrightarrow{G}$.

(d) The equivalence relation defined by $\xrightarrow{G}$ will be denoted by $\xleftrightarrow{G}$.

In the following proposition we give a fundamental property of rewrite relations.

Proposition 1.2.18. *Let $f \in P$, $g_1, \ldots, g_s \in P \setminus \{0\}$, and let $G = \{g_1, \ldots, g_s\}$. Then we have $f \xleftrightarrow{G} 0$ if and only if $f \in \langle g_1, \ldots, g_s \rangle$.*

Proof. See Proposition 2.2.2 in [KR00]. □

Unfortunately it is not clear how we could use Proposition 1.2.18 to check whether a given polynomial $f \in P$ is contained in the ideal $\langle g_1, \ldots, g_s \rangle$ (and thus solving the *Ideal Membership Problem*), because we do not know the direction of the reduction steps in $f \xleftrightarrow{G} 0$. In other words, if we use only the reduction steps $f = f_0 \xrightarrow{g_{i_1}} f_1 \xrightarrow{g_{i_2}} \ldots$, we might get stuck at some point with an irreducible element with respect to $\xrightarrow{G}$, as shown in the following example.

Example 1.2.19. Let $n = 2$, $P = \mathbb{Q}[x, y]$, $g_1 = x - 1$ and $g_2 = xy - 3$ be polynomials of P, $G = \{g_1, g_2\}$, and $\sigma = \texttt{DegLex}$. Consider the polynomial $f = x^2y - 3$; since $f = 3 \cdot g_1 + x \cdot g_2$ it follows that it is contained in the ideal $\langle g_1, g_2 \rangle$. But if we use the reduction step $f \xrightarrow{g_1} xy - 3 \xrightarrow{g_1} y - 3$, we arrive at an irreducible element with respect to $\xrightarrow{G}$.

We conclude that, only with the use of a set of rewrite rules, or equivalently with the Division Algorithm, it is *not* possible to solve the ideal membership problem, *i.e.* it is not always possible to decide whether an element $f \in P$ is contained in the ideal $\langle g_1, \ldots, g_s \rangle$. A positive answer in this direction is given in the following section by the theory of Gröbner bases.

1.2.3 Gröbner bases

At the end of Section 1.2.1 we observed that not all of systems of generators of a polynomial ideal are equal, that means there exist systems of generators that are more *special* than others. We called *special* the sets of polynomials which exhibited a particular behaviour with respect to the leading terms of the elements they generated. In this section we introduce another aspect under which a generating set of polynomials could be considered *special*. Let $P = K[x_1, \ldots, x_n]$ be a polynomial ring over a field K, $I \subseteq P$ an ideal, $\{g_1, \ldots, g_s\}$ be a generating set of I, and σ a term ordering on $\mathbb{T}^n$. We recall from Section 1.2.2 that every g_i can be viewed as a rewrite rule, that is a rule for replacing $\mathrm{LM}_\sigma(g_i)$ by its tail $\mathrm{LM}_\sigma(g_i) - g_i$, a polynomial that indeed represents the same residue class. If a polynomial $f \in P$ contains a term in its support which is a multiple of $\mathrm{LT}_\sigma(g_i)$ for some $i \in \{1, \ldots, s\}$ then we can use the rule associated to g_i and rewrite f. The element obtained in this way is congruent to f modulo I. The procedure of moving from one representative of this residue class to another resembles the division algorithm. However, if we apply the rewriting rules $g_1, \ldots, g_s$ in a different order we obtain a different result (see Example 1.2.12). The generating set $\{g_1, \ldots, g_s\}$ of I is considered *special* if, no matter which order we choose, we always arrive at the same final result. In this section we prove the equivalence of the different conditions under which some sets of polynomials are called special, and combine all these ideas with the very important notion of Gröbner bases.

As usual we let K be a field, $n \geq 1$, $P = K[x_1, \ldots, x_n]$ a polynomial ring, and σ a term ordering on $\mathbb{T}^n$. In the following theorem we provide some basic and equivalent characterizations of Gröbner bases (see also Proposition 1.2.8).

Theorem 1.2.20. (Characterization of Gröbner Bases)
Let $G = \{g_1, \ldots, g_s\}$ be a set of non-zero polynomials of P, let $I = \langle g_1, \ldots, g_s \rangle \subseteq P$ the ideal generated by G, and $\xrightarrow{G}$ the corresponding rewrite rule. The following conditions are equivalent.

(a) For every element $f \in I \setminus \{0\}$, there are $f_1, \ldots, f_s \in P$ such that $f = \sum_{i=1}^{s} f_i g_i$ and $\mathrm{LT}_\sigma(f) \geq \mathrm{LT}_\sigma(f_i g_i)$, for all $i = 1, \ldots, s$ such that $f_i g_i \neq 0$.

(b) The set $\{\mathrm{LT}_\sigma(g_1), \ldots, \mathrm{LT}_\sigma(g_s)\}$ generates the ideal $\mathrm{LT}_\sigma(I)$.

(c) For an element $f \in P$, we have $f \xrightarrow{G} 0$ if and only if $f \in I$.

(d) If $f \in I$ is irreducible with respect to $\xrightarrow{G}$ then we have $f = 0$.

(e) For every element $f_1 \in P$, there is a unique element $f_2 \in P$ such that $f_1 \xrightarrow{G} f_2$ and f_2 is irreducible with respect to $\xrightarrow{G}$.

Proof. See Theorem 2.4.1 in [KR00]. □

Definition 1.2.21. Let $G = \{g_1, \ldots, g_s\} \subseteq P$ be a set of non-zero polynomials which generates the ideal $I = \langle g_1, \ldots, g_s \rangle \subseteq P$. If the conditions of Theorem 1.2.20 are satisfied, then G is called a **Gröbner basis** of I with respect to σ, or equivalently a **σ-Gröbner basis** of I. In the case $I = \langle 0 \rangle$, the only possible Gröbner basis is $G = \emptyset$.

The following result concerns the existence of Gröbner bases. From Proposition 1.2.8 it follows that there are polynomials $g_1, \ldots, g_s \in I$ satisfying condition (b) of Theorem 1.2.20. The question is whether they generate the ideal I; the following proposition answers to it affirmatively.

Proposition 1.2.22. (Existence of a σ-Gröbner Basis)
Let I be a non-zero ideal of P.

(a) Given $g_1, \ldots, g_s \in I \setminus \{0\}$ such that $\mathrm{LT}_\sigma(I) = \langle \mathrm{LT}_\sigma(g_1), \ldots, \mathrm{LT}_\sigma(g_s) \rangle$, we have $I = \langle g_1, \ldots, g_s \rangle$, and the set G is a σ-Gröbner basis of I.

(b) The ideal I has a σ-Gröbner basis $G = \{g_1, \ldots, g_s\} \subseteq I \setminus \{0\}$.

Proof. See Proposition 2.4.3 in [KR00]. □

The above result does not give any guarantees about the uniqueness of Gröbner bases: given a term ordering σ, an ideal $I \subseteq P$ has many σ-Gröbner bases. For instance, we can add arbitrary non-zero elements of I to a σ-Gröbner basis and it remains a σ-Gröbner basis of I. The notion of σ-Gröbner basis can be refined to achieve uniqueness as follows.

Definition 1.2.23. Let $G = \{g_1, \ldots, g_s\} \subseteq P \setminus \{0\}$ and $I = \langle g_1, \ldots, g_s \rangle$. We say that G is a **reduced σ-Gröbner basis** of I if the following conditions are satisfied.

(a) For $i = 1, \ldots, s$, we have $\mathrm{LC}_\sigma(g_i) = 1$.

(b) $\{\mathrm{LT}_\sigma(g_1), \ldots, \mathrm{LT}_\sigma(g_s)\}$ is a minimal system of generators of $\mathrm{LT}_\sigma(I)$.

(c) For $i = 1, \ldots, s$, we have $\mathrm{Supp}(g_i - \mathrm{LT}_\sigma(g_i)) \cap \mathrm{LT}_\sigma\{I\} = \emptyset$.

Theorem 1.2.24. (Existence and uniqueness of reduced σ-Gröbner bases) *For every ideal $I \subseteq P$ there exists a unique reduced σ-Gröbner basis.*

Proof. See Theorem 2.4.13 in [KR00]. □

Let $G = \{g_1, \ldots, g_s\}$ be a σ-Gröbner basis of the ideal $I = \langle g_1, \ldots, g_s \rangle \subseteq P$, and let $f \in P$. By an abuse of notation we call G the tuple $(g_1, \ldots, g_s)$ as well; from Theorem 1.2.20 part (e) it follows that there exists a unique $f_G \in P$ such that $f \xrightarrow{G} f_G$ and f_G is irreducible w.r.t. $\xrightarrow{G}$. *A priori* this element seems to depend on the Gröbner basis chosen, though in facr it does not, as proved in [KR00], Proposition 2.4.7. We give a name to this important concept.

Definition 1.2.25. Let $I \subseteq P$ be a non-zero ideal, and let $f \in P$. The element $f_G \in P$ described above is called the **normal form** of f modulo I with respect to σ. It is denoted by $\mathrm{NF}_{\sigma,I}(f)$, or simply $\mathrm{NF}_\sigma(f)$ if it is clear which ideal is considered.

The following corollary contains the most important property of normal forms.

Corollary 1.2.26. *In the above situation $\mathrm{NR}_G(f)$ agrees with $\mathrm{NF}_{\sigma,I}(f)$; in particular, the normal remainder does not depend on the order of the elements $g_1, \ldots, g_s$.*

We observe that the Division Algorithm with respect to a Gröbner basis therefore provides an effective method for computing normal forms, and furnishes a direct solution to the ideal membership problem, that is the problem of checking whether a polynomial is contained in a given polynomial ideal (see [CLO92] and [KR00], Proposition 2.4.10).

1.3 Introduction to Border bases

In this section we summarize the theory of border bases of a zero-dimensional ideal I in the polynomial ring $P = K[x_1, \ldots, x_n]$. The main idea behind the notion of border basis is the search for more "general" systems of generators of I, which nevertheless give rise to a K-basis of the quotient ring P/I. In contrast to Gröbner bases theory which gives a description of P/I using the set of terms $\mathbb{T}^n \setminus \mathrm{LT}_\sigma\{I\}$ (see Macaulay's Basis Theorem), border basis theory aims

at describing the same quotient ring (as well as the ideal I) using a different and more general concept, *i.e.* a factor closed set of monomials $\mathcal{O}$ which may not depend on any term ordering σ. If there exists a term ordering σ such that $\mathcal{O} = \mathbb{T}^n \setminus \mathrm{LT}_\sigma\{I\}$, we will show that the $\mathcal{O}$-border basis of I contains the reduced σ-Gröbner basis of I. Nevertheless, we will see that there exist border bases that cannot be associated to any Gröbner basis. This means that, in the zero-dimensional case, the theory of border bases is indeed an extension of the Gröbner basis theory.

This section is organized as follows: firstly, we introduce the notion of zero-dimensional ideal; then, in Section 1.3.2 we define an order ideal $\mathcal{O}$ as a factor closed set of terms of $\mathbb{T}^n$. The advantages of using such a set rather than an arbitrary one are manifold. For instance, we can define the border $\partial\mathcal{O}$ of $\mathcal{O}$ by $\partial\mathcal{O} = \cup_{i=1}^n x_i\mathcal{O} \setminus \mathcal{O}$. To describe the multiplicative structure of P/I it suffices to know how products $x_i t \in \partial\mathcal{O}$ with $t \in \mathcal{O}$ are expressed as a linear combination of monomials in $\mathcal{O}$ plus an element of I. Moreover, by defining $\overline{\partial\mathcal{O}} = \mathcal{O} \cup \partial\mathcal{O}$ and $\partial^2\mathcal{O} = \partial(\overline{\partial\mathcal{O}})$ and repeating this procedure, we can introduce higher borders, and measure the "distance" of each term t from the set $\mathcal{O}$. Although this notion is not quite as well behaved as the degree, it enables us to define the Border Division Algorithm (see Section 1.3.3), which is a generalization of the Division Algorithm when no term ordering is given. Finally, Section 1.3.4 is dedicated to defining formally the concept of border bases, analyzing their characteristics, and comparing them with the theory of Gröbner bases.

1.3.1 Zero-dimensional ideals

For the rest of this chapter we let K be a field, $P = K[x_1, \ldots, x_n]$ a polynomial ring over K, $g_1, \ldots, g_s \in P$, and $I = \langle g_1, \ldots, g_s \rangle \subseteq P$ be the generated ideal. In order to introduce the notion of zero-dimensional ideal we recall the equivalent conditions which define the Finiteness Criterion.

Proposition 1.3.1. (Finiteness Criterion)
Let σ be a term ordering on $\mathbb{T}^n$. The following conditions are equivalent.

(a) For $i = 1, \ldots, n$ we have $I \cap K[x_i] \neq (0)$.

(b) The K-vector space P/I is finite-dimensional.

(c) The set $\mathbb{T}^n \setminus \mathrm{LT}_\sigma\{I\}$ is finite.

(d) For every $i \in \{1, \ldots, n\}$ there exists a number $\alpha_i \geq 0$ such that we have $x_i^{\alpha_i} \in \mathrm{LT}_\sigma(I)$.

Proof. See Proposition 3.7.1 in [KR00]. □

Definition 1.3.2. An ideal $I = \langle g_1, \ldots, g_s \rangle \subseteq P$ is called **zero-dimensional** if it satisfies the equivalent conditions of the Finiteness Criterion 1.3.1.

1.3.2 Order ideals

Let $\mathbb{T}^n$ be the monoid of terms in P. The following kind of subset of $\mathbb{T}^n$ will be central to this section.

Definition 1.3.3. Let $\mathcal{O}$ be a non-empty subset of $\mathbb{T}^n$.

(a) The **factor closure** (abbr. **closure**) of $\mathcal{O}$ is the set $\overline{\mathcal{O}}$ of all power products in $\mathbb{T}^n$ which divide some power product of $\mathcal{O}$.

(b) The set $\mathcal{O}$ is called an **order ideal** if $\mathcal{O} = \overline{\mathcal{O}}$, *i.e.* if $\mathcal{O}$ is factor closed.

Since the concept of order ideal has been used in several branches of mathematics, it appears in the literature with many different names: a collection of alternatives can be found in [KR05], Section 0.5. Given an order ideal $\mathcal{O}$, we introduce other useful subsets of $\mathbb{T}^n$ related to $\mathcal{O}$.

Definition 1.3.4. Let $\mathcal{O} \subseteq \mathbb{T}^n$ be an order ideal.

(a) The **border** $\partial\mathcal{O}$ of $\mathcal{O}$ is defined to be $\partial\mathcal{O} = (x_1\mathcal{O} \cup \cdots \cup x_n\mathcal{O}) \setminus \mathcal{O}$. The **first border closure** of $\mathcal{O}$ is the set $\overline{\partial\mathcal{O}} = \mathcal{O} \cup \partial\mathcal{O}$.

(b) For every $k \geq 1$, we inductively define the $(k+1)^{\text{st}}$ **border** of $\mathcal{O}$ by $\partial^{k+1}\mathcal{O} = \partial(\overline{\partial^k\mathcal{O}})$ and the $(k+1)^{\text{st}}$ **border closure** of $\mathcal{O}$ by the rule $\overline{\partial^{k+1}\mathcal{O}} = \overline{\partial^k\mathcal{O}} \cup \partial^{k+1}\mathcal{O}$. For convenience, we let $\partial^0\mathcal{O} = \overline{\partial^0\mathcal{O}} = \mathcal{O}$.

(c) The elements of the minimal set of generators of the monomial ideal corresponding to $\mathbb{T}^n \setminus \mathcal{O}$ are called the **corners** of $\mathcal{O}$.

We observe that the set of corners of $\mathcal{O}$ is a subset of its border $\partial\mathcal{O}$, and its $k-$th border closure is an order ideal for every $k \geq 0$.

Example 1.3.5. 1. Let $\mathcal{O} = \{1, x, y, y^2, xy^2\} \subset \mathbb{T}^2$; the set $\mathcal{O}$ is not an order ideal as it does not contain xy which is a factor of $xy^2 \in \mathcal{O}$. We visualize $\mathcal{O}$ in Figure 1.1.

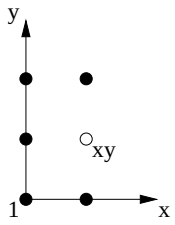


Figure 1.1: Representation of $\mathcal{O}$ in $\mathbb{T}^2$

2. Let $\mathcal{O} = \{1, x, y, x^2, xy, y^2, x^3, x^2y, y^3, y^4\} \subset \mathbb{T}^2$; the set $\mathcal{O}$ is an order ideal. Its border is $\partial\mathcal{O} = \{xy^2, x^4, x^3y, x^2y^2, xy^3, xy^4, y^5\}$, and its corners are the set $C = \{xy^2, x^4, x^3y, y^5\}$. We visualize $\mathcal{O}$, its border $\partial\mathcal{O}$, and the set of corners C in Figure 1.2.

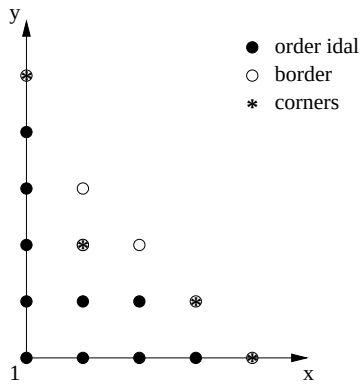


Figure 1.2: Representation of $\mathcal{O}$, $\partial\mathcal{O}$ and C in $\mathbb{T}^2$

The following definition allows us to measure the "distance" of a term from an order ideal.

Definition 1.3.6. Let $\mathcal{O} \subseteq \mathbb{T}^n$ be an order ideal.

(a) For every $t \in \mathbb{T}^n$, the unique number $k \in \mathbb{N}$ such that $t \in \partial^k\mathcal{O}$ is called the **index** of t with respect to $\mathcal{O}$ and is denoted by $\mathrm{ind}_{\mathcal{O}}(t)$.

(b) For a polynomial $f \in P \setminus \{0\}$, we define the **index** of f with respect to $\mathcal{O}$ (or the $\mathcal{O}$-index of f) by $\mathrm{ind}_{\mathcal{O}}(f) = \max\{\mathrm{ind}_{\mathcal{O}}(t) \mid t \in \mathrm{Supp}(f)\}$.

We point out that there is a drawback to using the $\mathcal{O}$-index to order terms in $\mathbb{T}^n$: the $\mathcal{O}$-index ordering is not always compatible with multiplication since $\mathrm{ind}_{\mathcal{O}}(t) \leq \mathrm{ind}_{\mathcal{O}}(t')$ does not, in general, imply $\mathrm{ind}_{\mathcal{O}}(tt'') \leq \mathrm{ind}_{\mathcal{O}}(t't'')$. We show such a situation in the following example.

Example 1.3.7. Let $\mathcal{O} = \{1, x, y, x^2, xy, x^2y\} \subseteq \mathbb{T}^2$. Then $\mathcal{O}$ is an order ideal with border $\partial\mathcal{O} = \{y^2, xy^2, x^3, x^2y^2, x^3y\}$. Figure 1.3 illustrates the situation. Consider the terms x^2 and y^2: we have $\mathrm{ind}_{\mathcal{O}}(x^2) < \mathrm{ind}_{\mathcal{O}}(y^2)$. However, multiplying both terms by x^2 we get $\mathrm{ind}_{\mathcal{O}}(x^2 \cdot x^2) > \mathrm{ind}_{\mathcal{O}}(x^2 \cdot y^2)$.

The following definition identifies those order ideals of particular interest to us.

Definition 1.3.8. Let $I \subseteq P$ be a zero-dimensional ideal, and $\mathcal{O} \subseteq \mathbb{T}^n$ be a factor closed set of terms. If the residue classes of the elements of $\mathcal{O}$ form a vector space basis of P/I then we call $\mathcal{O}$ a **quotient basis** for I.

1.3.3 The border division algorithm

We let $I \subseteq P$ be a zero-dimensional ideal and $\mathbb{T}^n$ be the monoid of terms of P. Further, we let $\mathcal{O} = \{t_1, \ldots, t_\mu\}$ be a finite order ideal in $\mathbb{T}^n$, and $\partial\mathcal{O} = \{b_1, \ldots, b_\nu\}$ be its border. When possible, we shall use the order ideal $\mathcal{O}$

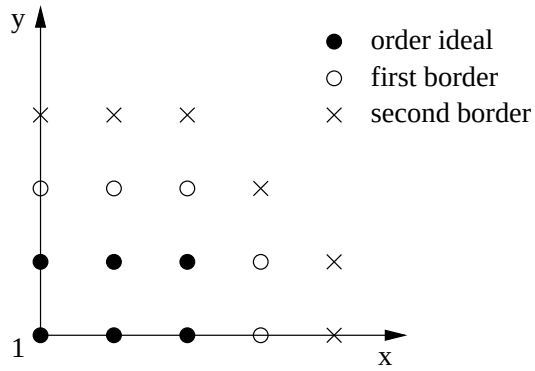


Figure 1.3: Representation of $\mathcal{O}$, $\partial\mathcal{O}$ and $\partial^2\mathcal{O}$ in $\mathbb{T}^2$

as a quotient basis for the zero-dimensional ideal I. Given a zero-dimensional ideal we shall look for a system of generators with the following shape.

Definition 1.3.9. A set of polynomials $G = \{g_1, \ldots, g_\nu\}$ is called an **$\mathcal{O}$-border prebasis** if the polynomials have the form $g_j = b_j - \sum_{i=1}^{\mu} \alpha_{ij} t_i$ with each coefficient $\alpha_{ij} \in K$.

Using border prebases it is possible to define an algorithm for polynomial division with remainder similar to the classical Division Algorithm (see Section 1.2.2).

Algorithm 1.3.10. (The Border Division Algorithm)
Let $\mathcal{O} = \{t_1, \ldots, t_\mu\}$ be an order ideal in $\mathbb{T}^n$, let $\partial\mathcal{O} = \{b_1, \ldots, b_\nu\}$ be its border, and $\{g_1, \ldots, g_\nu\}$ be an $\mathcal{O}$-border prebasis. Given a polynomial $f \in P$, consider the following sequence of instructions.

1. Let $f_1 = \ldots = f_\nu = 0$, $c_1 = \ldots = c_\mu = 0$, and $h = f$.
2. If $h = 0$, return $(f_1, \ldots, f_\nu, c_1, \ldots, c_\mu)$ and stop.
3. If $\mathrm{ind}_{\mathcal{O}}(h) = 0$ then write $h = c_1 t_1 + \ldots + c_\mu t_\mu$ with $c_1, \ldots, c_\mu \in K$. Return $(f_1, \ldots, f_\nu, c_1, \ldots, c_\mu)$ and stop.
4. If $\mathrm{ind}_{\mathcal{O}}(h) > 0$ then let $h = a_1 h_1 + \ldots + a_s h_s$ with $a_1, \ldots, a_s \in K \setminus \{0\}$ and distinct $h_1, \ldots, h_s \in \mathbb{T}^n$ such that $\mathrm{ind}_{\mathcal{O}}(h_1) = \mathrm{ind}_{\mathcal{O}}(h)$. Determine the smallest index $i \in \{1, \ldots, \nu\}$ for which h_1 factors as $h_1 = t\, b_i$ with a term $t \in \mathbb{T}^n$ of degree $\mathrm{ind}_{\mathcal{O}}(h) - 1$. Subtract $a_1 t\, g_i$ from h, add $a_1 t$ to f_i, and continue with step 2.

Theorem 1.3.11. *Algorithm 1.3.10 returns a tuple $(f_1, \ldots, f_\nu, c_1, \ldots, c_\mu) \in P^\nu \times K^\mu$ such that*

$$f = f_1 g_1 + \ldots + f_\nu g_\nu + c_1 t_1 + \ldots + c_\mu t_\mu$$

and $\deg(f_i) \leq \mathrm{ind}_{\mathcal{O}}(f) - 1$ for all $i \in \{1, \ldots, \nu\}$ with $f_i g_i \neq 0$. This representation does not depend on the choice of the term h_1 in step 4.

Proof. See Proposition 6.4.11 in [KR05]. □

As can happen with the Division Algorithm, the order of the polynomials in the $\mathcal{O}$-border prebasis can affect the outcome of the Border Division Algorithm. Further, just as we did in Section 1.2.2, the result of the Border Division Algorithm is used to define the **normal $\mathcal{O}$-remainder** $\mathrm{NR}_{\mathcal{O},G}(f) = c_1t_1 + \ldots + c_\mu t_\mu$ of a polynomial f with respect to the tuple $G = (g_1, \ldots, g_\nu)$. The elements f and $\mathrm{NR}_{\mathcal{O},G}(f)$ represent the same residue class in the ring $P/\langle g_1, \ldots, g_\nu\rangle$; in particular, the residue classes of the elements of $\mathcal{O}$ generate the K-vector space $P/\langle g_1, \ldots, g_\nu\rangle$, though they do not always represent a basis of it. We formalize this notion in Section 1.3.4.

Also in border basis theory it is possible to define the concept of rewrite relation associated to a border prebasis $G = \{g_1, \ldots, g_\nu\}$. Let $f \in P$, let $t \in \mathrm{Supp}(f)$ be a multiple of a border term $t = t'b_i$, and let $c \in K$ be the coefficient of t in f. Then $h = f - ct'g_i$ does not contain the term t anymore. We say that f **reduces to** h **in one step** using g_i and write $f \xrightarrow{g_i} h$. The reflexive, transitive closure of the relations $\xrightarrow{g_i}$, $i \in \{1, \ldots, \nu\}$, is called the **rewrite relation** associated to G and is denoted by $\xrightarrow{G}$. Rewrite relations associated to border prebases do not exhibit all the properties enjoyed by the ones associated to Gröbner bases, nevertheless the equivalence relation $\xleftrightarrow{G}$ captures equivalence modulo a zero-dimensional ideal, and this is enough for characterizing border bases (see Theorem 1.3.19).

1.3.4 Border bases

In this subsection we define formally the notion of border bases.

Definition 1.3.12. Let $\mathcal{B} = \{g_1, \ldots, g_\nu\}$ be an $\mathcal{O}$-border prebasis, and let $I \subseteq P$ be an ideal containing $\mathcal{B}$. If $\mathcal{O}$ is a quotient basis for I then $\mathcal{B}$ is called a **border basis** of I founded on $\mathcal{O}$, or more briefly, an **$\mathcal{O}$-border basis** of I.

From Definition 1.3.12 it is not very clear that an $\mathcal{O}$-border basis of I is indeed a system of generators of the ideal: we claim this fact in the following proposition.

Proposition 1.3.13. *Let $\mathcal{B}$ be an $\mathcal{O}$-border basis of a zero-dimensional ideal $I \subseteq P$. Then the ideal I is generated by $\mathcal{B}$.*

Proof. See Proposition 6.4.15 in [KR05]. □

A necessary condition for the existence of an $\mathcal{O}$-border basis of I is clearly given by $\#\mathcal{O} = \dim_K(P/I)$. However, this condition is not sufficient, as the following example shows.

Example 1.3.14. Let $P = \mathbb{Q}[x, y]$, and let

$$I = \langle xy - \frac{1}{2}y^2 - x + \frac{1}{2}y,\ x^2 - y,\ y^3 - 5y^2 + 4y\rangle$$

be a zero-dimensional ideal of P. It is easy to prove that $\dim_{\mathbb{Q}}(P/I) = 4$; in $\mathbb{T}^2$ there are exactly 5 order ideals containing 4 elements:

$$\mathcal{O}_1 = \{1, x, x^2, x^3\},\ \mathcal{O}_2 = \{1, x, x^2, y\},\ \mathcal{O}_3 = \{1, x, y, xy\}$$
$$\mathcal{O}_4 = \{1, x, y, y^2\},\ \mathcal{O}_5 = \{1, y, y^2, y^3\}$$

But not all of these are suitable for founding a border basis of I: in fact $\mathcal{O}_2$ cannot be a quotient basis for I since $y - x^2 \in I$. Similarly $\mathcal{O}_5$ cannot form a basis for P/I since $y^3 - 5y^2 + 4y \in I$.

The following fundamental proposition proves the existence and uniqueness of border bases.

Proposition 1.3.15. (Existence and Uniqueness of Border Bases) *Let $I \subseteq P$ be a zero-dimensional ideal, and let $\mathcal{O} = \{t_1, \ldots, t_\mu\}$ be a quotient basis for I.*

(a) There exists a unique $\mathcal{O}$-border basis $\mathcal{B}$ of I.

(b) Let $\mathcal{B}$ be an $\mathcal{O}$-border prebasis whose elements are in I. Then $\mathcal{B}$ is the $\mathcal{O}$-border basis of I.

(c) Let k be the field of definition of I. Then the $\mathcal{O}$-border basis of I is contained in $k[x_1, \ldots, x_n]$.

Proof. See Proposition 6.4.17 in [KR05]. □

The existence of a border basis for a zero-dimensional ideal I is related to the existence of an order ideal $\mathcal{O}$, which is also a quotient basis for I. From Macaulay's Basis Theorem we know that, given an ideal I, each term ordering σ on $\mathbb{T}^n$ defines a unique order ideal $\mathcal{O}_\sigma(I) = \mathbb{T}^n \setminus \mathrm{LT}_\sigma\{I\}$, whose residue classes form a vector space basis of P/I. The existence of border bases can thus be proved using a result of Gröbner bases theory. We show it and the relationship between Gröbner bases and border bases in the following proposition.

Proposition 1.3.16. *Let σ be a term ordering on $\mathbb{T}^n$, let $I \subseteq P$ be a zero-dimensional ideal, and $\mathcal{O}_\sigma(I)$ be the order ideal $\mathbb{T}^n \setminus \mathrm{LT}_\sigma\{I\}$. Then there exists a unique $\mathcal{O}_\sigma(I)$-border basis $\mathcal{B}$ of I, and the reduced σ-Gröbner basis of I is the subset of $\mathcal{B}$ corresponding to the corners of $\mathcal{O}_\sigma(I)$.*

Proof. See Proposition 6.4.18 in [KR05]. □

In general the ideal I does not necessarily have an $\mathcal{O}$-border basis for every order ideal $\mathcal{O}$ consisting of $\dim_K(P/I)$ elements, but there always exists an $\mathcal{O}$-border basis of I, when $\mathcal{O} = \mathbb{T}^n \setminus \mathrm{LT}_\sigma\{I\}$ and σ is some term ordering. On the other hand, we show in Example 1.3.17 that not every border basis arises from a term ordering. In particular, a border basis of an ideal I does not arise from a term ordering exactly when the order ideal on which the basis is founded is not of the form $\mathcal{O}_\sigma(I)$, for some term ordering σ. In this sense, the theory of border bases of zero-dimensional ideals is a generalization of the theory of Gröbner bases.

Example 1.3.17. Let $P = \mathbb{Q}[x, y]$, let

$$I = \langle 4xy - 5y^2 - 6x + 9y,\ x^2 - y^2 - 3x + 3y,\ 2y^3 - 9y^2 - 2x + 11y \rangle$$

be a zero-dimensional ideal of P, and $\mathcal{O} = \{1, x, y, xy\} \subseteq \mathbb{T}^2$ be an order ideal. It is easy to prove that $\mathcal{O}$ is a quotient basis for I, so Proposition 1.3.15 guarantees the existence of the unique $\mathcal{O}$-border basis $\mathcal{B}$ of I. However, $\mathcal{B}$ does not arise from any term ordering σ on $\mathbb{T}^2$: in fact, if $x <_\sigma y$ then $x^2 <_\sigma xy$ so that $\mathrm{LT}_\sigma(I) = \langle x^2, xy, y^3 \rangle$ or $\mathrm{LT}_\sigma(I) = \langle x^4, y \rangle$. Much the same happens if $y <_\sigma x$: then $y^2 <_\sigma xy$ and $\mathrm{LT}_\sigma(I) = \langle y^2, xy, x^3 \rangle$ or $\mathrm{LT}_\sigma(I) = \langle y^4, x \rangle$. In any case $\mathcal{O}_\sigma(I) = \mathbb{T}^2 \setminus \mathrm{LT}_\sigma\{I\} \neq \mathcal{O}$.

We want to highlight the properties of border bases just as we did for Gröbner bases (see Proposition 1.2.20). For this reason we first introduce the notion of border form which mimics the concept of leading term.

Definition 1.3.18. Given $f \in P$, we write $f = a_1 u_1 + \ldots + a_s u_s$ with coefficients $a_1, \ldots, a_s \in K \setminus \{0\}$ and terms $u_1, \ldots, u_s \in \mathbb{T}^n$ satisfying $\mathrm{ind}_\mathcal{O}(u_1) \geq \ldots \geq \mathrm{ind}_\mathcal{O}(u_s)$.

(a) For $f \neq 0$, we call $\mathrm{BF}_\mathcal{O}(f) = \sum_{\{i | \mathrm{ind}_\mathcal{O}(u_i) = \mathrm{ind}_\mathcal{O}(f)\}} a_i u_i \in P$ the **border form** of f with respect to $\mathcal{O}$. For $f = 0$, we let $\mathrm{BF}_\mathcal{O}(f) = 0$.

(b) Given an ideal $I \subseteq P$, we call $\mathrm{BF}_\mathcal{O}(I) = \langle \mathrm{BF}_\mathcal{O}(f) \mid f \in I \rangle$ the **border form ideal** of I with respect to $\mathcal{O}$.

The following result holds.

Theorem 1.3.19. (Characterization of Border Bases)
Let $I \subseteq P$ be a zero-dimensional ideal, $\mathcal{O} \subseteq \mathbb{T}^n$ be a quotient basis for I, $\mathcal{B} = \{g_1, \ldots, g_\nu\}$ be the $\mathcal{O}$-border basis of I, and $\xrightarrow{\mathcal{B}}$ be the corresponding rewrite rule. The following conditions are equivalent.

(a) For every $f \in I \setminus \{0\}$, there exist polynomials $f_1, \ldots, f_\nu \in P$ such that $f = f_1 g_1 + \ldots + f_\nu g_\nu$ and $\deg(f_i) \leq \mathrm{ind}_\mathcal{O}(f) - 1$ whenever $f_i \neq 0$.

(b) The set $\{\mathrm{BF}_\mathcal{O}(g_1), \ldots, \mathrm{BF}_\mathcal{O}(g_\nu)\}$ generates the ideal $\mathrm{BF}_\mathcal{O}(I)$.

(c) For an element $f \in P$, we have $f \xrightarrow{\mathcal{B}} 0$ if and only if $f \in I$.

(d) If $f \in I$ is irreducible with respect to $\xrightarrow{\mathcal{B}}$, then we have $f = 0$.

(e) For every element $f_1 \in P$, there is a unique element $f_2 \in P$ such that $f_1 \xrightarrow{\mathcal{B}} f_2$ and f_2 is irreducible with respect to $\xrightarrow{\mathcal{B}}$.

Proof. See Propositions 6.4.23, 6.4.25, and 6.4.28 in [KR05]. □

Similarly to Gröbner bases, border bases exhibit good behaviour with respect to the Border Division algorithm: in Proposition 6.4.19, [KR05] it is shown that the normal $\mathcal{O}$-remainder does not depend on the order of the elements of a border basis. The notion of normal form is thus generalized to border basis theory.

Definition 1.3.20. Let $\mathcal{B} = \{g_1, \ldots, g_\nu\}$ be the $\mathcal{O}$-border basis of a zero-dimensional ideal I. The $\mathcal{O}$-**normal form** of a polynomial $f \in P$ with respect to $\mathcal{O}$ is the polynomial $\mathrm{NF}_{\mathcal{O},I} = \mathrm{NF}_{\mathcal{O},\mathcal{B}}(f)$.

We observe that if there exists a term ordering σ such that $\mathcal{O} = \mathcal{O}_\sigma(I)$ we have $\mathrm{NF}_{\mathcal{O},I}(f) = \mathrm{NF}_{\sigma,I}(f)$.

Chapter 2

Ideals of exact points

In this chapter we introduce and characterize a finite set of points, the most important mathematical object of this thesis. Based on [KR05], Section 6.3, in the first section we illustrate the concept of the vanishing ideal $\mathcal{I}(\mathbb{X})$ of an affine point set $\mathbb{X}$, and show how some geometric properties of $\mathbb{X}$ are related to the algebraic properties of the affine coordinate ring $P/\mathcal{I}(\mathbb{X})$. In Section 2.2 we describe the Buchberger-Möller Algorithm[BM82], an efficient method which, starting from the coordinates of the points, computes a Gröbner basis of the vanishing ideal of an affine point set. This classical result is generalized in two different ways: in Section 2.3 we present some equivalent variants of it obtained simply by replacing the operation of row reduction by the least squares method; in Section 2.4 we perform a similar generalization for computing border bases of ideals of points.

2.1 The vanishing ideal of a set of points

Let $n \geq 1$, let K be a field, and $P = K[x_1, \ldots, x_n]$ be a polynomial ring over K. We introduce the basic definition of a finite set of points of K^n.

Definition 2.1.1. An element $p = (c_1, \ldots, c_n) \in K^n$ is called a **point** of K^n; the numbers $c_1, \ldots, c_n \in K$ are called the **coordinates** of p.
A finite set $\mathbb{X} = \{p_1, \ldots, p_s\}$ of distinct points $p_1, \ldots, p_s \in K^n$ is called an **affine point set**, or more simply a **finite set of points** of K^n.

We can generalize the notion of evaluation homomorphism (given in Section 1.1.1) to a finite set $\mathbb{X}$ of points of K^n, and define the following objects.

Definition 2.1.2. Let $\mathbb{X} = \{p_1, \ldots, p_s\}$ be an affine point set of K^n, and let $G = \{g_1, \ldots, g_k\} \subseteq P$ be a non-empty finite set of polynomials.

(a) The $K-$linear map $\text{eval}_{\mathbb{X}} : P \to K^s$ defined by

$$\text{eval}_{\mathbb{X}}(f) = \begin{pmatrix} f(p_1) \\ \vdots \\ f(p_s) \end{pmatrix}$$

is called the **evaluation map** associated to $\mathbb{X}$. For brevity we write $f(\mathbb{X})$ to mean $\text{eval}_{\mathbb{X}}(f)$.

(b) The **evaluation matrix** of G at $\mathbb{X}$, written $M_G(\mathbb{X}) \in \text{Mat}_{s\times k}(K)$, is defined as having entry (i, j) equal to $g_j(p_i)$, *i.e.* whose columns are the images of the polynomials g_i under the evaluation map.

An algebraic way to describe a finite set $\mathbb{X}$ of points of K^n is to give the set of all the polynomials in P which vanish at all the points of $\mathbb{X}$.

Definition 2.1.3. Let $\mathbb{X} \subseteq K^n$ be a finite set of points. The set of all polynomials $f \in P$ such that $f(p) = 0$ for all points $p \in \mathbb{X}$, forms an ideal of P. This ideal is called the **vanishing ideal** of $\mathbb{X}$ in P and is denoted by

$$\mathcal{I}(\mathbb{X}) = \{f \in P \mid f(p) = 0,\ \forall\, p \in \mathbb{X}\}$$

In the following example we compute the vanishing ideal in a very simple case: $s = 1$, *i.e.* the set $\mathbb{X}$ is made up of a single point of K^n.

Example 2.1.4. Let $p = (c_1, \ldots, c_n) \in K^n$ be a point of K^n and let $\mathbb{X} = \{p\}$. It is simple to prove that the vanishing ideal of $\mathbb{X}$ is given by

$$\mathcal{I}(\mathbb{X}) = \mathcal{I}(p) = \langle x_1 - c_1, \ldots, x_n - c_n\rangle \subseteq P$$

Note that, in this very particular case, the generators $\langle x_1 - c_1, \ldots, x_n - c_n\rangle$ are both the reduced σ-Gröbner basis (for any term ordering σ), and the $\mathcal{O}$-border basis (where $\mathcal{O} = \{1\}$) of the vanishing ideal $\mathcal{I}(\mathbb{X})$.

The basic properties of the vanishing ideal $\mathcal{I}(\mathbb{X})$ are collected in the following proposition.

Proposition 2.1.5. (Basic Properties of vanishing ideals)
Let $\mathbb{X} = \{p_1, \ldots, p_s\}$ be a finite set of distinct points of K^n.

(a) We have $\mathcal{I}(\mathbb{X}) = \mathcal{I}(p_1) \cap \ldots \cap \mathcal{I}(p_s)$.

(b) The map $\varphi : P/\mathcal{I}(\mathbb{X}) \to K^s$ defined by $\varphi(f + \mathcal{I}(\mathbb{X})) = \text{eval}_{\mathbb{X}}(f)$ is an isomorphism of K-algebras. In particular, the ideal $\mathcal{I}(\mathbb{X})$ is zero-dimensional.

(c) For any term ordering σ on $\mathbb{T}^n$, the set $\mathcal{O}_\sigma(I) = \mathbb{T}^n \setminus \text{LT}_\sigma\{\mathcal{I}(\mathbb{X})\}$ consists of precisely s terms.

Proof. See Proposition 6.3.3 in [KR05]. □

2.2 The Buchberger-Möller algorithm

In this section we present an algorithm, known as the Buchberger-Möller (BM) Algorithm, for computing a Gröbner basis of the vanishing ideal $\mathcal{I}(\mathbb{X})$, where $\mathbb{X} = \{p_1, \ldots, p_s\}$ is a finite set of distinct points of K^n. A very naive way to compute $\mathcal{I}(\mathbb{X})$ is to use the formula $\mathcal{I}(\mathbb{X}) = \mathcal{I}(p_1) \cap \ldots \cap \mathcal{I}(p_s)$ given in Proposition 2.1.5, part (a), and the result of Example 2.1.4. However, this approach is computationally not efficient for large sets of points, *i.e.* when $s \gg 0$, because the necessary intersection computations become very lengthly. To remedy this situation an algorithm of low complexity (as it depends polynomially on n and s) was presented by B. Buchberger and M. Möller in [BM82]. The success of the BM-Algorithm is due to its efficiency, its intrinsic simplicity and its bent to generalizations. Several ways to generalize it have since now been proposed: for instance, it has been applied to points with multiplicity (see [Lak91]), to points lying in projective spaces (see [ABKR00], [AKR05]), or to points known with limited accuracy (see [HKPP06], [AFT08], [Fas08]). Though the BM-Algorithm is already very efficient, there exist also variants to optimize it. For instance, if K is the field of rational numbers, a modular version of the BM-Algorithm is presented in [ABKR00], where a modular technique is used to tame the problem of coefficient growth. We present here the BM-Algorithm as it was presented in [BM82].

Remark 2.2.1. Before presenting the algorithm, we note that in [BM82] the evaluation vectors are represented as rows rather than as columns. The reason is that the authors decided to simplify the matrices by using row reductions, that is by performing elementary operations on the rows. We recall here that a matrix is in **row echelon form** if it satisfies the following requirements: 1) all nonzero rows are above any rows of all zeros; 2) the leading coefficient of a row (*i.e.* its first nonzero entry) is strictly to the right of the leading coefficient of the row above it. A matrix can be always reduced to row echelon form using elementary row operations.

Algorithm 2.2.2. (The Original Buchberger-Möller Algorithm)
Let K be a field, let $n \geq 1$, let σ be a term ordering on the power products $\mathbb{T}^n$, and let $\mathbb{X} = \{p_1, \ldots, p_s\}$ be a finite set of distinct points of K^n. Consider the following sequence of instructions.

BM1 Start with the empty lists $\mathcal{G} = [\,]$, $\mathcal{O} = [\,]$, $S = [\,]$, a list $L = [1]$, and the matrix $M = (m_{ij})$ over K with s columns and initially zero rows.

BM2 If $L = [\,]$, return the pair $(\mathcal{G}, \mathcal{O})$ and stop. Otherwise, choose the term $t = \min_\sigma(L)$, the smallest according to the ordering σ, and delete it from L.

BM3 Compute the evaluation vector $t(\mathbb{X}) = (t(p_1), \ldots, t(p_s)) \in K^s$, and reduce the matrix $M' = \binom{M}{v}$ to row echelon form by computing

$$v = t(\mathbb{X}) - \sum_i \alpha_i (m_{i1}, \ldots, m_{is}) \qquad \alpha_i \in K$$

BM4 If $v = (0, \ldots, 0)$ then append the polynomial $t - \sum \alpha_i s_i$ to $\mathcal{G}$, where s_i is the i-th element in S. Remove from L all multiples of t. Continue with step BM2.

BM5 Otherwise $v \neq (0, \ldots, 0)$, so append v as a new row of M and append $t - \sum_i \alpha_i s_i$ as a new element of S. Adjoin t to $\mathcal{O}$, and put into L those elements of $\{x_1 t, \ldots, x_n t\}$ which are neither multiples of an element of L nor of $\mathrm{LT}_\sigma(\mathcal{G})$. Continue with step BM2.

Theorem 2.2.3. *Algorithm 2.2.2 stops after finitely many iterations. It returns a pair $(\mathcal{G}, \mathcal{O})$ such that $\mathcal{G}$ is the reduced σ-Gröbner basis of the vanishing ideal $\mathcal{I}(\mathbb{X})$ and $\mathcal{O} = \mathbb{T}^n \setminus \mathrm{LT}_\sigma\{\mathcal{I}(\mathbb{X})\}$ is a quotient basis of $\mathcal{I}(\mathbb{X})$.*

Proof. See Theorem 6.3.10 in [KR05]. □

We illustrate the BM-Algorithm by reproducing a complete computation on a small affine point set $\mathbb{X}$.

Example 2.2.4. Let $\mathbb{X}$ be the finite set of points of $\mathbb{Q}^2$ consisting of the 4 points $p_1 = (-1, 1)$, $p_2 = (0, 0)$, $p_3 = (1, 1)$, and $p_4 = (2, 4)$ and let $\sigma =$DegRevLex. We compute $\mathcal{I}(\mathbb{X})$ and follow the steps of the algorithm.

BM1 Let $\mathcal{G} = [\,]$, $\mathcal{O} = [\,]$, $S = [\,]$, and $L = [1]$.

BM2 Choose $t = 1$, and let $L = [\,]$.

BM3 Compute $t(\mathbb{X}) = (t(p_1), \ldots, t(p_4)) = (1, 1, 1, 1) = (v_1, v_2, v_3, v_4)$.

BM5 Let $M = (1, 1, 1, 1)$, $S = [1]$, $\mathcal{O} = \{1\}$, and $L = [y,\ x]$.

BM2 Choose $t = y$ and let $L = [x]$.

BM3 Compute $t(\mathbb{X}) = (t(p_1), \ldots, t(p_4)) = (1, 0, 1, 4)$ and $v = (0, -1, 0, 3)$.

BM5 Let $M = \left(\begin{smallmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & 0 & 3 \end{smallmatrix}\right)$, $S = [1,\ y - 1]$, $\mathcal{O} = \{1,\ y\}$, and $L = [x,\ y^2]$.

BM2 Choose $t = x$ and let $L = [y^2]$.

BM3 Compute $t(\mathbb{X}) = (t(p_1), \ldots, t(p_4)) = (-1, 0, 1, 2)$ and $v = (0, 0, 2, 6)$.

BM5 Let $M = \left(\begin{smallmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & 0 & 3 \\ 0 & 0 & 2 & 6 \end{smallmatrix}\right)$, $S = [1,\ y - 1,\ x + y]$, $\mathcal{O} = \{1,\ y,\ x\}$, and $L = [y^2,\ xy,\ x^2]$.

BM2 Choose $t = y^2$ and let $L = [xy,\ x^2]$.

BM3 Compute $t(\mathbb{X}) = (t(p_1), \ldots, t(p_4)) = (1, 0, 1, 16)$ and $v = (0, 0, 0, 12)$.

BM5 Let $M = \left(\begin{smallmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & 0 & 3 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & 12 \end{smallmatrix}\right)$, $S = [1,\ y - 1,\ x + y,\ y^2 - y]$, $\mathcal{O} = \{1,\ y,\ x,\ y^2\}$, and $L = [xy,\ x^2,\ y^3]$.

BM2 Choose $t = xy$ and let $L = [x^2,\ y^3]$.

BM3 Compute $t(\mathbb{X}) = (t(p_1), \ldots, t(p_4)) = (-1, 0, 1, 8)$ and $v = (0, 0, 0, 0)$.

BM4 Let $\mathcal{G} = (xy - \frac{1}{2}y^2 - x + \frac{1}{2}y)$ and $L = [x^2, y^3]$.

BM2 Choose $t = x^2$ and let $L = [y^3]$.

BM3 Compute $t(\mathbb{X}) = (t(p_1), \ldots, t(p_4)) = (1, 0, 1, 4)$ and $v = (0, 0, 0, 0)$.

BM4 Let $\mathcal{G} = (xy - \frac{1}{2}y^2 - x + \frac{1}{2}y,\ x^2 - y)$ and $L = [y^3]$.

BM2 Choose $t = y^3$ and let $L = [\,]$.

BM3 Compute $t(\mathbb{X}) = (t(p_1), \ldots, t(p_4)) = (1, 0, 1, 64)$ and $v = (0, 0, 0, 0)$.

BM4 Let $\mathcal{G} = (xy - \frac{1}{2}y^2 - x + \frac{1}{2}y,\ x^2 - y,\ y^3 - 5y^2 + 4y)$ and $L = [\,]$.

BM2 Return $(\mathcal{G}, \mathcal{O})$ and stop.

The result of this computation is that $\mathcal{G} = \{xy - \frac{1}{2}y^2 - x + \frac{1}{2}y, x^2 - y, y^3 - 5y^2 + 4y\}$ is the reduced σ-Gröbner basis of $\mathcal{I}(\mathbb{X})$ and $\mathcal{O} = \{1, y, x, y^2\}$ represents a $\mathbb{Q}$-basis of $\mathbb{Q}[x, y]/\mathcal{I}(\mathbb{X})$.

2.3 Some variations to the BM-Algorithm

Though it is not evident from its design, from a conceptual point of view the BM-Algorithm can be split into two principal tasks: determining the quotient basis $\mathcal{O}$ (and implicitly $\mathrm{LT}_\sigma\{\mathcal{I}(\mathbb{X})\}$), and then computing the reduced σ-Gröbner basis $\mathcal{G}$ of $\mathcal{I}(\mathbb{X})$ from $\mathrm{LT}_\sigma\{\mathcal{I}(\mathbb{X})\}$ by linear algebra. The core of the first task stands in the computation of a vector space basis of $P/\mathcal{I}(\mathbb{X})$: the reduction to row echelon form performed in step BM3 is only a tool for testing the linear dependence of the evaluation vector $t(\mathbb{X})$ (in this particular case represented as a row) on the space spanned by the rows of the matrix M. In Section 2.3.2 we present an equivalent version of the BM-Algorithm which is essentially obtained by replacing the operation of row reduction by another procedure for testing the linear dependence of vectors, namely the method of least squares, introduced in the following section.

2.3.1 The method of least squares

Let $r, c \in \mathbb{N}$, with $r > c$, and let K be a field. Consider the linear system of equations:

$$Ax = b \tag{2.1}$$

where $A \in \mathrm{Mat}_{r\times c}(K)$ is a full rank matrix, and $b \in K^r$ is a vector. The linear system $Ax = b$ contains more equations than unknowns, *i.e.* it is **overdetermined**. Usually a full rank overdetermined system does not allow any exact

solution since a necessary condition for its existence is that the vector b belongs to the subspace generated by the columns of A, which is generically a proper subspace of K^r. In such cases the **Least Squares Problem** (LSP) comes into play:

$$\text{find } x^* \in K^c \quad \text{s.t. } \|Ax^* - b\|_2 = \min_{x \in K^c} \|Ax - b\|_2 \tag{2.2}$$

Solving (2.2) means looking for an "optimal" solution of $Ax = b$ or, equivalently, determining a vector $x^* \in \mathbb{R}^c$ which minimizes the Euclidean distance between Ax and b. If x^* is the solution of the LSP, then we have

$$A^t(Ax^* - b) = 0 \tag{2.3}$$

We prove it by contadiction; suppose $x \in K^c$, $z \in K^c$, and $\alpha \in K$ and consider the equality

$$\begin{aligned} \|A(x^* + \alpha z) - b\|_2^2 &= \|(Ax^* - b) + \alpha Az\|_2^2 = \\ &= \|Ax^* - b\|_2^2 + 2\alpha z^t A^t(Ax^* - b) + \alpha^2 \|Az\|_2^2 \end{aligned}$$

If we choose $z = -A^t(Ax - b)$ and $\alpha \in K$ such that $2\alpha\|z\|_2^2 < \|Az\|_2^2$ then we obtain the contradictory inequality $\|A(x^* + \alpha z) - b\|_2 < \|Ax^* - b\|_2$. From (2.3) we may also conclude that, if x^* is a solution of the LSP, then $x^* + w$ solves (2.2) if and only if $w \in \ker(A)$; since A has full column rank it follows that (2.2) admits a unique solution.

In order to give an explicit solution to (2.2) we introduce the pseudoinverse (also known as Moore-Penrose pseudoinverse) of a matrix which is a generalization of the notion of the inverse matrix.

Definition 2.3.1. Let $r, c \in \mathbb{N}$, and let $A \in \mathrm{Mat}_{r \times c}(K)$; the **Moore-Penrose pseudoinverse** (or more simply the **pseudoinverse**) of A is the unique matrix $A^\dagger \in \mathrm{Mat}_{c \times r}(K)$ satisfying the four Moore-Penrose conditions:

(i) $AA^\dagger A = A$

(ii) $A^\dagger A A^\dagger = A^\dagger$

(iii) $(AA^\dagger)^t = AA^\dagger$

(iv) $(A^\dagger A)^t = A^\dagger A$

If $K = \mathbb{C}$, parts (iii) and (iv) of the above definition hold with the conjugate transpose. In some special cases an explicit formula for $A^\dagger$ is provided.

Proposition 2.3.2. *Let $A \in \mathrm{Mat}_{r \times c}(K)$.*

(a) If A has orthonormal columns (that is $A^tA = I_c$) or orthonormal rows (that is $AA^t = I_r$) then $A^\dagger = A^t$.

(b) If A has full column rank then $A^\dagger = (A^tA)^{-1}A^t$, and it is a left inverse of A, that is $A^\dagger A = I_c$.

(c) *If A has full row rank then $A^\dagger = A^t(AA^t)^{-1}$, and it is a right inverse of A, that is $AA^\dagger = I_r$.*

(d) *If both columns and rows of A are linearly independent, that is if A is a non-singular square matrix, then $A^\dagger = A^{-1}$.*

Proof. See Section 5.5.4 in [GL89]. □

We have the following result.

Proposition 2.3.3. *Let $r, c \in \mathbb{N}$, with $r > c$, let $A \in \mathrm{Mat}_{r\times c}(K)$ be a full rank matrix, and $S \subseteq K^r$ be the vector subspace generated by the columns of A. Then the operation $\pi : K^r \to K^r$ defined by $\pi : v \mapsto AA^\dagger v$ is the* **orthogonal projection** *onto S, that is $\mathrm{Im}(\pi) = \mathrm{S}$ and $\pi^2 = \pi$.*

Proof. See Section 5.5.4 in [GL89]. □

Note that a similar result holds when $r < c$, that is in the case the matrix A has full row rank and the subspace S is generated by its rows.

Now, we want to relate the notion of pseudoinverse to the LSP (2.2). If we multiply each side of (2.1) by A^t we obtain

$$A^tAx = A^tb$$

which is a linear system whose exact solution $x^* \in K^c$ is

$$x^* = (A^tA)^{-1}A^tb = A^\dagger b \tag{2.4}$$

The vector x^* is indeed the unique solution of the LSP (2.2) since, from Proposition 2.3.3, we know that $AA^\dagger b = Ax^*$ is the projection of b onto the subspace S generated by the columns of A. The vector

$$\rho^* = b - Ax^* = (I - AA^\dagger)b \in K^r \tag{2.5}$$

is called the **minimal residual** of the least squares problem (2.2), and is often used to check whether b belongs to S, as the following statement holds true:

$$b \in S \iff \rho^* = 0$$

Note that $\rho^* = 0$ is also a necessary and sufficient condition to have an exact solution to the overdetermined linear system (2.1).

2.3.2 Equivalent algorithms

In this section we present some equivalent variants of the BM-Algorithm. As in Section 2.2 we let $\mathbb{X} = \{p_1, \ldots, p_s\}$ be a finite set of distinct points of K^n, and $\mathcal{I}(\mathbb{X})$ be its vanishing ideal. We start with a trivial variation of Algorithm 2.2.2: a small change in step BM5 is performed in order to obtain a matrix M whose transposed is the evaluation matrix $M_{\mathcal{O}}(\mathbb{X})$ instead of $M_S(\mathbb{X})$ as in the original version.

Algorithm 2.3.4. In the setting of Algorithm 2.2.2, replace step BM5 by the following instruction:

BM5' Otherwise $v \neq (0,\ldots,0)$, so append $t(\mathbb{X})$ as a new row of M and $t - \sum_i \alpha_i s_i$ as a new element of S. Adjoin t to $\mathcal{O}$, and put into L those elements of $\{x_1 t, \ldots, x_n t\}$ which are neither multiples of an element of L nor of $\mathrm{LT}_\sigma(\mathcal{G})$. Continue with step BM2.

It is easy to conclude that the above algorithm is equivalent to the BM-Algorithm, and thus it returns the same result. In the following algorithm we introduce deeper changes: the new version is obtained by replacing the operation of row reduction performed in the BM-Algorithm, step BM3, by a similar procedure for testing the linear dependence of vectors based on the least squares method (see Section 2.3.1).

Algorithm 2.3.5. (The least squares version)
In the setting of Algorithm 2.2.2 consider the following sequence of instructions.

LS1 Start with the empty lists $\mathcal{G} = [\,]$, $\mathcal{O} = [\,]$, $S = [\,]$, a list $L = [1]$, and the matrix $M = (m_{ij})$ over K with s rows and initially zero columns.

LS2 If $L = [\,]$, return the pair $(\mathcal{G}, \mathcal{O})$ and stop. Otherwise, choose the term $t = \min_\sigma(L)$, the smallest according to the ordering σ, and delete it from L.

LS3 Compute the evaluation vector $t(\mathbb{X}) = (t(p_1), \ldots, t(p_s)) \in K^s$, and solve the least squares problem $Mx = b$, where $b = t(\mathbb{X})$, by computing the vectors x^* and ρ^* (see formulas (2.4) and (2.5))

$$\begin{aligned} x^* &= M^\dagger b \\ \rho^* &= (I_s - MM^\dagger)b \end{aligned}$$

LS4 If $\rho^* = (0,\ldots,0)$ then append the polynomial $t - \sum x_i^* s_i$ to $\mathcal{G}$, where s_i is the i-th element in S. Remove from L all multiples of t. Continue with step LS2.

LS5 Otherwise $\rho^* \neq (0,\ldots,0)$, so append $\rho^*(\mathbb{X})$ as a new column of M and append $t - \sum_i x_i^* s_i$ as a new element of S. Adjoin t to $\mathcal{O}$, and put into L those elements of $\{x_1 t, \ldots, x_n t\}$ which are neither multiples of an element of L nor of $\mathrm{LT}_\sigma(\mathcal{G})$. Continue with step LS2.

Theorem 2.3.6. *Algorithm 2.3.5 stops after finitely many iterations. It returns a pair $(\mathcal{G}, \mathcal{O})$ such that $\mathcal{G}$ is the reduced σ-Gröbner basis of the vanishing ideal $\mathcal{I}(\mathbb{X})$ and $\mathcal{O} = \mathbb{T}^n \setminus \mathrm{LT}_\sigma\{\mathcal{I}(\mathbb{X})\}$ is a quotient basis of $\mathcal{I}(\mathbb{X})$.*

Proof. We observe that this algorithm has the same structure of the BM-Algorithm. Its termination and its correcteness can be simply proved by adapting the proof of the BM-Algorithm (*e.g.* see Proposition 6.3.10 in [KR05]) to this particular case. □

Note that, differently from the classical BM-Algorithm, which computes the matrix M making an incremental triangularization of the transposed of the evaluation matrix $M_{\mathcal{O}}(\mathbb{X})$ with operations of row reduction, Algorithm 2.3.5 builds M by computing, with the least squares method, an incremental orthogonalization of $M_{\mathcal{O}}(\mathbb{X})$. From a computational point of view the operations in step LS3 are not very demanding since $M^{\dagger} = (M^t M)^{-1} M^t$ and the matrix $M^t M$ is diagonal. Our decision to give here a detailed description of Algorithm 2.3.5 is due to the fact that the generalized version of the BM-Algorithm for sets of approximate points will be based on it (see Chapter 5).

2.4 Computation of border bases

In this section we describe a method for computing a border basis of the vanishing ideal $\mathcal{I}(\mathbb{X})$, where $\mathbb{X} = \{p_1, \ldots, p_s\}$ is a finite set of distinct points of K^n. The procedure consists essentially of two main tasks: determining a quotient basis $\mathcal{O}$ for $\mathcal{I}(\mathbb{X})$, and then computing the border basis of $\mathcal{I}(\mathbb{X})$ founded on it. We describe how to perform the first task using a strategy based on the BM-Algorithm (or its variants, see Sections 2.2 and 2.3.2). The quotient basis $\mathcal{O}$ is built stepwise: initially $\mathcal{O}$ is empty; then, at each iteration, a new power product t is considered. If the evaluation vector $t(\mathbb{X})$ is linearly independent of the set of evaluation vectors $\{t_i(\mathbb{X}) \mid t_i \in \mathcal{O}\}$ then t is added to $\mathcal{O}$; otherwise t is added to the set of corners of the order ideal.

Algorithm 2.4.1. (The Quotient Basis Algorithm)
Let K be a field, let $n \geq 1$, and let $\mathbb{X} = \{p_1, \ldots, p_s\}$ be a finite (non-empty) set of distinct points of K^n. Consider the following sequence of instructions.

QB1 Start with the empty lists $\mathcal{O} = [\,]$, $C = [\,]$, $V = [\,]$, and a list $L = [1]$.

QB2 If $L = [\,]$ return $\mathcal{O}$ and stop. Otherwise, choose any term $t \in L$, and delete it from L.

QB3 Compute the evaluation vector $t(\mathbb{X}) = (t(p_1), \ldots, t(p_s)) \in K^s$, and check whether $t(\mathbb{X})$ is linearly dependent on the elements of V.

QB4 If $t(\mathbb{X})$ is linearly dependent on the elements of V, then add t to C. Continue with step QB2.

QB5 Otherwise $t(\mathbb{X})$ is linearly independent of the elements of V, so add $t(\mathbb{X})$ to the list V. Append t to $\mathcal{O}$, and put into L those elements of $\{x_1 t, \ldots, x_n t\}$ which are neither multiples of an element of L nor of C. Continue with step QB2.

Theorem 2.4.2. *Algorithm 2.4.1 stops after finitely many iterations and returns a set of terms $\mathcal{O}$ which is a quotient basis of $\mathcal{I}(\mathbb{X})$.*

Proof. First we exhibit termination. In each iteration either step QB4 is performed or step QB5. By its construction, the list V is made up of linearly

independent vectors. Hence step QB5, which appends a vector to V, can be performed at most s times. Since the list L is enlarged only in step QB5 and each iteration removes one element of L, we arrive at $L = \emptyset$ after finitely many iterations.

Now we exhibit correctness. Since, by construction, the set $V = \{t(\mathbb{X}) | t \in \mathcal{O}\}$ is made up of linearly independent vectors, it follows that the residue classes of the elements of $\mathcal{O}$ form a vector space basis of $P/\mathcal{I}(\mathbb{X})$. To show that $\mathcal{O}$ is factor closed we prove by induction on the number of iterations of the algorithm that, after steps QB2-QB5 have been performed, $L \cup C$ is the set of corners of $\mathcal{O}$. This is clearly true after the first iteration: since $\mathbb{X}$ is not empty, the term $t = 1$ is added to $\mathcal{O}$ and $L \cup C = \{x_1, \ldots, x_n\}$ is the set of corners of $\mathcal{O}$. Now we follow the steps of one iteration. Let $t, t_1, \ldots, t_k$ be the elements of $L \cup C$, and suppose that the term $t \in L$ is considered at the current iteration. If step QB4 is performed, the claim is trivially true since neither $\mathcal{O}$ nor $L \cup C$ do change. Suppose that the step QB5 is performed. The set of corners of $\mathcal{O} \cup \{t\}$ is made up by the elements $z \in \mathbb{T}^n$ which belong to the minimal set of generators of the monomial ideal $\langle t_1, \ldots, t_k, x_1 t, \ldots, x_n t \rangle$, that is the terms $z = t_j$, $j = 1 \ldots k$, or $z = x_i t$, $i = 1 \ldots n$, and such that z is not a multiple of any t_j. Then the claim follows from the definition of L in step QB5. □

Note that in Algorithm 2.4.1 the main task of step QB3 is to check the linear dependence of a set of vectors, which can be easily done using the techniques we have already described, namely the row reduction (see Section 2.2) or the method of least squares (see Section 2.3.1). Further, in step QB2 the choice of the power product t to consider at each iteration is completely free: any strategy that chooses a term t from the list L can be applied, since L is a subset of the corners of $\mathcal{O}$, and this fact guarantees that $\mathcal{O} \cup \{t\}$ is factor closed. This complete freedom allows Algorithm 2.4.1 to compute any quotient basis for $\mathcal{I}(\mathbb{X})$ including those that do not arise from a term ordering σ. However, as a particular case, the strategy of the BM-Algorithm can be used, where the power product t is chosen according to a fixed term ordering σ, so that Algorithm 2.4.1 computes the quotient basis $\mathcal{O}_\sigma(\mathcal{I}(\mathbb{X}))$ associated to σ. We show this particular case in the following example.

Example 2.4.3. Let $\mathbb{X}$ be the finite set of points of $\mathbb{R}^2$ given in Example 2.2.4

$$\mathbb{X} = \{(-1, 1), (0, 0), (1, 1), (2, 4)\}$$

We apply Algorithm 2.4.1 on $\mathbb{X}$ and follow its steps.

QB1 Let $\mathcal{O} = [\,]$, $C = [\,]$, $V = [\,]$, and $L = [1]$.

QB2 Choose $t = 1$, and let $L = [\,]$.

QB3 Compute $t(\mathbb{X}) = (1, 1, 1, 1)$.

QB5 The vector $t(\mathbb{X})$ is linearly independent of the elements of V.
Let $V = [(1, 1, 1, 1)]$, $\mathcal{O} = \{1\}$, $C = [\,]$ and $L = [x,\ y]$.

QB2 Choose $t = x$, and let $L = [y]$.

QB3 Compute $t(\mathbb{X}) = (-1, 0, 1, 2)$.

QB5 The vector $t(\mathbb{X})$ is linearly independent of the elements of V.
Let $V = [(1, 1, 1, 1), (-1, 0, 1, 2)]$, $\mathcal{O} = \{1,\ x\}$, $C = [\,]$ and $L = [x^2,\ y]$.

QB2 Choose $t = y$, and let $L = [x^2]$.

QB3 Compute $t(\mathbb{X}) = (1, 0, 1, 4)$.

QB5 The vector $t(\mathbb{X})$ is linearly independent of the elements of V.
Let $V = [(1, 1, 1, 1), (-1, 0, 1, 2), (1, 0, 1, 4)]$, $\mathcal{O} = \{1,\ x,\ y\}$, $C = [\,]$ and $L = [x^2,\ xy,\ y^2]$.

QB2 Choose $t = xy$, and let $L = [x^2,\ y^2]$.

QB3 Compute $t(\mathbb{X}) = (-1, 0, 1, 8)$.

QB5 The vector $t(\mathbb{X})$ is linearly independent of the elements of V.
Let $V = [(1, 1, 1, 1), (-1, 0, 1, 2), (1, 0, 1, 4), (-1, 0, 1, 8)]$, $\mathcal{O} = \{1, x, y, xy\}$, $C = [\,]$ and $L = [x^2,\ y^2]$.

QB2 Choose $t = y^2$, and let $L = [x^2]$.

QB3 Compute $t(\mathbb{X}) = (1, 0, 1, 16)$.

QB4 The vector $t(\mathbb{X})$ is linearly dependent on the elements of V. Let $C = [y^2]$.

QB2 Choose $t = x^2$, and let $L = [\,]$.

QB3 Compute $t(\mathbb{X}) = (1, 0, 1, 4)$.

QB4 The vector $t(\mathbb{X})$ is linearly dependent on the set V. Let $C = [y^2,\ x^2]$.

QB2 The list L is empty, so return $\mathcal{O} = \{1,\ x,\ y,\ xy\}$ and stop.

Note that, although Algorithm 2.4.1 is very similar to the BM-Algorithm, it computes a different quotient basis of $\mathcal{I}(\mathbb{X})$ (in this case $\{1,\ x,\ y,\ xy\}$ whereas in Example 2.2.4 the returned order ideal was $\{1,\ y,\ x,\ y^2\}$), since in step QB2 it employs a random criterion to choose the term t to be considered.

Starting from a quotient basis $\mathcal{O}$ for the vanishing ideal of $\mathbb{X}$, the $\mathcal{O}$-border basis of $\mathcal{I}(\mathbb{X})$ can be determined with simple linear algebra computations, as described in the following algorithm.

Algorithm 2.4.4. (The Border Basis Algorithm)
Let K be a field, let $n \geq 1$, let $\mathbb{X} = \{p_1, \ldots, p_s\}$ be a finite set of distinct points of K^n, and let $\mathcal{O}$ be a quotient basis for $\mathcal{I}(\mathbb{X})$. Consider the following sequence of instructions.

BB1 Start with the empty list $\mathcal{B} = []$; compute the border $\partial\mathcal{O}$ of $\mathcal{O}$, *e.g.* directly from the formula in Definition 1.3.4, and the evaluation matrix $M_{\mathcal{O}}(\mathbb{X}) \in \mathrm{Mat}_{s,s}(K)$.

BB2 For each $t \in \partial\mathcal{O}$ compute the evaluation vector $t(\mathbb{X}) \in K^s$, compute the solution $\alpha \in K^s$ of the linear system $M_{\mathcal{O}}(\mathbb{X})\,\alpha = t(\mathbb{X})$, and append the polynomial $t - \sum_i \alpha_i t_i$ to $\mathcal{B}$, where t_i is the i-th element in $\mathcal{O}$.

BB3 Return the list $\mathcal{B}$ and stop.

Proposition 2.4.5. *Algorithm 2.4.4 stops after finitely many iterations and returns the $\mathcal{O}$-border basis $\mathcal{B}$ of the vanishing ideal $\mathcal{I}(\mathbb{X})$.*

Proof. The set $\mathcal{B}$ is an $\mathcal{O}$-border prebasis of $I(\mathbb{X})$ and its elements are in $\mathcal{I}(\mathbb{X})$ by construction. Then, our claim follws from Proposition 1.3.15, part (b). □

We observe that in the procedure for computing a border basis for an ideal of points, the main role is played by the quotient basis $\mathcal{O}$, and so the essential part is represented by Algorithm 2.4.1. From Proposition 1.3.16 we recall that, if $I \subseteq P$ is a zero-dimensional ideal, and $\mathcal{O}$ is equal to $\mathcal{O}_\sigma(I)$ for some term ordering σ, the $\mathcal{O}$-border basis contains the σ-Gröbner basis of I. Nevertheless Example 1.3.17 shows that not every border basis arises from a term ordering σ, and this happens exactly when the order ideal on which the basis is founded is not of the form $\mathcal{O}_\sigma(I)$, for some term ordering σ.

We end this section with the following example.

Example 2.4.6. Let $\mathbb{X}$ be the finite set of points of $\mathbb{R}^2$ given in Examples 2.2.4 and 2.4.3; let $\mathcal{O} = \{1, x, y, xy\}$ be the quotient basis computed in Example 2.4.3. We compute the $\mathcal{O}$-border basis of $\mathcal{I}(\mathbb{X})$ by applying the Algorithm 2.4.4 on $\mathbb{X}$.

BB1 Let $\mathcal{B} = [\,]$; compute $\partial\mathcal{O} = \{x^2, x^2y, xy^2, y^2\}$ and the evaluation matrix $M_{\mathcal{O}}(\mathbb{X}) = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{pmatrix}$

BB2 Let $t = x^2$; compute $t(\mathbb{X}) = (1, 0, 1, 4)$, $\alpha = (0, 0, 1, 0)$, and append $x^2 - y$ to $\mathcal{B} = [\,]$.
Let $t = x^2y$, compute $t(\mathbb{X}) = (1, 0, 1, 16)$, $\alpha = (0, -2, 1, 2)$, and append $y^2 - 2xy - y + 2x$ to $\mathcal{B} = [x^2 - y]$.
Let $t = xy^2$, compute $t(\mathbb{X}) = (-1, 0, 1, 32)$, $\alpha = (0, -4, 0, 5)$, and append $xy^2 - 5xy + 4x$ to $\mathcal{B} = [x^2 - y, x^2y - 2xy - y + 2x]$.
Let $t = y^2$, compute $t(\mathbb{X}) = (1, 0, 1, 16)$, $\alpha = (0, -2, 1, 2)$, and append $y^2 - 2xy - y + 2x$ to $\mathcal{B} = [x^2 - y, x^2y - 2xy - y + 2x, xy^2 - 5xy + 4x]$.

BB3 Return $\mathcal{B} = [x^2 - y, x^2y - 2xy - y + 2x, xy^2 - 5xy + 4x, y^2 - 2xy - y + 2x]$ and stop.

The result of this computation is the $\mathcal{O}$-border basis $\mathcal{B}$ of $\mathcal{I}(\mathbb{X})$; note that $\mathcal{B}$ arises from no term orderings on $\mathbb{T}^2$.

Chapter 3

Empirical points and empirical vectors

In this chapter we define a formal framework for dealing with indetermination in $\mathbb{R}^n$ (with $n \geq 1$). In Section 3.1 we introduce the basic definition of empirical point, discuss the analogies with Stetter's definition [Ste04], formalize the idea of redundancy in a set of points and hint at the possible methods to overcome it. In Section 3.2 we give the definition of empirical vector and empirical evaluation vector, we describe the concept of numerical linear dependence, and how to adapt it to the empirical case.

3.1 Finite sets of empirical points

Frequently a mathematical model of a physical phenomenon is derived from processing a large number of real-world measurements which are perturbed by noise, uncertainty and approximation. If each experimental test consists of the measurement of n different physical quantities, the empirical data can be organized as a set of points in $\mathbb{R}^n$. Each point corresponds to a single test, and each coordinate to a single measurement, so that the error affecting each component of the point most likely derives from the limits of accuracy of the measuring instruments. This fact leads to the following basic assumption: given an empirical datum $p \in \mathbb{R}^n$, the estimates $\varepsilon_1, \ldots, \varepsilon_n \in \mathbb{R}^+$ of the error in each component of p are known.

Given the tolerance $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ and a measurement $p \in \mathbb{R}^n$, we view the pair (p, ε) as an empirical point representing a neighbourhood of data which differ from p by less than the tolerance ε. In particular, any point $\widetilde{p} \in \mathbb{R}^n$ which differs from p componentwise by less than the corresponding ε_i can be considered equivalent to p from a numerical point of view. We can formalize this idea by means of the definition of empirical point, introduced by Stetter in [Ste04].

Definition 3.1.1. Let $p \in \mathbb{R}^n$ be a point and let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ with each $\varepsilon_i \in \mathbb{R}^+$, be the vector of the componentwise estimated errors. An **empirical point** p^ε is the pair (p, ε) where we call p the **specified value** and ε the **tolerance**.

We make the following assumption: given a set of empirical points whose specified values belong to $\mathbb{R}^n$, we shall suppose that a single common tolerance vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ exists for the entire set, *i.e.* each value ε_i estimates the uncertainty in the i-th component of all of the points. This is a natural assumption if the values of each single variable derive from real-world measurements of a physical quantity using instruments with the same accuracy. On the other hand different variables typically represent measurements of different physical quantities (*e.g.* temperature and pressure) with different instruments, so the various ε_i are probably all different.

Notation 3.1.2. We denote by $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$, with each $\varepsilon_i \in \mathbb{R}^+$, the common tolerance vector of a set of empirical points of $\mathbb{R}^n$. For any $p \in \mathbb{R}^n$ we write p^ε to mean the corresponding empirical point having p as specified value and ε as tolerance. We denote by $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_s^\varepsilon\}$ a set of empirical points each having tolerance ε, and by $\mathbb{X} = \{p_1, \ldots, p_s\} \subseteq \mathbb{R}^n$ the set of the specified values associated to $\mathbb{X}^\varepsilon$.

Let p^ε be an empirical point of $\mathbb{R}^n$; we define the positive diagonal matrix $E = \mathrm{diag}(1/\varepsilon_1, \ldots, 1/\varepsilon_n) \in \mathrm{Mat}_{n,n}(\mathbb{R})$ and use E-weighted norms $\|\cdot\|_{\alpha,E}$ on $\mathbb{R}^n$ (defined in [DBA74]) in order to "normalize" the distance between points w.r.t. the tolerance ε. For completeness, we recall here the definition:

$$\|v\|_{\alpha,E} := \|Ev\|_\alpha$$

Later on, when using the Euclidean norms $\|\cdot\|_2$ or $\|\cdot\|_{2,E}$, the index 2 will be omitted for simplicity of notation.

An empirical point p^ε defines its neighbourhood of perturbations according to the following definition.

Definition 3.1.3. Let p^ε be an empirical point of $\mathbb{R}^n$. We define its **neighbourhood of perturbations** as

$$N^\alpha(p^\varepsilon) = \{\widetilde{p} \in \mathbb{R}^n \ : \ \|\widetilde{p} - p\|_{\alpha,E} < 1\}$$

and we call each $\widetilde{p} \in N^\alpha(p^\varepsilon)$ an **admissible perturbation** of the specified value p.

As above when we use the Euclidean norm we leave out the index 2, so that we write $N(p^\varepsilon)$ instead of $N^2(p^\varepsilon)$. Note that each element in $N^\alpha(p^\varepsilon)$ can be obtained by perturbing the coordinates of the specified value p by amounts less than the tolerance ε; so we can say that each admissible perturbation $\widetilde{p} \in N^\alpha(p^\varepsilon)$ represents the same empirical information as p. Clearly the shape of $N^\alpha(p^\varepsilon)$, that is the shape of the unit ball of $(\mathbb{R}^n, \|\cdot\|_{\alpha,E})$, is different for different values of α. Figure 3.1 shows some examples of $N^\alpha(p^\varepsilon)$ in $\mathbb{R}^2$ with $\alpha = 1, 2, \infty$.

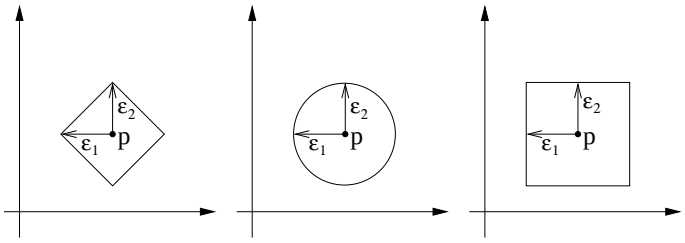


Figure 3.1: Neighbourhoods of perturbations of p^ε with $\varepsilon_1 = \varepsilon_2$

Note that, though deeply inspired by Stetter's work, our definition of neighbourhood of perturbations (see Definition 3.1.3) slightly differs from his idea of *family of neighborhoods* of data clouds. In his formalization, in order to preserve the vague character of indetermination present in every area of scientific computing, Stetter associates to each empirical quantity (p, ε) a family of neighborhoods parametrized by a positive real parameter δ in the following way:

$$N_\delta^\alpha(p, \varepsilon) = \{\widetilde{p} \in \mathbb{R}^n \ : \ \|\widetilde{p} - p\|_{\alpha,E} \leq \delta\}, \qquad \delta \in \mathbb{R} \setminus \{0\}$$

Stetter's concept of empirical data is closely related to that of *fuzzy data*, which is based on a probability distribution and is mainly used in engineering applications (see for instance [AFG+00]); but rather than dealing with formal probabilities Stetter associates with the parameter δ a *validity scale*: he considers the values $\widetilde{p} \in N_\delta^\alpha(p, \varepsilon)$ with $\delta = O(1)$, *i.e.* δ real number of order 1, as *valid instances* of the empirical quantity p^ε. The relation between the values of the parameter $\delta \in \mathbb{R} \setminus \{0\}$ and an intuitive concept of validity can be visualized using the following representation:

values of δ :	$0 \ldots 1$	$1 \ldots 3$	$3 \ldots 10$	$10 \ldots 30$	$10 \ldots 30$	$30 \ldots$
validity :	valid	probably valid	possibly valid	possibly invalid	probably invalid	invalid

This means that, as in our interpretation, any value $\widetilde{p}$ with $\|\widetilde{p} - p\|_{\alpha,E} \leq 1$ is a valid instance for the empirical quantity p^ε; nevertheless any point $\widetilde{p}$ with a larger distance from p may well occur, though this is assumed to be less and less probable. Unlike our deterministic point of view, Stetter considers the value 1 not as a strict boundary for the quantity $\|\widetilde{p} - p\|_{\alpha,E}$ but rather as a mark for the interpretation of the potential numerical values of p^ε. Figure 3.2 shows some examples of $N_\delta^\alpha(p, \varepsilon)$, in $\mathbb{R}^2$ with $\alpha = 1, 2, \infty$: the solid lines corresponds to $\delta = 1$, the dashed line corresponds to $\delta = 3$.

Though we believe that the formalization of the concept of empirical point given by Stetter in [Ste04] is the best one for describing the uncertain data, in this work we decided, for the sake of simplicity, to adopt a slightly different and less general point of view (see Definition 3.1.3).

The following definition concerns the notion of distinct empirical points, *i.e.* points carrying different empirical information.

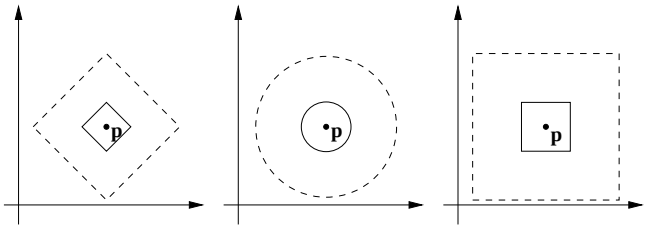


Figure 3.2: Families of neighborhoods $N_\delta^\alpha(p,\varepsilon)$ with $\varepsilon_1 = \varepsilon_2$

Definition 3.1.4. Let p_1^ε and p_2^ε be two empirical points whose specified values belong to $\mathbb{R}^n$; p_1^ε and p_2^ε are said to be **distinct** if

$$N^\alpha(p_1^\varepsilon) \cap N^\alpha(p_2^\varepsilon) = \emptyset$$

Given an empirical set $\mathbb{X}^\varepsilon$, we want to analyze the empirical information carried by its empirical points. Unfortunately it does not seem to be possible to produce a natural definition of "equivalence" between empirical points, as the intuitive relation generated by the nearness measure of points of $\mathbb{R}^n$ is reflexive and symmetric, but seldom transitive (a classical example where transitivity vanishes is sketched in Figure 3.3).

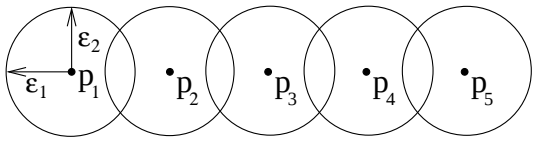


Figure 3.3: Chain configuration

In the following definition we introduce a condition which tells us when an empirical set of points can be represented by a single "equivalent" point (which we choose to be the average of the original set).

Definition 3.1.5. The set $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_s^\varepsilon\}$ of empirical points of $\mathbb{R}^n$ is **collapsable** if

$$\|p_i - q\|_{\alpha,E} \leq 1 \quad \forall\, i = 1, \ldots, s \tag{3.1}$$

where $q = \frac{1}{s}\sum_{i=1}^s p_i$ is the centroid of $\mathbb{X}$.

If $\mathbb{X}^\varepsilon$ is collapsable, the centroid q of $\mathbb{X}$ belongs to each set $N^\alpha(p_i^\varepsilon)$, so the intersection $\bigcap_i N^\alpha(p_i^\varepsilon) \neq \emptyset$. However, this condition alone does not guarantee that the set $\mathbb{X}^\varepsilon$ is collapsable. In fact, for it the intersection $N^\alpha(p_i^\varepsilon)$ must be non-empty *and* the centroid must lie in the intersection. Now, when $\mathbb{X}^\varepsilon$ is collapsable the empirical point q^ε is numerically equivalent to every point in $\mathbb{X}^\varepsilon$. We formalize this idea as follows.

Definition 3.1.6. The **empirical centroid** of a set $\mathbb{X}^\varepsilon$ is the empirical point q^ε where q is the centroid of the set $\mathbb{X}$. If $\mathbb{X}^\varepsilon$ is a collapsable set, its empirical centroid is called the **valid representative** of $\mathbb{X}^\varepsilon$.

In Figure 3.4 we show a simple example of a collapsable set of points and its valid representative (when $n = 2$, $\alpha = 2$, and $\varepsilon_1 = \varepsilon_2$).

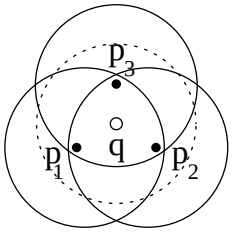


Figure 3.4: A collapsable set of empirical points and its valid representative

If a set of empirical points contains a collapsable subset, it contains some redundancy, *i.e.* it carries relatively little empirical information compared to the number of points in it. Based on this idea we design in Chapter 4 some methods to "thin out" such sets by finding a smaller set of empirical points with much lower redundancy yet which still contains essentially the same empirical information.

We end this section by generalizing the concept of admissible perturbation of an empirical point (given in Definition 3.1.3) to $\mathbb{X}^\varepsilon$. We define an **admissible perturbation** of $\mathbb{X}^\varepsilon$ to be any set made up of s distinct points, each being an admissible perturbation of a different empirical point of $\mathbb{X}^\varepsilon$. The admissible perturbations of $\mathbb{X}^\varepsilon$ are denoted by $\widetilde{\mathbb{X}} = \{\widetilde{p_1}, \ldots, \widetilde{p_s}\} \subseteq \mathbb{R}^n$, where each $\widetilde{p_i} \in N^\alpha(p_i^\varepsilon)$.

3.1.1 A parametric description of $\mathbb{X}^\varepsilon$

Let $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_s^\varepsilon\}$ be a finite set of distinct empirical points with specified values $\mathbb{X} \subseteq \mathbb{R}^n$, and let $\alpha = 2$ (that is we use a weighted 2-norm on $\mathbb{R}^n$); we represent an admissible perturbation of $\mathbb{X}^\varepsilon$ by using first order infinitesimals for the perturbation in each coordinate; that is, we express it as a function of sn **error variables**

$$\mathbf{e} = (e_{11}, \ldots, e_{s1}, e_{12}, \ldots, e_{s2}, \ldots, e_{1n}, \ldots, e_{sn})$$

Specifically, the admissible perturbation is $\widetilde{\mathbb{X}}(\mathbf{e}) = \{\widetilde{p}_1(\mathbf{e}), \ldots, \widetilde{p}_s(\mathbf{e})\}$ where

$$\widetilde{p_k} = (p_{k1} + e_{k1},\ p_{k2} + e_{k2}, \ldots p_{kn} + e_{kn})$$

The conditions on the values of the e_{kj} such that each $\widetilde{p}_k$ is an admissible perturbation of the point p_k are equivalent to the following:

$$\|(e_{k1}, \ldots, e_{kn})\|_{\alpha,E} \leq 1 \qquad \text{for each } k \tag{3.2}$$

We observe that the coordinates of each perturbed point $\widetilde{p}_k(\mathbf{e})$ are elements of the polynomial ring $R = \mathbb{R}[\mathbf{e}]$ and that each variable e_{kj} represents the perturbation in the j-th coordinate of the specified value p_k. The domain of the perturbed set $\widetilde{\mathbb{X}}(\mathbf{e})$, viewed as a function of sn variables, is denoted by D_ε. Obviously, if $\alpha = 2$, that is we use the Euclidean norm on $\mathbb{R}^n$, and if $\delta \in D_\varepsilon$ we have

$$\|\delta\|^2 = \sum_{j=1}^{n}\sum_{k=1}^{s}\delta_{kj}^2 \leq \sum_{j=1}^{n} s\varepsilon_j^2,$$

and consequently

$$\|\delta\| \leq \sqrt{s}\|\varepsilon\| \tag{3.3}$$

To keep evident the dependence on the error variables $\mathbf{e}$, we extend the concepts of Definition 2.1.2, namely the evaluation map of a polynomial $f \in P$ and the evaluation matrix of a set of polynomials $G = \{g_1, \ldots, g_k\} \subseteq P$, to a generic perturbed set $\widetilde{\mathbb{X}}(\mathbf{e})$, using the following notation:

$$\mathrm{eval}_{\widetilde{\mathbb{X}}(\mathbf{e})}(f) = (f(\widetilde{p}_1(\mathbf{e})), \ldots, f(\widetilde{p}_s(\mathbf{e}))) \in R^s$$

for brevity denoted by $f(\widetilde{\mathbb{X}}(\mathbf{e}))$; similarly we write the evaluation matrix

$$M_G(\widetilde{\mathbb{X}}(\mathbf{e})) = \Big(g_1(\widetilde{\mathbb{X}}(\mathbf{e})), \ldots, g_k(\widetilde{\mathbb{X}}(\mathbf{e}))\Big) \in \mathrm{Mat}_{s\times k}(R)$$

3.2 Finite sets of empirical vectors

In this section we define the notion of empirical vector and introduce a generalized concept of linear dependence. Let $n \geq 1$ and let $\mathbb{R}^n$ be the trivial n-dimensional vector space over $\mathbb{R}$. Since each vector $v \in \mathbb{R}^n$ is indeed a point of $\mathbb{R}^n$, the concept of empirical vector directly follows from that of empirical point. For completeness we give here the definition of empirical vector of $\mathbb{R}^n$.

Definition 3.2.1. Let $v \in \mathbb{R}^n$ be a vector and let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ with each $\varepsilon_i \in \mathbb{R}^+$, be the vector of the componentwise estimated errors. An **empirical vector** v^ε is the pair (v, ε) where we call v the **specified value** and ε the **tolerance**.

As for the empirical points, the empirical vectors are elements of $\mathbb{R}^n$ whose components are affected by errors, and only an estimate for the amount of uncertainty is known. Any vector $\widetilde{v} \in \mathbb{R}^n$ which differs from v componentwise by less than the corresponding ε_i can be considered equivalent to v from a numerical point of view. Based on Definition 3.1.3 we introduce the notion of admissible perturbation of an empirical vector.

Definition 3.2.2. Let v^{ε} be an empirical vector of $\mathbb{R}^n$. We define its **neighbourhood of perturbations** as

$$N^{\alpha}(v^{\varepsilon}) = \{\widetilde{v} \in \mathbb{R}^n \ : \ \|\widetilde{v} - v\|_{\alpha,E} < 1\}$$

and we call each $\widetilde{v} \in N^{\alpha}(v^{\varepsilon})$ an **admissible perturbation** of the specified value v.

A very important aspect when dealing with empirical vectors concerns the notion of linear dependence. Let's consider the following example.

Example 3.2.3. Let $v_1 = (1, 0)$ and $v_2 = (1, \delta)$, with $\delta \neq 0$, $|\delta| \ll 1$, be vectors of $\mathbb{R}^2$, and suppose that the components of v_2 are known with limited accuracy (see Figure 3.5). Clearly v_1 and v_2 are linearly independent vectors of $\mathbb{R}^2$. Nevertheless if a small perturbation on v_2 is allowed, the vectors v_1 and v_2 turn out to be linearly dependent.

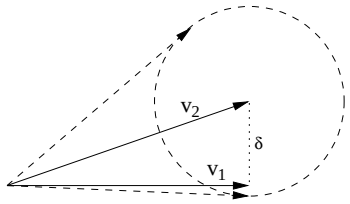


Figure 3.5: Two numerically linearly dependent vectors

Let $k \geq 1$, and let $v_1, \ldots, v_k$ be k vectors of $\mathbb{R}^n$ known with limited accuracy. Suppose that the vectors are linearly independent, so that none of them can be written as a linear combination of the others. In some cases a small perturbation on the components of each vector v_i could lead to a new family of linearly dependent vectors (see for instance Example 3.2.3). Therefore, the original vectors $v_1, \ldots, v_k$ should be considered linearly dependent from a numerical point of view. We formalize this idea in the following definition.

Definition 3.2.4. Let $v_1^{\varepsilon_1}, \ldots, v_k^{\varepsilon_k}$ be k empirical vectors of $\mathbb{R}^n$. The empirical vectors $v_1^{\varepsilon_1}, \ldots, v_k^{\varepsilon_k}$ are said to be **numerically linearly dependent** if there exist admissible perturbations $\widetilde{v}_1, \ldots, \widetilde{v}_k$ of $v_1^{\varepsilon_1}, \ldots, v_k^{\varepsilon_k}$ which are linearly dependent (according to the classical definition). If no such admissible perturbations exist then $v_1^{\varepsilon_1}, \ldots, v_k^{\varepsilon_k}$ are said to be **numerically linearly independent**.

We observe that the above definition generalizes the classical notion of linear dependence. If the empirical vectors $v_1^{\varepsilon_1}, \ldots, v_k^{\varepsilon_k}$ are numerically linearly independent, then, according to the classical definition, the specified values $v_1, \ldots, v_k$ are linearly independent, but the converse does not hold. If the real vectors $v_1, \ldots, v_k$ are linearly dependent, then the same holds, from a numerical point of view, for the empirical vectors $v_1^{\varepsilon_1}, \ldots, v_k^{\varepsilon_k}$ for any choice of the tolerances $\varepsilon_1, \ldots, \varepsilon_n$, that is $v_1^{\varepsilon_1}, \ldots, v_k^{\varepsilon_k}$ are numerically linearly dependent. On the

other hand the numerical linear dependence among a set of empirical vectors $v_1^{\varepsilon_1}, \ldots, v_k^{\varepsilon_k}$ does not imply the classical linear dependence among the set of specified values $v_1, \ldots, v_k$.

In the following definition we introduce a particular empirical vector: the empirical evaluation vector.

Definition 3.2.5. Let $\mathbb{X}^\varepsilon$ be a set of empirical points of $\mathbb{R}^n$, and let $f \in P$ be a polynomial. The empirical vector $f(\mathbb{X}^\varepsilon)$ having $f(\mathbb{X}) \in \mathbb{R}^n$ as its specified value and the set

$$N^\alpha(f(\mathbb{X}^\varepsilon)) = \{f(\widetilde{\mathbb{X}}) \ : \ \widetilde{\mathbb{X}} \in N^\alpha(\mathbb{X}^\varepsilon)\}$$

as its neighbourhood of perturbations is called the **empirical evaluation vector** of f associated to $\mathbb{X}^\varepsilon$.

Note that $N^\alpha(f(\mathbb{X}^\varepsilon))$ has not the same shape as the neighbourhoods of perturbation $N^\alpha(\mathbb{X}^\varepsilon)$ of the empirical points. When dealing with empirical evaluation vectors the notion of numerical linear dependence given in Definition 3.2.4 needs to be more precisely specified.

Definition 3.2.6. Let $\mathbb{X}^\varepsilon$ be a finite set of empirical points of $\mathbb{R}^n$, and let $f_1, \ldots, f_k$ be polynomials of P. If there exist an admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$ such that the evaluation vectors $f_1(\widetilde{\mathbb{X}}), \ldots, f_k(\widetilde{\mathbb{X}})$ are linearly dependent (according to the classical definition), then the empirical evaluation vectors $f_1(\mathbb{X}^\varepsilon), \ldots, f_k(\mathbb{X}^\varepsilon)$ are said to be **numerically linearly dependent**. If no such admissible perturbation exist, then $f_1(\mathbb{X}^\varepsilon), \ldots, f_k(\mathbb{X}^\varepsilon)$ are said to be **numerically linearly independent**.

In order to determine a numerical linear dependence among empirical vectors which derive from evaluation maps, different approaches based on classical linear algebra techniques have since now been proposed. Among them there exist methods which use the theory of the singular value decomposition. In the following example we show how such a method would be inadequate in our context.

Example 3.2.7. Let $P = \mathbb{R}[x, y]$, let $\alpha = 2$, and let $\mathbb{X}^\varepsilon$ be a set of 3 empirical points having $\mathbb{X} = \{(0, 0), (1, 1), (2, 2)\}$ as the set of specified values and $\varepsilon = (0.4, 0.4)$ as the tolerance (see Figure 3.6).

Consider the power products $t_1 = 1$, $t_2 = y$, $t_3 = y^2$, and the order ideal $\mathcal{O} = \{t_1, t_2, t_3\} = \{1, y, y^2\}$. Let $M_\mathcal{O}(\mathbb{X}) \in \mathrm{Mat}_{3\times 3}(\mathbb{R})$ be the evaluation matrix of $\mathcal{O}$ associated to $\mathbb{X}$

$$M_\mathcal{O}(\mathbb{X}) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix} \in \mathrm{Mat}_{3\times 3}(\mathbb{R})$$

Its minimun sigular value is about 0.35, which is less than the componentwise tolerances ε_i, so the empirical vectors $t_1(\mathbb{X}^\varepsilon)$, $t_2(\mathbb{X}^\varepsilon)$, $t_3(\mathbb{X}^\varepsilon)$ would be considered linearly dependent.

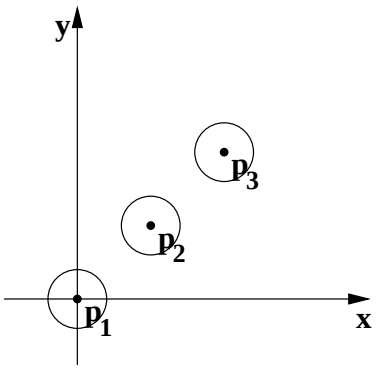


Figure 3.6: Three aligned points

On the contrary, from our point of view, the empirical evaluation vectors $t_1(\mathbb{X}^\varepsilon)$, $t_2(\mathbb{X}^\varepsilon)$, $t_3(\mathbb{X}^\varepsilon)$ are numerically linearly independent since there exists no perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$ such that $t_1(\widetilde{\mathbb{X}})$, $t_2(\widetilde{\mathbb{X}})$, $t_3(\widetilde{\mathbb{X}})$ are linearly dependent. In fact, consider the admissible perturbation $\widetilde{\mathbb{X}} = \{\widetilde{p}_1, \widetilde{p}_2, \widetilde{p}_3\}$ of $\mathbb{X}^\varepsilon$ where:

$$\widetilde{p}_1 = (\widetilde{p}_{11}, \widetilde{p}_{12}) \qquad \widetilde{p}_2 = (\widetilde{p}_{21}, \widetilde{p}_{22}) \qquad \text{and} \qquad \widetilde{p}_3 = (\widetilde{p}_{31}, \widetilde{p}_{32})$$

The evaluation matrix

$$M_{\mathcal{O}}(\widetilde{\mathbb{X}}) = \begin{pmatrix} 1 & \widetilde{p}_{12} & \widetilde{p}_{12}^2 \\ 1 & \widetilde{p}_{22} & \widetilde{p}_{22}^2 \\ 1 & \widetilde{p}_{32} & \widetilde{p}_{32}^2 \end{pmatrix} \in \mathrm{Mat}_{3\times 3}(\mathbb{R})$$

is a Vandermonde matrix having determinant $\det(M_{\mathcal{O}}(\mathbb{X})) = (\widetilde{p}_{22} - \widetilde{p}_{12})(\widetilde{p}_{32} - \widetilde{p}_{22})(\widetilde{p}_{32} - \widetilde{p}_{22})$. It follows that, for each perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$, the matrix $M_{\mathcal{O}}(\widetilde{\mathbb{X}})$ is invertible, and this concludes the proof.

Chapter 4

Reducing redundant empirical points

In this chapter we present some methods (introduced in [AFT07]) to "thin out" a large body of empirical points prior to applying the costly algebraic algorithms described in Chapter 5. Our approach is based on the idea of reducing "redundancy" in the original data: we regard subsets of original points which lie close to each other as repeat measurements, and replace them a single representative value. We illustrate this intuitive idea in the following example where an initial set of 12 points is thinned out to an equivalent set of 4 points.

Example 4.0.8. Given the set $\mathbb{X}$ of 12 points in $\mathbb{R}^2$

$$\begin{aligned}\mathbb{X} \;=\; \{&(-1,-1),\ (0,-1),\ (1,-1),\ (-1,0),\ (0,0),\ (1,0),\\ &(-1,1),\ (0,1),\ (1,1),\ (5,-2.9),\ (5,0),\ (5,2.9)\}\end{aligned}$$

we suppose that each coordinate is perturbed by an error less than 1.43 (the Euclidean norm on $\mathbb{R}^2$ is used). In this situation, the first nine points most likely derive from measurements of the same quantity; therefore it is quite reasonable (and appropriate) to collapse them onto a single representative, for example the point $(0,\ 0)$. In contrast, since the last three points are well separated, they should not be collapsed. This partition, shown in Figure 4.1, is found by our algorithms, as reported in Examples 4.1.3 and 4.1.8.

Based on the idea of clustering together empirical points which could derive from repeated measurements of the same datum, we design three algorithms which thin out a large set of redundant data to produce a smaller set of "equivalent" empirical points. Naturally, the degree of the reduction depends on how much redundancy is present in the original data.

This chapter is organized as follows. In Section 4.1 we present the Agglomerative and the Divisive Algorithms which thin out sets of empirical points while preserving their overall geometrical structure. The two algorithms are included

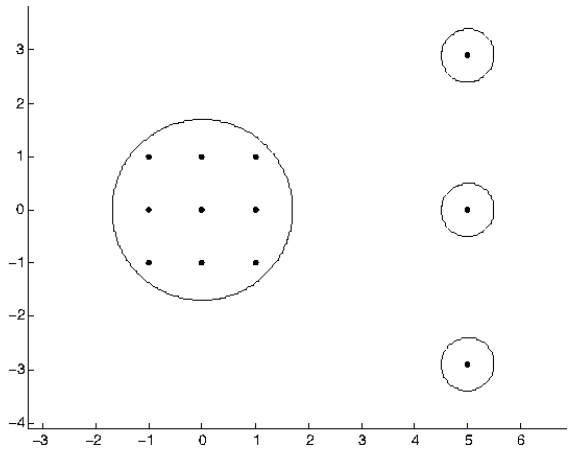


Figure 4.1: Appropriate partition of $\mathbb{X}$

in CoCoALib which is available from the web site [CoC]. In this section we also include a description of the third algorithm, the very simple Grid Algorithm. The relationship with the theory of Cluster Analysis is discussed in Section 4.2. In Section 4.3 we present some numerical examples to illustrate the behaviour of our algorithms on different geometrical configurations of points.

4.1 Algorithms

In this section we describe two algorithms that, given a set $\mathbb{X}^\varepsilon$ of empirical points of $\mathbb{R}^n$, compute a partition $\mathcal{L}^\varepsilon = \{L_1^\varepsilon, \ldots, L_k^\varepsilon\}$ of it, consisting of non-empty collapsable sets, and a set $\mathbb{Y}^\varepsilon = \{q_1^\varepsilon, \ldots, q_k^\varepsilon\}$ where each q_i^ε is the valid representative of L_i^ε. Our algorithms make use of the (weighted) Euclidean norm on $\mathbb{R}^n$ and differ in the strategies for building the partitions:

1. the **Agglomerative Algorithm** initially puts each point of $\mathbb{X}^\varepsilon$ into a different subset and then iteratively unifies pairs of subsets into larger collapsable sets;

2. the **Divisive Algorithm** initially puts all the points of $\mathbb{X}^\varepsilon$ into a single subset and then iteratively splits off the remotest outlier and "evens up" the new partition.

We observe that the partition $\mathcal{L}$ produced by these algorithms enables us to determine easily the **multiplicity** of each valid representative: indeed, the multiplicity of q_i^ε is just the cardinality of L_i^ε.

4.1.1 The Agglomerative Algorithm

The Agglomerative Algorithm (AA) implements a unifying method. The sets in the partition are determined by an iterative process. Initially each set contains a single original empirical point, then iteratively the two closest sets are unified provided their union is collapsable. This method is fast when the input points are well separated w.r.t. the tolerance, as then only a few set unifications are required.

Algorithm 4.1.1. (The Agglomerative Algorithm)
Let $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \dots, p_s^\varepsilon\}$ be a set of empirical points, with each $p_i \in \mathbb{R}^n$ and with a common tolerance $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$. Let $\|\cdot\|_E$ be the weighted 2-norm on $\mathbb{R}^n$ w.r.t. $E = \operatorname{diag}(1/\varepsilon_1, \dots, 1/\varepsilon_n)$. Consider the following sequence of instructions.

AA1 Start with the subset list $\mathcal{L} = [L_1, \dots, L_s]$ where each $L_i = \{p_i\}$, and the list $\mathbb{Y} = [q_1, \dots, q_s]$ of the centroids of the L_i.

AA2 Compute the symmetric matrix $M = (m_{ij})$ where $m_{ij} = \|q_i - q_j\|_E$ for each $q_i, q_j \in \mathbb{Y}$.

AA3 If $|\mathbb{Y}| = 1$ or $\min\{m_{ij} : i < j\} > 2$ then return the pair $(\mathcal{L}, \mathbb{Y})$, and stop.

AA4 Choose $\hat{\imath}$, $\hat{\jmath}$ s.t. $m_{\hat{\imath}\hat{\jmath}} = \min\{m_{ij} : i < j\}$ and compute the centroid q of $L_{\hat{\imath}} \cup L_{\hat{\jmath}}$

$$q = \frac{|L_{\hat{\imath}}| q_{\hat{\imath}} + |L_{\hat{\jmath}}| q_{\hat{\jmath}}}{|L_{\hat{\imath}}| + |L_{\hat{\jmath}}|}$$

AA5 If $\|p - q\|_E \le 1$ for every $p \in L_{\hat{\imath}} \cup L_{\hat{\jmath}}$ then in $\mathcal{L}$ replace $L_{\hat{\imath}}$ by $L_{\hat{\imath}} \cup L_{\hat{\jmath}}$ and remove $L_{\hat{\jmath}}$. Similarly, in $\mathbb{Y}$ replace $q_{\hat{\imath}}$ by q and remove $q_{\hat{\jmath}}$ and then go to step AA2. Otherwise put $m_{\hat{\imath}\hat{\jmath}} = \infty$ (any value greater than 2 will do) and go to step AA3.

Theorem 4.1.2. *Algorithm 4.1.1 computes a pair* $(\mathcal{L}, \mathbb{Y})$ *such that:*

- $\{L_i^\varepsilon : L_i \in \mathcal{L}\}$ *is a partition of* $\mathbb{X}^\varepsilon$ *into collapsable sets, no two of which can be unified into a collapsable set;*
- *for each* $q_i \in \mathbb{Y}$ *the empirical point* q_i^ε *is the valid representative of* L_i^ε.

Proof. First we prove finiteness. Step AA2 is performed only finitely many times and so a finite number of matrices M is computed. In fact, after the first computation of M, this step is performed only when the algorithm removes an element from $\mathbb{Y}$, *i.e.* at most $s-1$ times. Now, also step AA4 is performed only finitely many times on the same matrix M, because it is performed only when the minimal element $m_{\hat{\imath}\hat{\jmath}}$ of the matrix M is less than or equal to 2, in which case either two subsets are unified or $m_{\hat{\imath}\hat{\jmath}}$ is replaced by ∞; but this can happen at most $s^2/2$ times.

Next we show correctness. First, note that the elements of $\mathcal{L}$ define a partition of $\mathbb{X}$. In fact, in step AA1 we set $\mathcal{L} = [\{p_1\}, \dots, \{p_s\}]$; the only place

where $\mathcal{L}$ changes is in Step AA5 when we unite two of its elements, and so a new partition of $\mathbb{X}$ is obtained. Consequently $\mathcal{L}^\varepsilon$ is also a partition of $\mathbb{X}^\varepsilon$.

For each $L_i \in \mathcal{L}$, the corresponding empirical set L_i^ε is collapsable. This is clearly true in step AA1. Step AA5 unites two elements of $\mathcal{L}$ only if their union is collapsable: step AA4 computes the centroid q of $L_i \cup L_j$ and step AA5 tests condition (3.1) for each point in $L_i \cup L_j$.

Now we prove that upon termination the union of any pair of elements of $\mathcal{L}$ is not collapsable. If the algorithm stops because $\mathbb{Y}$ (and $\mathcal{L}$ too) contains a single element, the conclusion is trivial. Otherwise, the algorithm ends because $m_{ij} > 2$ for all $i < j$. We observe that the elements m_{ij} of the final matrix M are such that either $m_{ij} = \|q_i - q_j\|_E$ or $m_{ij} = \infty$ but $\|q_i - q_j\|_E \leq 2$. The case where $m_{ij} = \infty$ is trivial: an entry in M can become ∞ only in step AA5 after having verified that $L_i^\varepsilon \cup L_j^\varepsilon$ is not collapsable. In the case where m_{ij} is finite we show by contradiction that the union of L_i^ε and L_j^ε is not a collapsable set. We suppose that $\|p - q\|_E \leq 1$ for each $p \in L_i \cup L_j$, where q is the centroid of $L_i \cup L_j$. Let $m = |L_i|$ and $n = |L_j|$, then we have

$$\begin{aligned}
\|q_i - q_j\|_E &= \left\| \frac{1}{m}\left(\sum_{p\in L_i} p - mq\right) + \frac{1}{n}\left(nq - \sum_{p\in L_j} p\right)\right\|_E = \\
&= \left\| \frac{1}{m}\sum_{p\in L_i}(p-q) + \frac{1}{n}\sum_{p\in L_j}(q-p)\right\|_E \leq \\
&\leq \frac{1}{m}\sum_{p\in L_i}\|p-q\|_E + \frac{1}{n}\sum_{p\in L_j}\|q-p\|_E
\end{aligned}$$

From the hypothesis, we deduce that $\|q_i - q_j\|_E \leq 2$, a contradiction.

Finally, we can conclude the proof since, by construction, each element $q_i \in \mathbb{Y}$ is the centroid of L_i and L_i^ε is collapsable, so the empirical centroid q_i^ε is indeed the valid representative of L_i^ε. □

Note that, in step AA5, we must check the condition that $\|p - q\|_E \leq 1$ for each $p \in L_{\hat{\imath}} \cup L_{\hat{\jmath}}$. In fact, if we check only the condition $\|q_{\hat{\imath}} - q_{\hat{\jmath}}\|_E \leq 1$, there are pathological examples where not collapsable sets are built in the final partition (see Example 4.3.3).

The algorithm as presented here can easily be improved from the computational point of view: in step AA2 it is not necessary to compute a new matrix M after uniting $L_{\hat{\imath}}$ and $L_{\hat{\jmath}}$; it suffices to remove the $\hat{\jmath}$-th column and update the $\hat{\imath}$-th row.

For completeness, we include a brief complexity analysis; but as the timings in Table 4.1 show, computation time depends greatly on the density of the input points, with AA performing best when the density is low. With the improvement described in the previous paragraph the worst case complexity of AA is $O(s^2(n+s))$ arithmetic operations. The principal contributions to the complexity are $O(ns^2)$ for the creation of the matrix M in step AA2, $O(s^2)$ for

finding the minimum in step AA3 and $O(ns)$ in step AA5 to test the condition and also to update the matrix M; steps AA3 to AA5 are in a loop which may perform as many as s iterations. In the best case, no iterations are performed, and the complexity is just that of step AA2, namely $O(ns^2)$.

In the following example we apply the Agglomerative Algorithm to the points of Example 4.0.8 to show that the desired partition is obtained (see Figure 4.1).

Example 4.1.3. Let $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_{12}^\varepsilon\}$ be a set of empirical points with tolerance $\varepsilon = (1.43,\ 1.43)$, whose specified values coincide with the set $\mathbb{X}$ of Example 4.0.8:

$$\begin{aligned}\mathbb{X} \;=\; \{&(-1,-1),\ (0,-1),\ (1,-1),\ (-1,0),\ (0,0),\ (1,0),\\ &(-1,1),\ (0,1),\ (1,1),\ (5,-2.9),\ (5,0),\ (5,2.9)\}\end{aligned}$$

The AA computes, at each step, the following partitions, clustering together only the first nine points. We indicate in bold face the union created in step AA5.

1. $\mathcal{L} = \big\{\{p_1\}, \{p_2\}, \{p_3\}, \{p_4\}, \{p_5\}, \{p_6\}, \{p_7\}, \{p_8\}, \{p_9\}, \{p_{10}\}, \{p_{11}\}, \{p_{12}\}\big\}$
2. $\mathcal{L} = \big\{\{\mathbf{p_1}, \mathbf{p_2}\}, \{p_3\}, \{p_4\}, \{p_5\}, \{p_6\}, \{p_7\}, \{p_8\}, \{p_9\}, \{p_{10}\}, \{p_{11}\}, \{p_{12}\}\big\}$
3. $\mathcal{L} = \big\{\{\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_4}\}, \{p_3\}, \{p_5\}, \{p_6\}, \{p_7\}, \{p_8\}, \{p_9\}, \{p_{10}\}, \{p_{11}\}, \{p_{12}\}\big\}$
4. $\mathcal{L} = \big\{\{p_1, p_2, p_4\}, \{\mathbf{p_3}, \mathbf{p_6}\}, \{p_5\}, \{p_7\}, \{p_8\}, \{p_9\}, \{p_{10}\}, \{p_{11}\}, \{p_{12}\}\big\}$
5. $\mathcal{L} = \big\{\{p_1, p_2, p_4\}, \{p_3, p_6\}, \{\mathbf{p_5}, \mathbf{p_8}\}, \{p_7\}, \{p_9\}, \{p_{10}\}, \{p_{11}\}, \{p_{12}\}\big\}$
6. $\mathcal{L} = \big\{\{\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_4}, \mathbf{p_5}, \mathbf{p_8}\}, \{p_3, p_6\}, \{p_7\}, \{p_9\}, \{p_{10}\}, \{p_{11}\}, \{p_{12}\}\big\}$
7. $\mathcal{L} = \big\{\{\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}, \mathbf{p_4}, \mathbf{p_5}, \mathbf{p_6}, \mathbf{p_8}\}, \{p_7\}, \{p_9\}, \{p_{10}\}, \{p_{11}\}, \{p_{12}\}\big\}$
8. $\mathcal{L} = \big\{\{\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}, \mathbf{p_4}, \mathbf{p_5}, \mathbf{p_6}, \mathbf{p_7}, \mathbf{p_8}\}, \{p_9\}, \{p_{10}\}, \{p_{11}\}, \{p_{12}\}\big\}$
9. $\mathcal{L} = \big\{\{\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}, \mathbf{p_4}, \mathbf{p_5}, \mathbf{p_6}, \mathbf{p_7}, \mathbf{p_8}, \mathbf{p_9}\}, \{p_{10}\}, \{p_{11}\}, \{p_{12}\}\big\}$

4.1.2 The Divisive Algorithm

The Divisive Algorithm (DA) implements a "subdivision" method. The sets in the partition are determined by an iterative process. Initially the partition consists of a single set containing all the points. Then iteratively DA seeks the original point farthest from the centroid of its set. If the distance between them is below the tolerance threshold then the algorithm stops, because all original points are sufficiently well represented by the centroids of their sets. Otherwise it splits off the worst represented original point into a new set initially containing just itself. Then DA proceeds with a redistribuition phase with the aim of associating each original point to the current best representative subset (locally) minimizing the total central sum of squares, defined as follows [RR03].

Definition 4.1.4. Let $\mathbb{X}$ be a subset of $\mathbb{R}^n$ and let q be its centroid. The **central sum of squares** of $\mathbb{X}$ is defined to be:

$$\sum_{p\in\mathbb{X}} \|p-q\|^2$$

Definition 4.1.5. Let $\mathcal{L}=\{L_1,\ldots,L_k\}$ be a partition of the set $\mathbb{X}$. The **total central sum of squares** of the partition $\mathcal{L}$ is defined to be:

$$I(\mathcal{L})=\sum_{j=1}^{k} I_j$$

where I_j is the central sum of squares of L_j.

If $\mathbb{X}^\varepsilon$ contains many closely packed empirical points, DA turns out to be more efficient than AA, since only a few subdivisions are required.

Algorithm 4.1.6. (The Divisive Algorithm)
Let $\mathbb{X}^\varepsilon=\{p_1^\varepsilon,\ldots,p_s^\varepsilon\}$ be a set of empirical points, with each $p_i\in\mathbb{R}^n$ and a common tolerance $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)$. Let $\|\cdot\|_E$ be the weighted 2-norm on $\mathbb{R}^n$ w.r.t. $E=\mathrm{diag}(1/\varepsilon_1,\ldots,1/\varepsilon_n)$. Consider the following sequence of instructions.

DA1 Start with the list $\mathcal{L}=[L_1]$ where $L_1=\mathbb{X}$, and the centroid list $\mathbb{Y}=[q_1]$ where q_1 is the centroid of L_1.

DA2 Let $\mathcal{L}=[L_1,\ldots,L_r]$ and $\mathbb{Y}=[q_1,\ldots,q_r]$, the list of the centroids of the elements of $\mathcal{L}$. For each $p_i\in\mathbb{X}$ set $d_i=\|p_i-q_j\|_E$ where L_j is the subset (of $\mathbb{X}$) to which p_i belongs. Build the list $D=[d_1,\ldots,d_s]$.

DA3 If $\max(D)\le 1$ then return the pair $(\mathcal{L},\ \mathbb{Y})$, and stop.

DA4 Choose an index $\hat{\imath}$ such that $d_{\hat{\imath}}=\max(D)$, and compute the index $\hat{\jmath}$ of the subset $L_{\hat{\jmath}}$ to which $p_{\hat{\imath}}$ belongs. Remove p_i from $L_{\hat{\jmath}}$ and compute the new centroid $q_{\hat{\jmath}}$ of $L_{\hat{\jmath}}$; append $L_{r+1}=\{p_i\}$ to $\mathcal{L}$ and $q_{r+1}=p_i$ to $\mathbb{Y}$.

DA5 Compute the total central sum of squares $I(\mathcal{L})$ of the new partition $\mathcal{L}$.

DA6 For each $p\in\mathbb{X}$ and for each $L_k\in\mathcal{L}$, denote by $\mathcal{L}_{p,k}$ the partition $\mathcal{L}$ but with p moved into L_k. Compute each total central sum of squares $I(\mathcal{L}_{p,k})$.

DA7 Choose a point $\hat{p}\in\mathbb{X}$ and an index $\hat{k}$ s.t.

$$I(\mathcal{L}_{\hat{p},\hat{k}})=\min\{I(\mathcal{L}_{p,k}) : p\in\mathbb{X},\ L_k\in\mathcal{L}\}$$

DA8 If $I(\mathcal{L}_{\hat{p},\hat{k}})\ge I(\mathcal{L})$ then go to **DA2**. Otherwise set $\mathcal{L}=\mathcal{L}_{\hat{p},\hat{k}}$. Compute the centroids of the new partition $\mathcal{L}$. Go to **DA5**.

Theorem 4.1.7. *Algorithm 4.1.6 computes a pair* $(\mathcal{L},\ \mathbb{Y})$ *such that:*

- $\{L_i^\varepsilon : L_i \in \mathcal{L}\}$ *is a partition of* $\mathbb{X}^\varepsilon$ *into collapsable sets;*
- *for each* $q_i \in \mathbb{Y}$*, the empirical point* q_i^ε *is the valid representative of* L_i^ε*.*

Proof. Later on we shall refer to the loop DA5–DA8 as "the redistribution phase": points are moved from one subset to another in order to strictly decrease the total central sum of squares. Note that in the redistribution phase the cardinality of $\mathcal{L}$ does not change as the algorithm never eliminates any set in $\mathcal{L}$. Indeed, if the singleton set $L_j = \{p\}$ belongs to $\mathcal{L}$, the point p will not be moved to another set $L_k \in \mathcal{L}$ leaving L_j empty, since this new configuration cannot have smaller total central sum of squares: the combined central sum of squares of the sets $L_j = \{p\}$ and L_k is

$$I_j + I_k = 0 + \sum_{r \in L_k} \|r - q_k\|^2$$

where q_k is the centroid of L_k, whereas the combined central sum of squares of the new sets $L'_j = \emptyset$ and $L'_k = L_k \cup \{p\}$ is

$$I'_j + I'_k = 0 + \left(\|p - q'_k\|^2 + \sum_{r \in L_k} \|r - q'_k\|^2 \right)$$

where q'_k is the centroid of $L'_k = L_k \cup \{p\}$. And since q_k is the centroid of L_k, we have $\sum_{r \in L_k} \|r - q'_k\|^2 \geq \sum_{r \in L_k} \|r - q_k\|^2$. Consequently the new total central sum of squares cannot be smaller.

Now we prove finiteness. The algorithm comprises two nested loops: the outer loop spanning steps DA2–DA8, and the redistribution phase (steps DA5–DA8). The outer loop cannot perform more than s iterations because step DA4 can be performed at most s times; anyway, after s iterations the termination criterion in step DA3 will surely be satisfied as all the d_i would be zero.

The redistribution loop will perform only finitely many iterations. Each iteration strictly reduces the total central sum of squares, and since $\mathbb{X}$ is finite it has only finitely many partitions. Consequently there are only finitely many possible values for the total central sum of squares.

Next we show correctness. The elements of $\mathcal{L}$ define a partition of $\mathbb{X}$. This is trivially true in step DA1. The creation of a new subset in step DA4 clearly maintains the property. The redistribution phase merely moves points between subsets (in step DA8), so also preserves the property.

The test in step DA3 guarantees that upon completion of the algorithm each $L_i \in \mathcal{L}$ corresponds to a collapsable L_i^ε. By construction, each element $q_i \in \mathbb{Y}$ is the centroid of L_i. Thus q_i^ε is the valid representative of L_i^ε. □

For completeness, we include a brief complexity analysis; but as the timings in Table 4.1 show, computation time depends greatly on the density of

the input points, with DA performing best when the density is high. The algorithm contains two nested loops: DA2–DA8 and DA5–DA8. The outer loop can perform at most $O(s)$ iterations since each iteration increases the number of subsets in the partition. It seems to be tricky to bound the number of iterations the inner loop performs; based on experience, we conjecture that the inner loop performs $O(s)$ iterations. Now, steps DA6 and DA7 lie inside both loops and so are clearly dominant; their combined complexity is $O(ns^2)$ per iteration. Hence we obtain $O(ns^4)$ arithmetic operations as the worst case complexity for DA as a whole. In the best case, when no iterations are performed, the complexity is $O(ns)$.

In the following example we apply the Divisive Algorithm to the points of Example 4.0.8 to show that the desired partition is obtained (see Figure 4.1).

Example 4.1.8. Let $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_{12}^\varepsilon\}$ be a set of empirical points with tolerance $\varepsilon = (1.43,\ 1.43)$, whose specified values coincide with the set $\mathbb{X}$ of Example 4.0.8:

$$\begin{aligned}\mathbb{X} \;=\; \{&(-1,-1),\ (0,-1),\ (1,-1),\ (-1,0),\ (0,0),\ (1,0),\\ &(-1,1),\ (0,1),\ (1,1),\ (5,-2.9),\ (5,0),\ (5,2.9)\}\end{aligned}$$

The DA computes, at each step, after the redistribution phase, the following partitions. We use the bold face to indicate the newly sundered sets.

1. $\mathcal{L} = \big\{\{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}, p_{11}, p_{12}\}\big\}$
2. $\mathcal{L} = \big\{\{\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}, \mathbf{p_4}, \mathbf{p_5}, \mathbf{p_6}, \mathbf{p_7}, \mathbf{p_8}, \mathbf{p_9}\}, \{\mathbf{p_{10}}, \mathbf{p_{11}}, \mathbf{p_{12}}\}\big\}$
3. $\mathcal{L} = \big\{\{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9\}, \{\mathbf{p_{10}}\}, \{\mathbf{p_{11}}, \mathbf{p_{12}}\}\big\}$
4. $\mathcal{L} = \big\{\{p_1, p_2, p_4, p_5, p_8, p_3, p_6, p_7, p_9\}, \{p_{10}\}, \{\mathbf{p_{11}}\}, \{\mathbf{p_{12}}\}\big\}$

As mentioned before, DA performs fewer iterations than AA since many input points are close together w.r.t. the tolerance.

4.1.3 A particularly quick method: the Grid Algorithm

We recall the ∞-norm and its corresponding E-weighted norm on $\mathbb{R}^n$ (see also [DBA74]):

$$\|v\|_\infty = \max_{i=1\ldots n} |v_i| \quad \text{and} \quad \|v\|_{\infty,E} = \|Ev\|_\infty$$

where $E = \operatorname{diag}(1/\varepsilon_1, \ldots, 1/\varepsilon_n)$, as before.

A particularly quick method for decreasing the cardinality of the set $\mathbb{X}^\varepsilon$ can be designed using a regular grid, consisting of half-open balls of radius $1/2$ w.r.t. the E-weighted norm $\|\cdot\|_{\infty,E}$. We arbitrarily choose one ball to have the origin as its centre then tessellate to cover the whole space; note that the balls are actually cuboids. We shall use the notation $[z]$ to mean the closest integer to z (rounding up in the case of a tie).

Algorithm 4.1.9. (The Grid Algorithm)
Let $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_s^\varepsilon\}$ be a set of empirical points, with each $p_i \in \mathbb{R}^n$ and a common tolerance $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$. Consider the following sequence of instructions.

GA1 Create the set of balls $B = \{b(p_1), \ldots, b(p_s)\}$ where $b(x_1, \ldots, x_n)$ is the ε-ball centred on $(\varepsilon_1[x_1/\varepsilon_1], \ldots, \varepsilon_n[x_n/\varepsilon_n])$, that is the grid ball containing the point $(x_1, \ldots, x_n)$.

GA2 For each grid ball $g \in B$, create the subset L_g containing exactly those p_i for which $b(p_i) = g$.

GA3 Compute the list $\mathbb{Y} = [q_1, \ldots, q_t]$ of the centroids of the L_g. Return the pair $(\mathcal{L}, \mathbb{Y})$, where $\mathcal{L} = [L_1, \ldots, L_t]$.

Theorem 4.1.10. *Algorithm 4.1.9 computes a pair* $(\mathcal{L}, \mathbb{Y})$ *such that:*

- $\{L_i^\varepsilon : L_i \in \mathcal{L}\}$ *is a partition of* $\mathbb{X}^\varepsilon$*;*
- *each* $q_i \in \mathbb{Y}$ *gives a good representative* q_I^ε *of* L_i^ε*.*

This algorithm computes a partition of $\mathbb{X}^\varepsilon$ by gathering all the empirical points whose specified values lie in the same ball into the same subset. Suppose that one of these subsets comprises the empirical points $p^\varepsilon{}_1, \ldots, p^\varepsilon{}_m$, and let q^ε be their empirical centroid, then q^ε is a "good" representative of each $p^\varepsilon{}_i$ because

$$\|p_i - q\|_{\infty,E} = \left\| p_i - \frac{1}{m}\sum_{j=1}^{m} p_j \right\|_{\infty,E} \leq \frac{1}{m}\sum_{j=1}^{m} \|p_i - p_j\|_{\infty,E} < 1$$

However, in general such a subset is not collapsable, a notion defined in terms of the 2-norm.

Note that, since the separations of the empirical points are ignored by this method, unsatisfactory partitions can be obtained, *e.g.* close points may happen to belong to different balls and so be assigned to different subsets in the partition. Nevertheless, this drawback is compensated by the speed and simplicity of the method. In particular, this grid method (with a smaller radius) can be used to reduce the bulk of a very large body of dense data before applying one of the more sophisticated but slower algorithms, *i.e.* AA or DA. Another application of the grid method is to help choose the more suitable algorithm between AA and DA by estimating the numbers of sets in the partitions which would be produced. The complexity of GA is $O(ns \log s)$ if we sort the list B and the set of input points $\mathbb{X}$.

4.2 Relationship with Cluster Analysis

The idea of analyzing a large body of empirical data and of partitioning it into sets of "similar values" has been well studied in the theory of Cluster Analysis

(*e.g.* see [KM00]). The overall aim of Cluster Analysis is to separate the original data into clusters where the members of each cluster are much more similar to each other than to members of other clusters. In contrast, our methods are more concerned with thinning out groups of very close values while ignoring more distant points. Below we show how Ward's "classical" algorithm [RR03], an agglomerative hierarchical method, and Li's more recent algorithm [Li06], a divisive hierarchical method, partition the empirical points of Example 4.0.8.

Example 4.2.1. Let $\mathbb{X}^\varepsilon$ be the set of empirical points whose set of specified values is given in Example 4.0.8; similarly, let $\varepsilon = (1.43, 1.43)$ as given there. We recall that in Examples 4.1.3 and 4.1.8 both our algorithms AA and DA obtained the minimal partition into collapsable sets, as illustrated in Figure 4.1.

Ward's and Li's algorithms do not obtain this minimal partition. In fact, after 8 steps, Ward's algorithm puts the points $(5, -2.9)$ and $(5, 0)$ into the same cluster, while the first nine points of $\mathbb{X}$ still belong to different clusters. Since this is an agglomerative method no set of points is split during the computation, so Ward's algorithm fails to recognise the collapsable set of nine points. In a similar vein, Li's algorithm goes astray at the third step: it divides the first nine points of $\mathbb{X}$ into two subsets while the points $(5, -2.9)$ and $(5, 0)$ still belong to the same cluster. Since this is a hierarchical divisive method, once a set is split it can never be joined together again, so Li's algorithm needlessly splits the collapsable set of nine points.

Now we consider another method of Cluster Analysis, the QT Clustering (see [HKY99]), because it has a number of similarities to our methods, especially AA. QT Clustering computes a partition of the input data using a given limit on the diameter of the clusters. It works by building clusters according to their cardinality, while we are primarily interested in the local geometrical separations of the input data.

Example 4.2.2. Let $\mathbb{X}^\varepsilon$ be a set of empirical points with tolerance $\varepsilon = (0.5)$ and with specified values $\mathbb{X} = \{0,\ 0.05,\ 0.9,\ 1,\ 1.2\} \subseteq \mathbb{R}$. Applying the QT Clustering algorithm with maximum cluster diameter equal to 2ε, we obtain the partition $\big\{\{0, 0.05, 0.9, 1\},\ \{1.2\}\big\}$ where $\{0, 0.05, 0.9, 1\}^\varepsilon$ is a not collapsable set. In contrast, if we apply AA or DA to $\mathbb{X}^\varepsilon$, we obtain the more balanced partition $\big\{\{0, 0.05\},\ \{0.9, 1, 1.2\}\big\}$ whose elements consist of specified values of collapsable sets. We maintain that our partition is more plausible as a grouping of noisy data.

4.3 Numerical tests and illustrative examples

In this section we present some numerical examples to show the effectiveness and the potential of our techniques. Both AA and DA have been implemented using the C++ language, and are included in CoCoALib [CoC] (see also [Abb06]). All computations in the following examples have been performed on an Intel

Pentium M735 processor (at 1.7 GHz) running GNU/Linux and using the implementation in CoCoALib. For simplicity of presentation, the data in the following artificial examples are prescaled so that the tolerance is isotropic, *i.e.* all the ε_i are equal.

Example 4.3.1. Clouds of empirical points.
In this example we consider an empirical set $\mathbb{X}^\varepsilon$ containing two well separated empirical points and three clusters, two big and one small. Both AA and DA compute five valid representatives for $\mathbb{X}^\varepsilon$, but because the result comprises very few points DA is faster than AA.
Let $\mathbb{X}^\varepsilon$ be a set of empirical points, with tolerance $\varepsilon = (20, 20)$ and specified values $\mathbb{X} = \cup_{i=1}^5 \mathbb{X}_i \subseteq \mathbb{R}^2$, where

- $\mathbb{X}_1$ consists of 82 points lying inside the disk of radius 10 centered on $(0, 0)$,
- $\mathbb{X}_2$ consists of 64 points lying inside the disk of radius 10 centered on $(40, 50)$,
- $\mathbb{X}_3 = \{(49, 0), (50, 0), (50, 1)\}$, $\mathbb{X}_4 = \{(9, 41)\}$ and $\mathbb{X}_5 = \{(-10, 80)\}$.

Both AA and DA compute the "intuitive" partition consisting of 5 subsets $L_i = \mathbb{X}_i$ for $i = 1, \ldots, 5$, as shown in Figure 4.2.

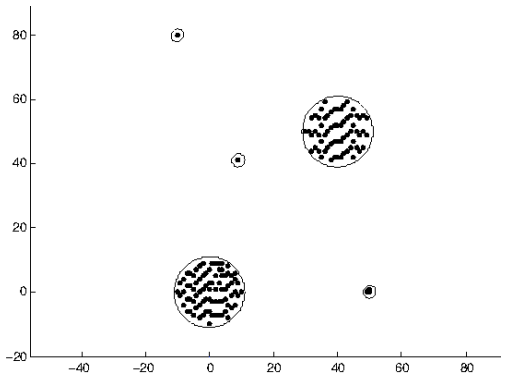


Figure 4.2: Appropriate partition of $\mathbb{X}$

Example 4.3.2. Empirical points close to a circle
In this example we compare the behaviour of AA and DA on a family of artificial test cases, comprising sets of empirical points with similar geometrical configurations but with differing "densities". Let $\mathbb{X}_1$, $\mathbb{X}_2 \subset \mathbb{R}^2$ be two sets of points lying close to the circle of radius 200 and centered at the origin. They contain 2504 and 5032 points, respectively. The numerical tests are performed by applying both AA and DA to the empirical sets $\mathbb{X}_1^\varepsilon$ and $\mathbb{X}_2^\varepsilon$ for various (isotropic) values of ε: viz. $\varepsilon_1 = 2^k$ for $k = 0, \ldots, 6$, and note that for a fixed set of points increasing ε effectively increases the density of the points.

In Table 4.1 we summarize the results obtained from processing $\mathbb{X}_1$ and $\mathbb{X}_2$ respectively. The first column contains the value of the tolerance, the columns labeled with "#VR" contain the number of the valid representatives computed by AA and DA respectively, while those labeled with "Time" show the timings (in seconds) of each algorithm. The results show that DA runs quickly if ε is large, that is when the set of empirical points is dense enough, since only a few splittings of the original set are needed. On the other hand, when the points are well separated, AA is preferable since the final partition consists of a large number of sets.

	2504 empirical points				**5032 empirical points**			
	AA		**DA**		**AA**		**DA**	
ε	#VR	Time	#VR	Time	#VR	Time	#VR	Time
1	911	1 s	727	293 s	2096	6 s	1460	2306 s
2	462	3 s	347	184 s	734	31 s	587	1250 s
4	224	8 s	173	114 s	263	118 s	185	577 s
8	108	18 s	87	66 s	121	317 s	86	314 s
16	56	50 s	41	33 s	61	733 s	41	166 s
32	29	117 s	20	15 s	28	1680 s	21	79 s
64	13	2633 s	10	6 s	14	3695 s	10	25 s

Table 4.1: Points close to a circle

Figure 4.3 shows a subset of $\mathbb{X}_1$ (the crosses) and its valid representatives (the dots) w.r.t. the tolerance $\varepsilon = (16, 16)$.

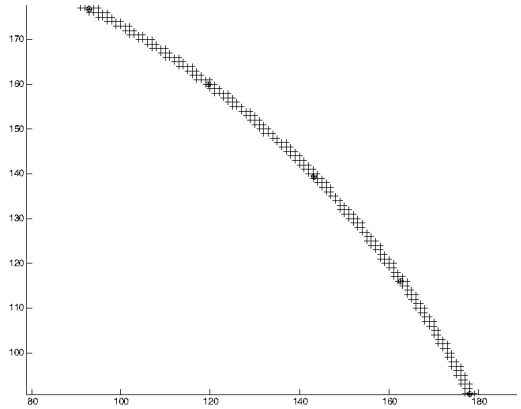


Figure 4.3: Valid representatives of $\mathbb{X}_1$

The computational timings can be drastically reduced if we apply GA (see Section 4.1.3) before applying AA or DA. Let us consider two cases where computation time was high: AA with $\varepsilon = 64$, and DA with $\varepsilon = 2$. In the case AA with $\varepsilon = 64$, we make a first reduction of the data using a grid whose balls have a weighted radius of 1/4; the computation takes 0.14 seconds and

produces 48 points. Now AA is applied to this result, and produces an output of 13 points in 0.01 seconds — overall far faster than applying AA directly, but the final result is less accurate.
Analogous remarks hold for the test with DA and $\varepsilon = 2$: using a grid whose balls have a weighted radius of $1/2$ we obtain 1657 points in 0.2 seconds; then the execution of DA on this output takes 83 seconds to return 466 points. Once again, a drastic reduction in time at the cost of a lower quality result.

Example 4.3.3. Example of the "zip"
This first example illustrates the necessity of the test at Step AA5. Indeed, if the condition is not checked the algorithm builds a partition consisting of not collapsable sets.

Let $\mathbb{X}^\varepsilon$ be a set of empirical points whose tolerance is $\varepsilon = (2.199, 2.199)$ and whose set of specified values $\mathbb{X} \subseteq \mathbb{R}^2$ is given by:

$$\mathbb{X} = \{(0.1, 2),\ (2, 0),\ (4.2, 0),\ (6.4, 0),\ (8.6, 0),\ (3.1, 3)\ (5.3, 3),\ (7.5, 3)\}$$

Applying AA to $\mathbb{X}^\varepsilon$ we obtain the following partition

$$\big\{\{(0.1, 2), (3.1, 3)\}, \{(2, 0), (4.2, 0)\}, \{(6.4, 0), (8.6, 0)\}, \{(5.3, 3), (7.5, 3)\}\big\}$$

for which the set of specified values of the valid representatives is

$$\mathbb{Y} = \{(1.6, 2.5),\ (3.1, 0),\ (7.5, 0),\ (6.4, 3)\}$$

In Figure 4.4 we represent the points of $\mathbb{X}$ and $\mathbb{Y}$ using big and small dots respectively.

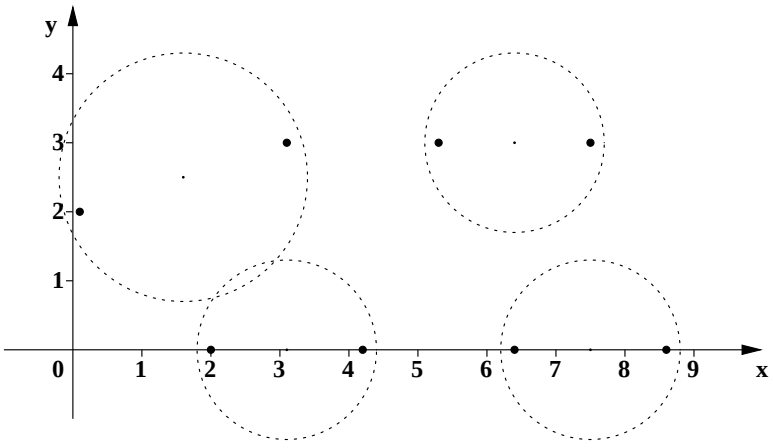


Figure 4.4: Example of the "zip"

However, if we check only the distance between the centroids in step AA5, all the elements of $\mathbb{X}^\varepsilon$ are placed in a single set which is obviously not collapsable.

Example 4.3.4. Example of the "three-pointed star"
We have seen that AA always produces a partition into collapsable sets such that no pair can be unified into a collapsable set. In most cases the partition produced by DA also enjoys this property; however, this is not true in general. Such a situation is shown in this example.

Let $\mathbb{X}^\varepsilon$ be a set of 6 empirical points whose tolerance is $\varepsilon = (1, 1)$ and whose set of specified values $\mathbb{X} \subseteq \mathbb{R}^2$ is given by:

$$\begin{aligned}\mathbb{X} &= \{(0.577, 0.99),\ (0.577, -0.99),\ (0, 0.0001),\ (0, 0),\\ &\quad (-1.1551, 0),\ (-1.155, 0)\}\end{aligned}$$

Applying both AA and DA we obtain the two different partitions $\mathcal{L}_A$ and $\mathcal{L}_D$:

$$\begin{aligned}\mathcal{L}_A &= \big\{\{(0.577, -0.99)\},\\ &\quad \{(0.577, 0.99), (0, 0.0001), (0, 0)\},\\ &\quad \{(-1.1551, 0), (-1.155, 0)\}\big\}\\ \mathcal{L}_D &= \big\{\{(0.577, -0.99)\},\\ &\quad \{(0.577, 0.99)\},\\ &\quad \{(0, 0.0001), (0, 0), (-1.1551, 0), (-1.155, 0)\}\big\}\end{aligned}$$

associated to the valid representatives whose specified values are

$$\begin{aligned}\mathbb{Y}_A &= \{(0.577, -0.99), (0.192333, 0.330033), (-1.15505, 0)\}\\ \mathbb{Y}_D &= \{(0.577, 0.99), (0.577, -0.99), (-0.577525, 0.000025)\}\end{aligned}$$

respectively. To highlight the different partitions obtained, in Figure 4.5 we plot the points of $\mathbb{X}$ and $\mathbb{Y}_A$ using big and small dots; in Figure 4.6 we use the same symbols for $\mathbb{X}$ and $\mathbb{Y}_D$.

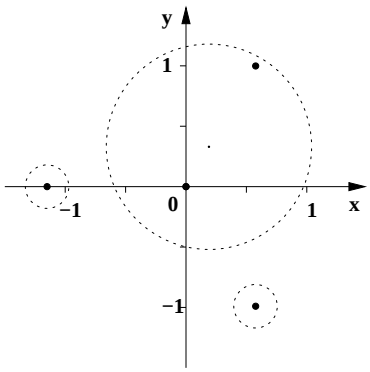


Figure 4.5: Representation of the sets $\mathbb{X}$ and $\mathbb{Y}_A$

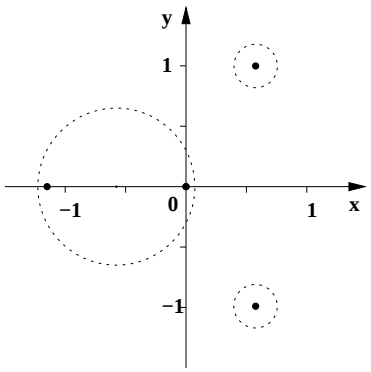


Figure 4.6: Representation of the sets $\mathbb{X}$ and $\mathbb{Y}_D$

It is trivial to verify that the elements of $\mathcal{L}_A^\varepsilon$ are pairwise not unifiable into a collapsable set, while the same property does not hold for the partition $\mathcal{L}_D^\varepsilon$ since $\{(0.577, -0.99)^\varepsilon\} \cup \{(0.577, 0.99)^\varepsilon\}$ is a collapsable set.

Example 4.3.5. Example with experimental data
In this example we use a set $\mathbb{X}^{\varepsilon}$ of 1000 empirical points in $\mathbb{R}^2$ which is a subset of a time series collected by A. Jessup, (Applied Physics Laboratory, University of Washington), and others and described in [JMK91] (with permission). This time series records the height of ocean waves as a function of time, measured via an infrared wave gauge.

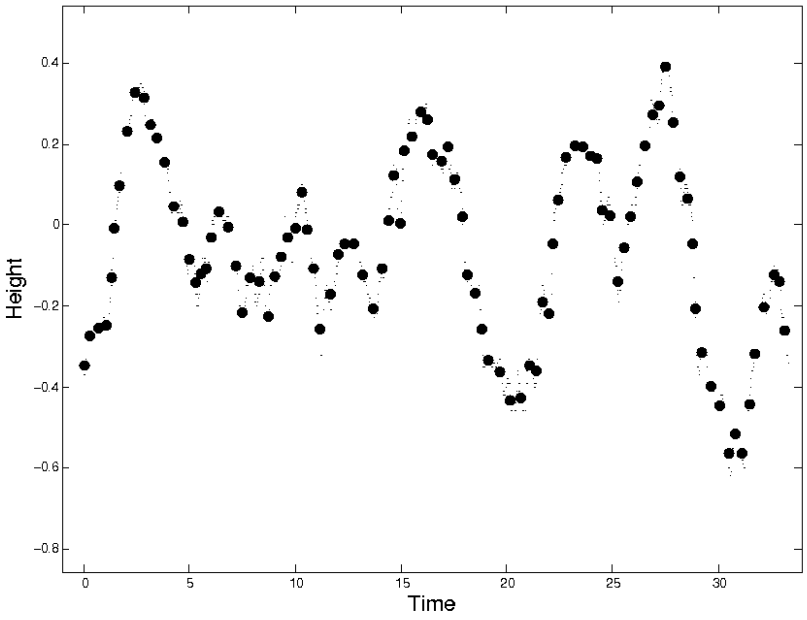


Figure 4.7: Valid representatives (99 points) of wave data

We applied both AA and DA (with anisotropic tolerance $\varepsilon = (0.25,\ 0.1)$) to $\mathbb{X}$, and obtained two different partitions having respectively 99 and 90 valid

representatives. In this case the data is relatively sparse, so AA proves to be faster than DA (taking 1.3 vs. 9.4 seconds). As AA was considerably faster we selected its output to be illustrated in Figure 4.7 together with the original points (we use small dots for the original data, large dots for the output of AA). Qualitatively the output of DA is very similar to that produced by AA.

Chapter 5

A global characterization of a set of empirical points

This chapter is based on [AFT08] and aims at giving a global characterization of the vanishing ideals $\mathcal{I}(\widetilde{\mathbb{X}})$, where $\widetilde{\mathbb{X}}$ is any admissible perturbation of a set $\mathbb{X}^\varepsilon$ of empirical points, independently of the data uncertainty.

Since any admissible perturbation $\widetilde{\mathbb{X}}$ is made up of points differing by less than the uncertainty from the corresponding elements of $\mathbb{X}$, then any set $\widetilde{\mathbb{X}}$ can be considered equivalent to any other admissible perturbation of $\mathbb{X}^\varepsilon$, in particular to $\mathbb{X}$. Nevertheless, given two distinct admissible perturbations $\widetilde{\mathbb{X}}_1$ and $\widetilde{\mathbb{X}}_2$ of $\mathbb{X}^\varepsilon$, it can happen that their affine coordinate rings $P/\mathcal{I}(\widetilde{\mathbb{X}}_1)$ and $P/\mathcal{I}(\widetilde{\mathbb{X}}_2)$ as well as their vanishing ideals $\mathcal{I}(\widetilde{\mathbb{X}}_1)$ and $\mathcal{I}(\widetilde{\mathbb{X}}_2)$ have very different bases – this is a well known phenomenon in Gröbner basis theory. In order to "emphasize" the numerical equivalence among the admissible perturbations $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$, we look for a common characterization of $P/\mathcal{I}(\widetilde{\mathbb{X}})$ and, when possible, we give a structurally stable representation of the vanishing ideals $\mathcal{I}(\widetilde{\mathbb{X}})$. More precisely our goal is to determine a basis $\mathcal{O}$ of the quotient ring $P/\mathcal{I}(\widetilde{\mathbb{X}})$ valid for every admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$. Unfortunately, not every quotient basis $\mathcal{O}$ for $\mathcal{I}(\mathbb{X})$ exhibits this good numerical behaviour. We consider the following example.

Example 5.0.6. Let $\mathbb{X}^\varepsilon$ be the set of empirical points having

$$\mathbb{X} = \{(-1,-5),\ (0,-2),\ (1,1),\ (2,4.1)\} \subseteq \mathbb{R}^2$$

as the set of specified values and $\varepsilon = (3/20, 3/20)$ as the tolerance; let $\|\cdot\|_E$ be the weighted 2-norm on $\mathbb{R}^2$ w.r.t. $E = \mathrm{diag}(20/3, 20/3)$. Consider the order ideal $\mathcal{O} = \{1, y, x, y^2\}$: since the evaluation matrix $M_{\mathcal{O}}(\mathbb{X})$ of $\mathcal{O}$ associated to $\mathbb{X}$,

$$M_{\mathcal{O}}(\mathbb{X}) = \begin{pmatrix} 1 & -5 & -1 & 25 \\ 1 & -2 & 0 & 4 \\ 1 & 1 & 1 & 1 \\ 1 & 4.1 & 2 & 16.81 \end{pmatrix}$$

is non singular, then $\mathcal{O}$ is a quotient basis of $\mathcal{I}(\mathbb{X})$.

Now consider the slightly perturbed set of points $\widetilde{\mathbb{X}}$ (which is indeed an admissible perturbation of $\mathbb{X}^\varepsilon$):

$$\widetilde{\mathbb{X}} = \{(-1,-5),\ (0,-2),\ (1,1),\ (2,4)\}$$

In this case, the order ideal $\mathcal{O}$ does not form a vector space basis of $P/\mathcal{I}(\widetilde{\mathbb{X}})$ as the evaluation matrix

$$M_{\mathcal{O}}(\widetilde{\mathbb{X}}) = \begin{pmatrix} 1 & -5 & -1 & 25 \\ 1 & -2 & 0 & 4 \\ 1 & 1 & 1 & 1 \\ 1 & 4 & 2 & 16 \end{pmatrix}$$

is singular. We conclude observing that a *small* (admissible) change in the coordinates of the points of $\mathbb{X}$ has led to a *drastic* change in the associated vector space basis of $P/\mathcal{I}(\mathbb{X})$.

Example 5.0.6 suggests that, when dealing with a set $\mathbb{X}^\varepsilon$ of empirical points, a notion of "numerically stable" quotient basis is necessary: a quotient basis $\mathcal{O}$ for $\mathcal{I}(\mathbb{X})$ is stable if exhibits good numerical behaviour, that is if its residue classes form a vector space basis of $P/\mathcal{I}(\widetilde{\mathbb{X}})$, where $\widetilde{\mathbb{X}}$ is any admissible perturbation of the empirical set $\mathbb{X}^\varepsilon$ (see Definition 5.1.3). Stable quotient bases provide a common characterization of the ideals $\mathcal{I}(\mathbb{X})$ and $\mathcal{I}(\widetilde{\mathbb{X}})$, highlight the geometrical properties of the empirical set $\mathbb{X}^\varepsilon$ and, by using the border basis theory, guarantee the existence of a structurally stable representation of $\mathcal{I}(\mathbb{X})$.

Border bases appeared for the first time in connection with problems arising in numerical analysis during the 1980s, thanks to the work of Hans J. Stetter (see [AS88] and [Ste04]); then, during the 1990s, the importance of these results for computer algebra was pointed out by H. Michael Möller (see [MS95] and [Möl93]). In 1999 the first algebraic properties of border bases were presented by B. Mourrain (see [Mou99]). In 2005, A. Kehrein, M. Kreuzer and L. Robbiano wrote a survey devoted to laying the algebraic foundations of the border basis theory for the zero-dimensional ideals (see [KKR05] and [KR05]). Recently, M. Kreuzer and L. Robbiano (see [KR08]), and later L. Robbiano (see [Rob08]) examined a natural link between border bases and Hilbert schemes which provides a futher improvement to the solid mathematical foundations of the border basis theory.

Our decision to use border bases for describing the vanishing ideal $\mathcal{I}(\mathbb{X})$ is based on two main reasons: firstly, the works mentioned above certify border bases as a good tool for dealing with numerical problems; secondly, border bases are easy to compute since, once a quotient basis $\mathcal{O}$ is fixed, the corresponding border basis can be obtained by simple combinatorical and linear algebra computations; so, in the empirical frame we focus our attention on determining a stable quotient basis $\mathcal{O}$.

Unfortunately, as we will show in Example 5.1.5, stable quotient bases (and consequently stable border bases) do not always exist: in such cases we turn to

the wider notion of stable order ideal given in Definition 5.1.1. In fact, though stable order ideals do not define a monomial basis of the vector space $P/\mathcal{I}(\mathbb{X})$, they nevertheless provide $P/\mathcal{I}(\widetilde{\mathbb{X}})$ with a common structure, for any admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$, and thus give information on the geometrical configuration of the original points.

This chapter is organized as follows: in Section 5.1 we introduce the concept of numerical stability for structures associated to a set of empirical points, namely stable order ideals, stable quotient bases and stable border bases; then, in Section 5.2 we present a theoretical method and a practical algorithm for computing them.

5.1 Stable structures for $\mathbb{X}^\varepsilon$

For the rest of this chapter we let $n \geq 1$, P be the polynomial ring $\mathbb{R}[x_1, \ldots, x_n]$, and $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_s^\varepsilon\}$ be a finite set of distinct empirical points with specified values $\mathbb{X} = \{p_1, \ldots, p_s\} \subseteq \mathbb{R}^n$ and tolerance $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$, where each $\varepsilon_i \in \mathbb{R}^+$. We let $E = \operatorname{diag}(1/\varepsilon_1, \ldots, 1/\varepsilon_n)$ and use the weighted 2-norm $\| \cdot \|_E$ on $\mathbb{R}^n$.

5.1.1 Stable order ideals

Definition 5.1.1. Let $\mathcal{O}$ be an order ideal of $\mathbb{T}^n$, then $\mathcal{O}$ is **stable** w.r.t. $\mathbb{X}^\varepsilon$ if the evaluation matrix $M_\mathcal{O}(\widetilde{\mathbb{X}})$ has full rank for each admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$.

Given any finite set $\mathbb{X}^\varepsilon$ of empirical points, an order ideal $\mathcal{O}$ stable w.r.t. $\mathbb{X}^\varepsilon$ always exists: in fact, the trivial order ideal $\mathcal{O} = \{1\}$ is stable w.r.t. any empirical set $\mathbb{X}^\varepsilon$. In the following example we show a case of a *non*-stable order ideal.

Example 5.1.2. Let $\mathbb{X}^\varepsilon$ be the set of empirical points given in Example 5.0.6. Consider the order ideal $\mathcal{O} = \{1, y, x\}$: it is easy to prove that $M_\mathcal{O}(\mathbb{X})$ is a full rank matrix. Nevertheless $\mathcal{O}$ is not stable w.r.t. $\mathbb{X}^\varepsilon$: in fact, let

$$\widetilde{\mathbb{X}} = \{(-1, -5),\ (0, -2),\ (1, 1),\ (2, 4)\}$$

be an admissible perturbation of $\mathbb{X}^\varepsilon$. Now, $\operatorname{rank}(M_\mathcal{O}(\widetilde{\mathbb{X}})) = 2$, and so the order ideal $\mathcal{O}$ is not stable w.r.t. $\mathbb{X}^\varepsilon$.

5.1.2 Stable quotient bases

We specialize to quotient bases the notion of stability given in Definition 5.1.1 for simple order ideals.

Definition 5.1.3. Let $\mathcal{O}$ be a quotient basis for $\mathcal{I}(\mathbb{X})$, then $\mathcal{O}$ is **stable** w.r.t. $\mathbb{X}^\varepsilon$ if the evaluation matrix $M_\mathcal{O}(\widetilde{\mathbb{X}})$ is non singular for each admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$.

Our assumption to deal with sets of distinct empirical points is due to the following result.

Proposition 5.1.4. *Let $\mathbb{X}^\varepsilon$ be a finite set of empirical points of $\mathbb{R}^n$, not necessarily distinct, and let $\mathcal{O}$ be a quotient basis of $\mathcal{I}(\mathbb{X})$. If the set $\mathbb{X}^\varepsilon$ contains at least a pair of non-distinct empirical points then $\mathcal{O}$ is non-stable w.r.t. $\mathbb{X}^\varepsilon$.*

Proof. Let $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_s^\varepsilon\}$, and $\mathbb{X} = \{p_1, \ldots, p_s\} \subseteq \mathbb{R}^n$ be its set of specified values; let $\mathcal{O} = \{t_1, \ldots, t_s\} \subseteq \mathbb{T}^n$ be a quotient basis of $\mathcal{I}(\mathbb{X})$. Let $p_i^\varepsilon, p_j^\varepsilon$, with $i \neq j$, be two non-distinct empirical points of $\mathbb{X}^\varepsilon$. Since $N^2(p_i^\varepsilon) \cap N^2(p_j^\varepsilon) \neq \emptyset$ there exist admissible perturbations $\widetilde{p}_i$ of p_i and $\widetilde{p}_j$ of p_j such that $\widetilde{p}_i = \widetilde{p}_j$. As a consequence the evaluation matrix of $\mathcal{O}$ at $\widetilde{\mathbb{X}} = \{\widetilde{p}_1, \ldots, \widetilde{p}_i, \ldots, \widetilde{p}_j, \ldots, \widetilde{p}_s\}$ is singular. □

However note that there are sets of distinct empirical points $\mathbb{X}^\varepsilon$ for which the vanishing ideal $\mathcal{I}(\mathbb{X})$ has no stable quotient basis.

Example 5.1.5. Let $P = \mathbb{R}[x, y]$, and let $\mathbb{X}^\varepsilon$ be the set of empirical points having

$$\mathbb{X} = \{(0, 0),\ (-5, 12),\ (12, -5)\}$$

as the set of specified values and $\varepsilon = (2.51, 2.51)$ as the tolerance. It is simple to verify that the points of $\mathbb{X}^\varepsilon$ are distinct. There are only 3 order ideals in $\mathbb{T}^2$ containing exactly 3 elements:

$$\mathcal{O}_1 = \{1, x, x^2\}, \quad \mathcal{O}_2 = \{1, x, y\} \quad \text{and} \quad \mathcal{O}_3 = \{1, y, y^2\}$$

We prove that none of them is a quotient basis for $\mathcal{I}(\mathbb{X})$ stable w.r.t. $\mathbb{X}^\varepsilon$. In fact, let $\widetilde{\mathbb{X}}_1, \widetilde{\mathbb{X}}_2, \widetilde{\mathbb{X}}_3$ be 3 admissible perturbations of $\mathbb{X}^\varepsilon$:

$$\begin{aligned}
\widetilde{\mathbb{X}}_1 &= \Big\{\Big(-\frac{5}{2}, 0\Big),\ \Big(-\frac{5}{2}, 12\Big),\ \Big(-12, 5\Big)\Big\} \\
\widetilde{\mathbb{X}}_2 &= \Big\{\Big(\frac{7}{4}, \frac{7}{4}\Big),\ \Big(-\frac{27}{4}, \frac{41}{4}\Big),\ \Big(\frac{41}{4}, -\frac{27}{4}\Big)\Big\} \\
\widetilde{\mathbb{X}}_3 &= \Big\{\Big(0, -\frac{5}{2}\Big),\ \Big(-5, 12\Big),\ \Big(12, -\frac{5}{2}\Big)\Big\}
\end{aligned}$$

The evaluation matrices $M_{\mathcal{O}_1}(\widetilde{\mathbb{X}}_1)$ and $M_{\mathcal{O}_3}(\widetilde{\mathbb{X}}_3)$ are Vandermonde matrices, so it is easy to verify that $\det(M_{\mathcal{O}_1}(\widetilde{\mathbb{X}}_1)) = \det(M_{\mathcal{O}_3}(\widetilde{\mathbb{X}}_3)) = 0$. Further

$$\det\Big(M_{\mathcal{O}_2}(\mathbb{X}_2)\Big) = \det\begin{pmatrix} 1 & 7/4 & 7/4 \\ 1 & -27/4 & 41/4 \\ 1 & 41/4 & -27/4 \end{pmatrix} = 0$$

This concludes the proof.

We end this section by observing that, given any finite set of points $\mathbb{X}$, any quotient basis $\mathcal{O}$ for $\mathcal{I}(\mathbb{X})$ is stable w.r.t. $\mathbb{X}^\delta$ for a sufficiently small value of the tolerance δ; this is equivalent to saying that $\mathcal{O}$ has a "region of stability" w.r.t. $\mathbb{X}$, as the following proposition shows.

Proposition 5.1.6. *Let $\mathbb{X}$ be a finite set of points of $\mathbb{R}^n$ and $\mathcal{I}(\mathbb{X})$ be its vanishing ideal; let $\mathcal{O}$ be a quotient basis for $\mathcal{I}(\mathbb{X})$. Then there exists a tolerance $\delta = (\delta_1, \ldots, \delta_n)$, with each $\delta_i > 0$, such that $\mathcal{O}$ is stable w.r.t. $\mathbb{X}^\delta$.*

Proof. Let $M_{\mathcal{O}}(\mathbb{X})$ be the evaluation matrix of $\mathcal{O}$ associated to the set $\mathbb{X}$; then $M_{\mathcal{O}}(\mathbb{X})$ is a structured matrix whose ntries depend continuously on the points in $\mathbb{X}$. Since, by hypothesis, $\mathcal{O}$ is a quotient basis for $\mathcal{I}(\mathbb{X})$, it follows that $M_{\mathcal{O}}(\mathbb{X})$ is invertible. Recalling that the determinant is a polynomial function in the matrix entries, and noting that the entries of $M_{\mathcal{O}}(\mathbb{X})$ are polynomials in the points' coordinates, we can conclude that there exists a tolerance $\delta = (\delta_1, \ldots, \delta_n)$, with each $\delta_i > 0$, such that $\det(M_{\mathcal{O}}(\widetilde{\mathbb{X}})) \neq 0$ for any perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\delta$. □

Nevertheless, since the tolerance ε of the empirical points in $\mathbb{X}^\varepsilon$ is given *a priori* by the measurements, Proposition 5.1.6 does not give a direct answer to the problem of stability. If the given tolerance ε on the points is larger than the "region of stability" of a chosen quotient basis $\mathcal{O}$, then $\mathcal{O}$ will not be stable w.r.t. $\mathbb{X}^\varepsilon$; such a situation is shown in the following example.

Example 5.1.7. Let $\mathbb{X}^\varepsilon$ be the set of empirical points given in Example 5.0.6; Consider the order ideal $\mathcal{O}_1 = \{1, y, x, y^2\}$, which is a quotient basis for $\mathcal{I}(\mathbb{X})$; $\mathcal{O}_1$ is not stable w.r.t. $\mathbb{X}^\varepsilon$. Indeed, consider again the perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$

$$\widetilde{\mathbb{X}} = \{(-1, -5),\ (0, -2),\ (1, 1),\ (2, 4)\}$$

The evaluation matrix $M_{\mathcal{O}_1}(\widetilde{\mathbb{X}})$ is singular, so $\mathcal{O}_1$ is not stable w.r.t. $\mathbb{X}^\varepsilon$ since its "region of stability" is too small w.r.t. the given tolerance ε.

Now consider the quotient basis $\mathcal{O}_2 = \{1, y, y^2, y^3\}$, and let

$$\widetilde{\mathbb{X}} = \{(-1 + e_1, -5 + e_2),\ (e_3, -2 + e_4),\ (1 + e_5, 1 + e_6),\ (2 + e_7, 4.1 + e_8)\}$$

be a generic admissible perturbation of $\mathbb{X}^\varepsilon$, where the parameters $e_i \in \mathbb{R}$ satisfy

$$\|(e_1, e_2)\|_E \leq 1 \qquad \|(e_3, e_4)\|_E \leq 1 \qquad \|(e_5, e_6)\|_E \leq 1 \qquad \|(e_7, e_8)\|_E \leq 1$$

We prove that $\mathcal{O}_2$ is stable w.r.t. $\mathbb{X}^\varepsilon$. In fact, for each perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$, we see that $M_{\mathcal{O}_2}(\widetilde{\mathbb{X}})$ is a Vandermonde matrix whose determinant is equal to $(e_4 - e_2 + 3)(e_6 - e_2 + 6)(e_8 - e_2 + 9.1)(e_6 - e_4 + 3)(e_8 - e_4 + 6.1)(e_8 - e_6 + 3.1)$. Since each $|e_i| \leq 0.15$, it follows that, for any admissible perturbation $\widetilde{\mathbb{X}}$, the matrix $M_{\mathcal{O}_2}(\widetilde{\mathbb{X}})$ is invertible, and so it is always possible to compute an $\mathcal{O}_2$-border basis of the ideal $\mathcal{I}(\widetilde{\mathbb{X}})$. In fact $\mathcal{O}_2$ is stable w.r.t. $\mathbb{X}^{(\delta_1, \delta_2)}$, where δ_1 is unlimited and $\delta_2 < 1.5$.

5.1.3 Stable border bases

Intuitively, a basis $\mathcal{B}$ of the vanishing ideal $\mathcal{I}(\mathbb{X})$ is considered to be structurally stable if, for each admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$, it is possible to produce

a basis $\widetilde{\mathcal{B}}$ of $\mathcal{I}(\widetilde{\mathbb{X}})$ only by means of a slight and continuous variation of the coefficients of the polynomials of $\mathcal{B}$, that is if there exists a basis $\widetilde{\mathcal{B}}$ of $\mathcal{I}(\widetilde{\mathbb{X}})$ whose polynomials have the same support as the corresponding polynomials of $\mathcal{B}$. We know that the supports of the polynomials of a border basis are easily computable and completely determined by the quotient basis $\mathcal{O}$ upon which the border basis is founded (see Definition 1.3.12). We start this section with the following result which highlights the importance of stable quotient bases and make a connection with border basis theory (see also Proposition 4.20 in [MR02]).

Proposition 5.1.8. *Let $\mathbb{X}^\varepsilon$ be a set of s distinct empirical points, and let $\mathcal{O} = \{t_1, \ldots, t_s\}$ be a quotient basis for $\mathcal{I}(\mathbb{X})$ which is stable w.r.t. $\mathbb{X}^\varepsilon$. Then, for each admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$, the vanishing ideal $\mathcal{I}(\widetilde{\mathbb{X}})$ has an $\mathcal{O}$-border basis $\widetilde{\mathcal{B}}$. Furthermore, if $\partial\mathcal{O} = \{b_1, \ldots, b_\nu\}$ is the border of $\mathcal{O}$ then $\widetilde{\mathcal{B}}$ consists of ν polynomials of the form*

$$g_j = b_j - \sum_{i=1}^{s} \alpha_{ij} t_i \qquad \text{for } j = 1 \ldots \nu \tag{5.1}$$

where the coefficients $a_{ij} \in \mathbb{R}$ satisfy the linear systems

$$b_j(\widetilde{\mathbb{X}}) = \sum_{i=1}^{s} \alpha_{ij} t_i(\widetilde{\mathbb{X}})$$

Proof. Let $\widetilde{\mathbb{X}}$ be an admissible perturbation of $\mathbb{X}^\varepsilon$ and let $\mathrm{eval}_{\widetilde{\mathbb{X}}} : P \to \mathbb{R}^s$ be the $\mathbb{R}$-linear evaluation map associated to the set $\widetilde{\mathbb{X}}$. It is easy to prove that $\mathcal{I}(\widetilde{\mathbb{X}}) = \ker(\mathrm{eval}_{\widetilde{\mathbb{X}}})$ and, consequently, that the quotient ring $P/\mathcal{I}(\widetilde{\mathbb{X}})$ is isomorphic to $\mathbb{R}^s$ as a vector space. Since $\mathcal{O}$ is stable w.r.t. the empirical set $\mathbb{X}^\varepsilon$, it follows that $\{t_1(\widetilde{\mathbb{X}}), \ldots, t_s(\widetilde{\mathbb{X}})\}$ are linearly independent vectors. Moreover $\#\widetilde{\mathbb{X}} = \#\mathcal{O}$, so the residue classes of the elements of $\mathcal{O}$ form a vector space basis of $P/\mathcal{I}(\widetilde{\mathbb{X}})$.

For each power product b_j lying in the border $\partial\mathcal{O}$ let $v_j = b_j(\widetilde{\mathbb{X}})$ be the associated evaluation vector. Now v_j can be expressed as

$$v_j = \sum_{i=1}^{s} \alpha_{ij} t_i(\widetilde{\mathbb{X}}) \qquad \text{for some } \alpha_{ij} \in \mathbb{R}$$

We define the polynomial $g_j = b_j - \sum_{i=1}^{s} \alpha_{ij} t_i$; by construction $\mathrm{eval}_{\widetilde{\mathbb{X}}}(g_j) = 0$, and so $\widetilde{\mathcal{B}} = \{g_1, \ldots, g_\nu\}$ is contained in $\mathcal{I}(\widetilde{\mathbb{X}})$; it follows that $\widetilde{\mathcal{B}}$ is the $\mathcal{O}$-border basis of the ideal $\mathcal{I}(\widetilde{\mathbb{X}})$. □

We observe that the coefficients α_{ij} of each polynomial $g_j \in \widetilde{\mathcal{B}}$ are just the components of the solution α_j of the linear system $M_{\mathcal{O}}(\widetilde{\mathbb{X}})\alpha_j = b_j(\widetilde{\mathbb{X}})$. It follows that α_{ij} are continuous functions of the points of the set $\widetilde{\mathbb{X}}$ and so, since the order ideal $\mathcal{O}$ is stable w.r.t. $\mathbb{X}^\varepsilon$, they undergo only continuous variations as $\widetilde{\mathbb{X}}$ changes. Now, the definition of stable border basis follows naturally.

Definition 5.1.9. Let $\mathbb{X}^\varepsilon$ be a finite set of distinct empirical points, let $\mathcal{O}$ be a quotient basis for the vanishing ideal $\mathcal{I}(\mathbb{X})$. If $\mathcal{O}$ is stable w.r.t. $\mathbb{X}^\varepsilon$ then the $\mathcal{O}$-border basis $\mathcal{B}$ for $\mathcal{I}(\mathbb{X})$ is said to be **stable** w.r.t. the set $\mathbb{X}^\varepsilon$.

Given a finite set $\mathbb{X}^\varepsilon$ of empirical points, a border basis for $\mathcal{I}(\mathbb{X})$ stable w.r.t. $\mathbb{X}^\varepsilon$ does not always exist (see Example 5.1.5). However, Proposition 5.1.8 and the subsequent observation on the continuity of the coefficients α_{ij} prove that, if $\mathcal{O}$ is a quotient basis for $\mathcal{I}(\mathbb{X})$ stable w.r.t. $\mathbb{X}^\varepsilon$, then the stable $\mathcal{O}$-border basis $\mathcal{B}$ of the ideal $\mathcal{I}(\mathbb{X})$ exists and is indeed structurally stable: namely, there exists a border basis $\widetilde{\mathcal{B}}$ for the perturbed ideal $\mathcal{I}(\widetilde{\mathbb{X}})$, sharing the same structure as $\mathcal{B}$, and whose coefficients differ only slightly, provided that $\widetilde{\mathbb{X}}$ is an admissible perturbation of $\mathbb{X}^\varepsilon$.

5.2 A method for computing stable structures

In this section we address the problem of finding an order ideal $\mathcal{O}$ stable w.r.t. a given finite set of s distinct empirical points, $\mathbb{X}^\varepsilon$. If $\mathcal{O}$ contains s power products, that is if $\mathcal{O}$ is a quotient basis of $\mathcal{I}(\mathbb{X})$, the corresponding stable border basis is also computed. The numerical examples show that $\mathcal{O}$ can have cardinality less than s when the tolerance on the points is, in some sense, too large; this phenomenon is illustrated in Examples 5.1.5 and 5.3.9.

Here we present the theoretical approach to the problem; Section 5.2.1 is dedicated to some technical results necessary for a practical implementation of the theoretical procedure, which is described in detail in Section 5.2.2.

Based on Algorithm 2.4.1, which is designed to determine a quotient basis for an ideal of exact points, our method starts from $\mathbb{X}^\varepsilon$ and computes a stable order ideal $\mathcal{O}$ by performing tests on the numerical linear dependence of a set of empirical evaluation vectors (see Definition 3.2.6). The strategy for computing a stable order ideal $\mathcal{O}$ is the following. The order ideal $\mathcal{O}$ is built stepwise: initially $\mathcal{O}$ comprises just the power product 1; then, at each iteration, a new power product t is considered. If the empirical evaluation vector $t(\mathbb{X}^\varepsilon)$ is numerically linearly independent of the family of vectors $\mathcal{O}(\mathbb{X}^\varepsilon) := \{t_i(\mathbb{X}^\varepsilon) \mid t_i \in \mathcal{O}\}$, then t is added to $\mathcal{O}$; otherwise t is added to the set of generators of a monomial ideal J. We sum up this procedure in the Table 5.1.

We observe that the procedure described in Table 5.1 indeed defines an algorithm for computing an order ideal $\mathcal{O}$ stable w.r.t. $\mathbb{X}^\varepsilon$. The procedure ends after a finite number of iterations: in fact, at each iteration, the algorithm performs either step 3 or step 4. We observe that step 3 can be executed at most $s-1$ times since in $\mathbb{R}^s$ any set of $s+1$ vectors is (numerically) linearly dependent. Moreover, step 4 is the only place where a term $t \notin J$ is added to the set of generators of J, but, since P is Noetherian, this can happen only finitely many times. We claim that the set $\mathcal{O}$ is an order ideal stable w.r.t. $\mathbb{X}^\varepsilon$: by construction $\mathcal{O}$ is factor closed, and, from step 3, $\mathcal{O}$ is stable w.r.t. $\mathbb{X}^\varepsilon$ as it is associated to a set of numerically linearly independent empirical vectors.

Another observation concerns the choice (in step 2) of the power product t

1. Input a finite non-empty set $\mathbb{X}^\varepsilon$. Let $\mathcal{O} = \{1\}$ and $J = (0)$.
2. Choose a term t in the set of corners of $\mathcal{O}$ not belonging to J, and compute the empirical evaluation vector $t(\mathbb{X}^\varepsilon)$. If no such term t exists, return $\mathcal{O}$ and stop.
3. If $t(\mathbb{X}^\varepsilon)$ is numerically linearly independent of the set $\mathcal{O}(\mathbb{X}^\varepsilon)$ then add t to $\mathcal{O}$. Continue with step 2.
4. Otherwise, add t to the set of generators of J. Continue with step 2.

Table 5.1: Theoretical approach for computing a stable order ideal

to consider at each iteration: any strategy that chooses a term t in the set of corners can be applied. A possible technique is the one used in the Buchberger-Möller algorithm (see Algorithm 2.2.2), where the chosen power product t is the smallest candidate according to a fixed term ordering σ. Note that σ is only used as a computational tool for choosing t; in fact the final computed set $\mathcal{O}$ is not, in general, the same as that which would be obtained processing the set $\mathbb{X}$ using Algorithm 2.4.1 with the same term ordering. Consider the following example (and see also Examples 5.3.1 and 5.3.3).

Example 5.2.1. Let $\mathbb{X}^\varepsilon$ be the set of empirical points having

$$\mathbb{X} = \{(0,0),\ (1,1),\ (2,2.1)\} \subseteq \mathbb{R}^2$$

as the set of specified values and $\varepsilon = (0.15, 0.15)$ as the tolerance; let σ be the degree lexicographic term ordering on $\mathbb{T}^2$ with $x > y$.

We apply Algorithm 2.4.1 to $\mathbb{X}$ and the above procedure to $\mathbb{X}^\varepsilon$ by using in steps Q2 and step 2 the criterion for choosing the power product t based on σ. Algorithm 2.4.1 returns the order ideal $\mathcal{O}_1 = \{1, y, x\}$, which is indeed equal to $\mathcal{O}_\sigma(\mathcal{I}(\mathbb{X}))$, while the above procedure computes $\mathcal{O}_2 = \{1, y, y^2\} \neq \mathcal{O}_1$.

We conclude this section with a brief observation about the computation of stable quotient bases and stable border bases: if the output $\mathcal{O}$ of the above algorithm contains exactly $\#\mathbb{X}$ elements, then $\mathcal{O}$ is necessarily a stable quotient basis for $\mathcal{I}(\mathbb{X})$, so the corresponding $\mathcal{O}$-border basis $\mathcal{B}$ exists and is stable w.r.t. $\mathbb{X}^\varepsilon$. To determine $\mathcal{B}$ it suffices to apply Algorithm 2.4.4, that is find the border of $\mathcal{O}$ (a simple combinatorical computation), and then for each element of the border solve a linear system (as described in the proof of Proposition 5.1.8).

5.2.1 Remarks on first order approximation

In this section we prove some results on the first order approximation of the rational functions of the field $F = K(x_1, \ldots, x_n)$, where $K = \mathbb{Q}$ or $K = \mathbb{R}$. We use multi-index notation to give the formal Taylor expansion of $f \in F$ at 0:

$$f = \sum_{|\alpha| \geq 0} \frac{D^\alpha f(0)}{\alpha!} \mathbf{x}^\alpha$$

We recall that given $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, we have $|\alpha| = \alpha_1 + \cdots + \alpha_n$ and $\alpha! = \alpha_1! \ldots \alpha_n!$ Similarly $D^\alpha = D_1^{\alpha_1} \ldots D_n^{\alpha_n}$ (where $D_i^j = \partial^j / \partial x_i^j$) and $\mathbf{x}^\alpha = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$. Each $f \in F$ can be decomposed into components of homogeneous degree in the following way:

$$f = \sum_{k \geq 0} f_k \quad \text{where } f_k = \sum_{|\alpha|=k} \frac{D^\alpha f(0)}{\alpha!} \mathbf{x}^\alpha$$

and where, by convention, $D^{(0\ldots0)} f = f$. Each polynomial f_k is called the **homogeneous component** of degree k of f.

Analogously, we can decompose a matrix $M \in \mathrm{Mat}_{r\times c}(F)$ into homogeneous parts in the following way.

Definition 5.2.2. Let $M = \big(m(i,j)\big)$ be a matrix in $\mathrm{Mat}_{r\times c}(F)$; we define M_k, the **homogeneous component** of degree k of M, to be the matrix whose (i,j) entry is the homogeneous component of degree k of $m(i,j)$.

Let $v \in \mathrm{Mat}_{r\times 1}(F)$ and $M \in \mathrm{Mat}_{r\times c}(F)$ be a full rank matrix, with $r \geq c$. We define $\alpha \in \mathrm{Mat}_{c\times 1}(F)$ and $\rho \in \mathrm{Mat}_{r\times 1}(F)$ via the following formulas:

$$\begin{array}{rcl} \alpha & = & (M^t M)^{-1} M^t v \\ \rho & = & v - M\alpha \end{array} \tag{5.2}$$

We observe that for any point $\delta \in K^n$ which lies in the domain of α, we can evaluate to obtain $x = \alpha(\delta)$ as the least squares solution to $M(\delta)\ x = v(\delta)$, and that the corresponding residual is $\rho(\delta)$ (see Section 2.3.1).

In our application, the matrix M comprises only polynomial entries, so the domain of α contains precisely those points $\delta \in K^n$ at which $\det(M(\delta)^t M(\delta)) \neq 0$, *i.e.* at which $M(\delta)$ has full rank (in $\mathrm{Mat}_{r\times c}(K)$).

The following proposition characterizes the homogeneous components of degrees 0 and 1 of α and ρ.

Proposition 5.2.3. *Let $r, c \in \mathbb{N}$ with $r \geq c$; let v be a vector in $\mathrm{Mat}_{r\times 1}(F)$ and let M be a full rank matrix in $\mathrm{Mat}_{r\times c}(F)$. Let $\alpha \in \mathrm{Mat}_{c\times 1}(F)$ and $\rho \in \mathrm{Mat}_{r\times 1}(F)$ be defined by (5.2). Then the homogeneous components of degrees 0 and 1 of α are*

$$\begin{array}{rcl} \alpha_0 & = & (M_0^t M_0)^{-1} M_0^t v_0 \\ \alpha_1 & = & (M_0^t M_0)^{-1} (M_0^t v_1 + M_1^t v_0 - M_0^t M_1 \alpha_0 - M_1^t M_0 \alpha_0) \end{array} \tag{5.3}$$

and the homogeneous components of degrees 0 and 1 of ρ are

$$\begin{array}{rcl} \rho_0 & = & v_0 - M_0 \alpha_0 \\ \rho_1 & = & v_1 - M_0 \alpha_1 - M_1 \alpha_0 \end{array} \tag{5.4}$$

Proof. First we prove a simple result about the homogeneous components of degrees 0 and 1 of the inverse of a matrix. Let A be a non singular element

of $\mathrm{Mat}_{c\times c}(F)$, and let B be its inverse. The homogeneous components B_0 and B_1 are given by

$$B_0 = A_0^{-1} \qquad B_1 = -A_0^{-1}A_1A_0^{-1} = -B_0A_1B_0 \tag{5.5}$$

We show this by decomposing A and B into sums of homogeneous components:

$$A = A_0 + A_1 + A_{2+} \quad \text{and} \quad B = B_0 + B_1 + B_{2+}$$

where $A_{2+} = \sum_{i\geq 2} A_i$ and $B_{2+} = \sum_{i\geq 2} B_i$. Now, since $AB = I_c$, the $c \times c$ identity matrix, we have

$$(A_0 + A_1 + A_{2+})(B_0 + B_1 + B_{2+}) = I_c$$

and our claim is immediate after expanding the product into a sum of homogeneous components.

Now we prove the result of the proposition. Since M is full rank, the matrix $A = M^tM$ is non singular and so we can define

$$\alpha = A^{-1}M^tv \tag{5.6}$$
$$\rho = v - M\alpha \tag{5.7}$$

Applying to (5.7) the homogeneous degree decomposition up to degree 1 we have

$$\rho_0 + \rho_1 = (v_0 - M_0\alpha_0) + (v_1 - M_0\alpha_1 - M_1\alpha_0)$$

thus (5.4) follows.

Since $A_0 = M_0^tM_0$ and $A_1 = M_0^tM_1 + M_1^tM_0$, from formula (5.5) we have the first two homogeneous components of $B = A^{-1} \equiv A_0^{-1} - A_0^{-1}A_1A_0^{-1}$. Up to degree 1, formula (5.6) becomes

$$\begin{aligned}\alpha_0 + \alpha_1 &= B_0(M_0^tv_0 + M_0^tv_1 + M_1^tv_0) + B_1M_0^tv_0 = \\ &= B_0\Big(M_0^tv_0 + M_0^tv_1 + M_1^tv_0 - A_1B_0M_0^tv_0\Big)\end{aligned}$$

and so

$$\begin{aligned}\alpha_0 &= (M_0^tM_0)^{-1}M_0^tv_0 \\ \alpha_1 &= (M_0^tM_0)^{-1}(M_0^tv_1 + M_1^tv_0 - M_0^tM_1\alpha_0 - M_1^tM_0\alpha_0)\end{aligned}$$

thus the proof is concluded. □

5.2.2 The SOI Algorithm

In this section we present the Stable Order Ideal (SOI) Algorithm, a practical implementation of the theoretical method for computing stable order ideals described in Section 5.2.

Given a finite set $\mathbb{X}^\varepsilon$ of distinct empirical points the SOI Algorithm computes an order ideal $\mathcal{O}$ stable w.r.t the empirical set $\mathbb{X}^\varepsilon$. The practical approach

used is based on a first order error analysis of the problem, as our interest is essentially focused on small perturbations $\widetilde{\mathbb{X}}$ of the empirical set $\mathbb{X}^\varepsilon$. In order to investigate the stability of $\mathcal{O}$ the SOI Algorithm uses some results on the first order approximation of rational functions (see Section 5.2.1) and exploits the parametric description of an admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$ defined in Section 3.1.1, so that the notion of numerical linear dependence of empirical evaluation vectors given in Definition 3.2.6 and used above is greatly simplified. As a consequence the check on the numerical linear dependence is carried out only by using the classical least squares method (as in Algorithms 2.3.5).

A first observation concerns the choice of the power product t to analyze at each iteration (step 2): as already pointed out any strategy that chooses t such that $\mathcal{O} \cup \{t\}$ is factor closed can be applied, also the one which exploits a fixed term ordering σ. For the sake of simplicity, the version of the SOI Algorithm presented below employs this latter strategy.

Another observation concerns the main loop of the algorithm (steps 3 and 4): note that the check on the numerical linear dependence of empirical evaluation vectors in terms of the parametric description $\widetilde{\mathbb{X}}(\delta)$, is equivalent to checking whether $\rho(\delta)$, the component of the evaluation vector $t(\widetilde{\mathbb{X}}(\delta))$ orthogonal to the column space of the matrix $M_{\mathcal{O}}(\widetilde{\mathbb{X}}(\delta))$, vanishes for some $\delta \in D_\varepsilon$. This check, greatly simplified by our restriction to first order error terms, requires a real parameter γ depending on the norm of $\rho_{2+} = \sum_{k\geq 2} \rho_k$, where each ρ_k is the homogeneous component of degree k of ρ (see Theorem 5.2.5).

Algorithm 5.2.4. (The Stable Order Ideal Algorithm)
Let $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_s^\varepsilon\}$ be a finite set of distinct empirical points, with $\mathbb{X} \subseteq \mathbb{R}^n$ and a common tolerance $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$, and let $\mathbf{e} = (e_{11}, \ldots, e_{sn})$ be the error variables whose constraints are given in (3.2). Let σ be a term ordering on $\mathbb{T}^n$ and $\gamma \geq 0$ (see Theorem 5.2.5). Consider the following sequence of instructions.

S1 Start with the lists $\mathcal{O} = [1]$, $L = [x_1, \ldots, x_n]$, and $C = [\ ]$; create the matrices $M_0 \in \mathrm{Mat}_{s\times 1}(\mathbb{R})$ initially with all the elements equal to 1, and $M_1 \in \mathrm{Mat}_{s\times 1}(\mathbb{R})$ initially with all the elements equal to 0.

S2 If $L = [\ ]$ return the set $\mathcal{O}$ and stop. Otherwise let $t = \min_\sigma(L)$ and delete it from L.

S3 Let v_0 and v_1 be the homogeneous components of degrees 0 and 1 of the evaluation vector $v = t(\widetilde{\mathbb{X}}(\mathbf{e}))$. Compute the vectors (see Proposition 5.2.3)

$$\begin{aligned} \rho_0 &= v_0 - M_0\alpha_0 \\ \rho_1 &= v_1 - M_0\alpha_1 - M_1\alpha_0 \end{aligned}$$

where

$$\begin{aligned} \alpha_0 &= (M_0^t M_0)^{-1} M_0^t v_0 \\ \alpha_1 &= (M_0^t M_0)^{-1}(M_0^t v_1 + M_1^t v_0 - M_0^t M_1 \alpha_0 - M_1^t M_0 \alpha_0). \end{aligned}$$

S4 Let $C_t \in \mathrm{Mat}_{s\times sn}(\mathbb{R})$ be such that $\rho_1 = C_t\mathbf{e}$. Let k be the maximum integer such that the matrix $\widehat{C}_t$, formed by selecting the first k rows of C_t, has minimum singular value $\widehat{\sigma}_k$ greater than $\|\varepsilon\|$. Let $\widehat{\rho}_0$ be the vector comprising the first k elements of ρ_0 and let $\widehat{C}_t^\dagger$ be the pseudoinverse of $\widehat{C}_t$. Compute $\widehat{\delta} = -\widehat{C}_t^\dagger\widehat{\rho}_0$, which is the minimal 2-norm solution of the underdetermined system $\widehat{C}_t\widehat{\delta} = -\widehat{\rho}_0$ (see [DH93]).

S5 If $\|\hat{\delta}\| > (1+\gamma)\sqrt{s}\|\varepsilon\|$ then adjoin the vector v_0 as a new column of M_0 and the vector v_1 as a new column of M_1. Append the power product t to $\mathcal{O}$, and add to L those elements of $\{x_1t, \ldots, x_nt\}$ which are not multiples of an element of L or C. Continue with step S2.

S6 Otherwise append t to the list C, and remove from L all multiples of t. Continue with step S2.

Theorem 5.2.5. *Algorithm 5.2.4 stops after finitely many steps and returns a factor closed set $\mathcal{O} \subset \mathbb{T}^n$. If γ satisfies $\sup_{\delta\in D_\varepsilon} \|\rho_{2+}(\delta)\| \leq \gamma\sqrt{s}\|\varepsilon\|^2$ then $\mathcal{O}$ is an order ideal stable w.r.t. the empirical set $\mathbb{X}^\varepsilon$. In particular, when $\#\mathcal{O} = s$ then $\mathcal{I}(\mathbb{X})$ has a corresponding stable border basis w.r.t. $\mathbb{X}^\varepsilon$.*

Proof. First we claim that ρ_0, ρ_1, α_0, α_1 computed in step S3 are the homogeneous components of degrees 0 and 1 of the residual ρ and of the solution α as defined in equations (5.2), where $M = M_{\mathcal{O}}(\widetilde{\mathbb{X}}(\mathbf{e}))$. To prove this claim it is sufficient to apply Proposition 5.2.3 and to observe that the matrices M_0 and M_1 coincide with the homogeneous components of degrees 0 and 1 of M. Clearly, this is true at the first iteration, since M has all entries equal to 1. We apply induction on the number of iterations. Assume that M_0 and M_1 are the components of degrees 0 and 1 of M and suppose that the power product t is added to $\mathcal{O}$. Since the last column of $M_{\mathcal{O}\cup\{t\}}(\widetilde{\mathbb{X}}(\mathbf{e}))$ is given by $t(\widetilde{\mathbb{X}}(\mathbf{e}))$, whose components of degrees 0 and 1 are t_0 and t_1 respectively, the new matrices $[M_0, v_0]$ and $[M_1, v_1]$ are the components of degrees 0 and 1 of $M_{\mathcal{O}\cup\{t\}}(\widetilde{\mathbb{X}}(\mathbf{e}))$. We conclude that the vectors $\rho_0 + \rho_1$ and $\alpha_0 + \alpha_1$ coincide with ρ and α, up to first order.

Now we prove the finiteness and the correctness of Algorithm 5.2.4. First we show finiteness. At each iteration the algorithm performs either step S5 or step S6. We observe that step S5 can be executed at most $s-1$ times; in fact, once M_0 becomes a square matrix, *i.e.* after $s-1$ iterations of step S5, the residual vector ρ_0 will always be zero, and consequently the minimal 2-norm solution $\hat{\mathbf{e}}$ of the linear system $C_t\mathbf{e} = -\rho_0$ is also zero. Moreover, step S5 is the only place where the set L is enlarged (with a finite number of terms), while each iteration removes from L at least one element; we conclude that the algorithm reaches the condition $L = [\,]$ after finitely many iterations.

In order to show correctness we prove, by induction on the number of iterations, that the output set $\mathcal{O}$ is an order ideal stable w.r.t. $\mathbb{X}^\varepsilon$. This is clearly true after zero iterations, *i.e.* after step S1 has been executed. By induction assume that a number of iterations has already been performed and that the order

ideal $\mathcal{O}$ is stable; let us follow the steps of the new iteration, in which a power product t is considered. If step S6 is performed the claim is true because $\mathcal{O}$ does not change. Otherwise, if step S5 is performed, the set $\mathcal{O}^* = \mathcal{O} \cup \{t\}$ is factor closed by the restriction on the choice of t. In order to prove that $\mathcal{O}^*$ is stable w.r.t. $\mathbb{X}^\varepsilon$ we simply show that $\rho(\delta)$ does not vanish for any $\delta \in D_\varepsilon$, since $\rho(\delta)$ is the component of $t(\widetilde{\mathbb{X}}(\delta))$ orthogonal to the columns of $M_{\mathcal{O}}(\widetilde{\mathbb{X}}(\delta))$, and $\rho(\delta) \neq 0$ implies that $M_{\mathcal{O}^*}(\widetilde{\mathbb{X}}(\delta))$ has full rank. Define $\widehat{\rho}(\delta)$ to be the vector comprising the first k elements of $\rho(\delta)$. Clearly $\widehat{\rho}(\delta) \neq 0$ implies $\rho(\delta) \neq 0$, so it suffices to prove that $\widehat{\rho}(\delta)$ does not vanish on D_ε. Suppose by contradiction that there exists $\widetilde{\delta} \in D_\varepsilon$ satisfying

$$\widehat{\rho}(\widetilde{\delta}) = \widehat{\rho}_0 + \widehat{C}_t\widetilde{\delta} + \widehat{\rho}_{2+}(\widetilde{\delta}) = 0 \tag{5.8}$$

Let $\xi = \widehat{C}_t^\dagger \widehat{\rho}_{2+}(\widetilde{\delta})$ be the minimal 2-norm solution of the linear system

$$\widehat{C}_t\xi = \widehat{\rho}_{2+}(\widetilde{\delta}) \tag{5.9}$$

Substituting (5.9) into (5.8), we obtain $\widehat{C}_t(\widetilde{\delta} + \xi) = -\widehat{\rho}_0$. Since $\widetilde{\delta} \in D_\varepsilon$ we have $\|\widetilde{\delta}\| \leq \sqrt{s}\|\varepsilon\|$ and since $\widehat{\delta}$ is the minimal 2-norm solution of $\widehat{C}_t\widehat{\delta} = -\widehat{\rho}_0$ we have $\|\widehat{\delta}\| \leq \|\widetilde{\delta} + \xi\|$. Thus we obtain

$$\begin{aligned} \|\widehat{\delta}\| \leq \|\widetilde{\delta} + \xi\| &\leq \sqrt{s}\|\varepsilon\| + \|\widehat{C}_t^\dagger\|\|\widehat{\rho}_{2+}(\widetilde{\delta})\| = \\ &= \sqrt{s}\|\varepsilon\| + \frac{\|\widehat{\rho}_{2+}(\widetilde{\delta})\|}{\widehat{\sigma}_k} \leq \\ &\leq (1+\gamma)\sqrt{s}\|\varepsilon\| \end{aligned}$$

This contradicts the condition at the start of step S5, and so we conclude that $\rho(\delta)$ does not vanish for any $\delta \in D_\varepsilon$. The final comment is immediate by Proposition 5.1.8. □

In order to implement the SOI Algorithm a value of γ has to be chosen even if an estimate of $\sup_{\delta \in D_\varepsilon} \|\rho_{2+}(\delta)\|$ is unknown. Since we consider small perturbations $\widetilde{\mathbb{X}}$ of the empirical set $\mathbb{X}^\varepsilon$, in most cases $\rho_0 + \rho_1(\delta)$ is a good linear approximation of $\rho(\delta)$ for every $\delta \in D_\varepsilon$. For this reason $\sup_{\delta \in D_\varepsilon} \|\rho_{2+}(\delta)\|$ is small and a value of $\gamma \ll 1$ can be chosen to obtain a set $\mathcal{O}$ stable w.r.t. $\mathbb{X}^\varepsilon$. On the other hand, if ρ is not well approximated by its homogeneous components of degrees 0 and 1 then our strategy loses its meaning, since it is based on the first order analysis.

5.3 Numerical examples

In this section we present some numerical examples to show the effectiveness of the SOI Algorithm. Our algorithm is implemented using the C++ language and the CoCoALib (see [CoC]) and all computations have been performed on an Intel Pentium M735 processor (at 1.7 GHz) running GNU/Linux. In all

the examples, the SOI Algorithm was run using a fixed precision of 1024 bits for the twin-float arithmetic implemented in CoCoALib (called RingTwinFloat, see [Abb07]), with the parameter $\gamma = 0.1$ and using the degree lexicographic term ordering σ; in addition, the coefficients of the polynomials are displayed as truncated decimals.

The following example shows that even though we use a term ordering σ in the SOI Algorithm, the resulting $\mathcal{O}$-border basis need not contain the τ-Gröbner basis of $\mathcal{I}(\mathbb{X})$ for any term ordering τ.

Example 5.3.1. The quotient basis $\mathcal{O}$ is not of Gröbner type
Let $\mathbb{X}^\varepsilon$ be a set of distinct empirical points having

$$\mathbb{X} = \{(1.1, 1.1),\ (0.9, -1.1),\ (-0.9, 0.9),\ (-1.1, -0.9)\}$$

as the set of specified values and $\varepsilon = (0.1, 0.1)$ as the tolerance. Applying the SOI algorithm to $\mathbb{X}^\varepsilon$, we obtain the quotient basis $\mathcal{O} = \{1, x, y, xy\}$ which is stable w.r.t. $\mathbb{X}^\varepsilon$.

Let τ be any term ordering on $\mathbb{T}^n$ and $\mathcal{O}_\tau(\mathcal{I}(\mathbb{X})) = \mathbb{T}^n \setminus \mathrm{LT}_\tau\{\mathcal{I}(\mathbb{X})\}$ be the quotient basis associated to τ. We observe that $\mathcal{O} \neq \mathcal{O}_\tau(\mathcal{I}(\mathbb{X}))$: in fact, since τ is a term ordering we have either $x^2 <_\tau xy$ or $y^2 <_\tau xy$; furthermore, the evaluation vectors $x^2(\mathbb{X})$ and $y^2(\mathbb{X})$ are each linearly independent of $\{1(\mathbb{X}),\ x(\mathbb{X}),\ y(\mathbb{X})\}$ so that one of x^2 or y^2 must belong to $\mathcal{O}_\tau(\mathcal{I}(\mathbb{X}))$. We conclude that the $\mathcal{O}$-border basis of $\mathcal{I}(\mathbb{X})$ does not contain any Gröbner basis of $\mathcal{I}(\mathbb{X})$.

Nevertheless, if there is no perturbation on the coordinates of the original points, that is if $\varepsilon = 0$, the SOI Algorithm returns the set $\mathcal{O} = \mathrm{LT}_\sigma(\mathcal{I}(\mathbb{X}))$, which is a quotient basis of $\mathcal{I}(\mathbb{X})$; the $\mathcal{O}$-border basis of $\mathcal{I}(\mathbb{X})$ thus contains the σ-Gröbner basis.

Example 5.3.2. Let $\mathbb{X}^\varepsilon$ be the set of 11 distinct empirical points having

$$\begin{aligned}\mathbb{X} \;=\; &\{(-15.625, -2.5),\ (-8, -2),\ (-3.375, -1.5),\ (-1, -1),\ (-0.125, -0.5),\\ &(0, 0),\ (0.125, 0.5),\ (1, 1),\ (3.375, 1.5),\ (8, 2),\ (15.625, 2.5)\} \subseteq \mathbb{R}^2\end{aligned}$$

as the set of specified values and $\varepsilon = (0, 0)$ as the tolerance. Applying the SOI Algorithm to $\mathbb{X}^\varepsilon$, we obtain the quotient basis

$$\mathcal{O} = \{1, y, x, y^2, xy, x^2, xy^2, x^2y, x^3, x^2y^2, x^3y\}$$

which is stable w.r.t. $\mathbb{X}^\varepsilon$. The stable $\mathcal{O}$-border basis $\mathcal{B}$ of $\mathcal{I}(\mathbb{X})$ is given by:

$$\mathcal{B} \approx \begin{cases} y^3 - x \\ xy^3 - x^2 \\ x^2y^3 - x^3 \\ x^4 - 13.75x^3y + 63.938x^2y^2 - 119.45x^2 + 82.328xy - 14.063y^2 \\ x^4y - 125.13x^3 + 759.69x^2y - 1560.15xy^2 + 1117.95x - 193.36y \\ x^3y^2 - 13.75x^3 + 63.94x^2y - 119.45xy^2 + 82.328x - 14.063y \end{cases}$$

The σ-Gröbner basis is:

$$\mathcal{G} = \begin{cases} y^3 - x \\ x^4 - 55/4x^3y + 1023/16x^2y^2 - 7645/64x^2 + 5269/64xy - 225/16y^2 \\ x^3y^2 - 55/4x^3 + 1023/16x^2y - 7645/64xy^2 + 5269/64x - 225/16y \end{cases}$$

It is simple to observe that the polynomials of $\mathcal{B}$ whose leading form belongs to the corners of $\mathcal{O}$, namely the first, the third and the last elements, coincide with the σ-Gröbner basis of $\mathcal{I}(\mathbb{X})$.

The following three examples show how the SOI Algorithm detects the simplest geometrical configuration almost satisfied by the empirical set $\mathbb{X}^\varepsilon$.

Example 5.3.3. Four almost aligned points
We consider the empirical set $\mathbb{X}^\varepsilon$ given in Example 5.0.6; we recall here the points in $\mathbb{X}$

$$\mathbb{X} = \{(-1,-5),\ (0,-2),\ (1,1),\ (2,4.1)\} \subseteq \mathbb{R}^2$$

and the tolerance $\varepsilon = (0.15, 0.15)$. Applying the SOI Algorithm to $\mathbb{X}^\varepsilon$ we obtain the quotient basis $\mathcal{O} = \{1, y, y^2, y^3\}$ which is stable w.r.t. $\mathbb{X}^\varepsilon$, as we proved in Example 5.0.6. As $\mathcal{O}$ is a quotient basis for $\mathcal{I}(\mathbb{X})$ we can compute the border basis $\mathcal{B}$ founded on it:

$$\mathcal{B} \approx \begin{cases} x + 0.0002y^3 + 0.0012y^2 - 0.3328y - 0.6686 \\ xy + 0.0008y^3 - 0.3286y^2 - 0.6643y - 0.0079 \\ xy^2 - 0.3301y^3 - 0.6471y^2 + 0.0098y - 0.0326 \\ xy^3 - 0.0199y^3 - 7.1199y^2 - 7.3933y + 13.533 \\ y^4 + 1.9y^3 - 21.6y^2 - 22.3y + 41 \end{cases}$$

Note that the lowest degree polynomial of $\mathcal{B}$,

$$f_1 = x + 0.0002y^3 + 0.0012y^2 - 0.3328y - 0.6686$$

highlights the fact that $\mathbb{X}$ contains "almost aligned" points. In fact, if we neglect the terms with smallest coefficients, f simplifies to $x - 0.3328y - 0.6686$. Since the coefficients of a polynomial are continuous functions of its zeros and the quotient basis $\mathcal{O}$ is stable w.r.t. $\mathbb{X}^\varepsilon$, we can conclude that there exists a small perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}$ containing aligned points and for which the associated evaluation matrix $M_{\mathcal{O}}(\widetilde{\mathbb{X}})$ is invertible. A simple example of such a set is given by $\widetilde{\mathbb{X}} = \{(-1,-5),\ (0,-2),\ (1,1),\ (2,4)\}$. Note that the algorithm also yields the almost vanishing polynomial $f_2 = x - 0.330y - 0.656$ which also highlights tha fact that $\mathbb{X}$ contains "almost aligned" points. A further interesting example is obtained by taking the difference of f_1 and f_2. The resulting polynomial $h = 0.0002y^3 + 0.0012y^2 - 0.0027y - 0.0118$ has small values at the points of $\mathbb{X}$. But this is not a contradiction to the "almost linear independence" of the terms contained in $\mathcal{O}$, since there is no admissible perturbation of $\mathbb{X}$ for which h vanishes.

A completely different result is obtained by applying to the set $\mathbb{X}$ the Buchberger-Möller algorithm w.r.t. the same term ordering σ. The σ-Gröbner basis $\mathcal{G}$ of $\mathcal{I}(\mathbb{X})$ is:

$$\mathcal{G} = \begin{cases} x^2 - 1/9y^2 - 121/30x + 9/10y + 101/45 \\ xy - 1/3y^2 - 41/10x + 7/10y + 41/15 \\ y^3 + 6y^2 + 516243/100x - 171781/100y - 172581/50 \end{cases}$$

and the associated quotient basis is $\mathcal{O}_\sigma(\mathcal{I}(\mathbb{X})) = \mathbb{T}^2 \setminus \mathrm{LT}_\sigma\{\mathcal{I}(\mathbb{X})\} = \{1, y, x, y^2\}$. We observe that $\mathcal{O}_\sigma(\mathcal{I}(\mathbb{X}))$ is not stable (see Example 5.0.6) because the evaluation matrix $M_{\mathcal{O}_\sigma}(\widetilde{\mathbb{X}})$ is singular for some admissible perturbations of $\mathbb{X}$, for instance $\widetilde{\mathbb{X}}$ just above. In particular, the information that the points of $\mathbb{X}$ are "almost aligned" is not at all evident from $\mathcal{G}$.

Example 5.3.4. Empirical points close to an ellipse
Let $\mathbb{X} \subseteq \mathbb{R}^2$ be a set of points created by perturbing by less than 0.1 the coordinates of 10 points lying on the ellipse $4x^2 + y^2 - 100 = 0$,

$$\begin{aligned} \mathbb{X} \;=\; & \{(-5.07, 0.02), (4.98, 0), (3.05, 8.07), (3.01, -8.02), (-3.02, 7.99), \\ & (-2.98, -8), (4.01, 5.94), (3.98, -6.06), (-3.92, 6.03), (-4.01, -6)\} \end{aligned}$$

Let $\mathbb{X}^\varepsilon$ be the set of empirical points whose set of specified values is $\mathbb{X}$ and whose common tolerance is $\varepsilon = (0.1, 0.1)$. Applying the SOI Algorithm to $\mathbb{X}^\varepsilon$ we obtain, after 11 iterations, the stable quotient basis

$$\mathcal{O} = \{1, y, x, y^2, xy, y^3, xy^2, y^4, xy^3, xy^4\}$$

We use linear algebra to compute the corresponding stable border basis $\mathcal{B}$ of $\mathcal{I}(\mathbb{X})$. We can identify the "almost elliptic" configuration of the points of $\mathbb{X}$ by looking at f the lowest degree polynomial contained in $\mathcal{B}$:

$$\begin{aligned} f \;\approx\; & x^2 + 0.273y^2 - 25.250 + 10^{-2}(0.004xy^4 + 0.020xy^3 - 0.034y^4 \\ & -0.489xy^2 - 0.177y^3 - 1.371xy + 9.035x + 9.810y) \end{aligned}$$

We observe that f highlights the fact that $\mathbb{X}$ contains points close to an ellipse. In fact, if we neglect the terms with smallest coefficients, f simplifies to $x^2 + 0.273y^2 - 25.250$. Since the coefficients of a polynomial are continuous functions of its zeros and the quotient basis $\mathcal{O}$ is stable w.r.t. $\mathbb{X}^\varepsilon$, we can conclude that there exists a small perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}$ containing points lying on an ellipse and such that the associated evaluation matrix $M_{\mathcal{O}}(\widetilde{\mathbb{X}})$ is invertible. A simple example of such a set is given by

$$\begin{aligned} \widetilde{\mathbb{X}} \;=\; & \{(-5, 0), (5, 0), (3, 8), (3, -8), (-3, 8), \\ & (-3, -8), (4, 6), (4, -6), (-4, 6), (-4, -6)\} \end{aligned}$$

Example 5.3.5. In this example we apply the SOI Algorithm to two sets of points which are the output of the Agglomerative and the Divisive Algorithms. We consider the set $\mathbb{X}_2 \subseteq \mathbb{R}^2$ given in Example 4.3.2 as the original set of points; we recall that $\mathbb{X}_2$ is made up of 5032 points lying close to the circle of radius 200 and centered at the origin. We call $\mathbb{Y}_1$ and $\mathbb{Y}_2$ the set of valid representatives obtained by applying the AA and the DA to $\mathbb{X}_2^\varepsilon$, with $\varepsilon = (64, 64)$. We have

$$\begin{aligned}
\mathbb{Y}_1 &= \{(-198.46, -17.8768), (-193.366, 46.0792), (-181.034, -81.8125),\\
&\quad (-145.47, 132.429), (-129.486, -149.977), (-62.6498, 188.546),\\
&\quad (-66.367, -187.972), (21.3435, -195.551), (35.3222, 193.707),\\
&\quad (129.221, -148.831), (124.263, 155.095), (173.186, 97.5654),\\
&\quad (190.032, -52.137), (196.248, 34.453)\}\\
\mathbb{Y}_2 &= \{(109.38, -163.448), (-123.76, -152.857), (-109.347, 163.608),\\
&\quad (121.045, 155.095), (-9.06163, -196.388), (7.07879, 196.572),\\
&\quad (188.826, 54.8762), (-189.924, -50.954), (184.636, -67.6265),\\
&\quad (-183.875, 70.1091)\}
\end{aligned}$$

Note that $\mathbb{Y}_1$ consist of 14 and 10 points respectively (see also Table 4.1). We choose $\varepsilon = (5, 5)$. Applying the SOI Algorithm to $\mathbb{Y}_1^\varepsilon$ and $\mathbb{Y}_2^\varepsilon$ we obtain the stable quotient bases $\mathcal{O}_1$ and $\mathcal{O}_2$

$$\begin{aligned}
\mathcal{O}_1 &= \{1, y, x, y^2, xy, y^3, xy^2, y^4, xy^3, y^5, xy^4, y^6, xy^5, y^7\}\\
\mathcal{O}_2 &= \{1, y, x, y^2, xy, y^3, xy^2, y^4, xy^3, y^5\}
\end{aligned}$$

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be the stable border bases of $\mathcal{I}(\mathbb{X}_1)$ and $\mathcal{I}(\mathbb{X}_2)$ built upon $\mathcal{O}_1$ and $\mathcal{O}_2$; let f_1 and f_2 be the lowest degree polynomial of $\mathcal{B}_1$ and $\mathcal{B}_2$:

$$\begin{aligned}
f_1 &= x^2 - 0.036xy + 1.029y^2 + 1.632x - 10.030y - 39466.67 +\\
&\quad +10^{-3}(2 \cdot 10^{-9}y^7 - 10^{-8}xy^5 + 2 \cdot 10^{-8}y^6 + 2 \cdot 10^{-5} - 10^{-4}xy^4 +\\
&\quad +10^{-4}y^5 + 3 \cdot 10^{-3}xy^3 - 10^{-3}y^4 - 5 \cdot 10^{-1}xy^2 + 2y^3)\\
f_2 &= x^2 + 0.001xy + 0.997y^2 + 0.080x - 0.290y - 38667.78 +\\
&\quad +10^{-5}(10^{-5}y^5 - 3 \cdot 10^{-3}xy^3 + 9 \cdot 10^{-3}y^2 - 10^{-1}xy^2 + y^3)
\end{aligned}$$

Note that f_1 and f_2 highlight the fact that $\mathbb{Y}_1$ and $\mathbb{Y}_2$ contain points lying close to a circle. In fact, if we neglect the terms with smallest coefficients, f_1 and f_2 simplify to $x^2 - 0.036xy + 1.029y^2 + 1.632x - 10.030y - 39466.67$ and $x^2 + 0.001xy + 0.997y^2 + 0.080x - 0.290y - 38667.78$.

In the next example we show the behaviour of the SOI Algorithm when applied to several sets of points with similar geometrical configuration but with different cardinality.

Example 5.3.6. Empirical points close to a circle
Let $\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \mathbb{X}_4 \subset \mathbb{R}^2$ be sets of points created by perturbing by less than 0.01 the coordinates of $8, 16, 32$ and 64 points lying on the circumference $x^2 +$

$y^2 - 1 = 0$, and let $\varepsilon = (0.01, 0.01)$ be the tolerance. We summarize in Table 5.2 the numerical tests performed by applying the SOI algorithm to the empirical set $\mathbb{X}_i^\varepsilon$, for $i = 1 \dots 4$. The first two columns of the table contain the name of the processed set and the value of its cardinality. The column labelled with "Corners" refers to the set of corners of the stable order ideal computed by the algorithm.

Input	$\#\mathbb{X}_i$	Corners
$\mathbb{X}_1$	8	$\{x^2, xy^3, y^5\}$
$\mathbb{X}_2$	16	$\{x^2, y^4\}$
$\mathbb{X}_3$	32	$\{x^2, y^5, xy^4\}$
$\mathbb{X}_4$	64	$\{x^2, y^4\}$

Table 5.2: Output of SOI computed on sets of points close to a circle

Note that the set of corners of the stable order ideals computed by the SOI Algorithm always contain the power product x^2: this means that there is a numerical linear dependence among the empirical vectors associated to the power products $\{1, y, x, y^2, xy, x^2\}$ and that some useful information on the geometrical configuration of the points could be found.

The numerical tests suggest that in most cases the SOI Algorithm computes a stable quotient basis, allowing us to determine a stable border basis of $\mathcal{I}(\mathbb{X})$. The following example shows a phenomenon of "premature termination": the stable order ideal computed by the SOI Algorithm contains less than $\#\mathbb{X}$ terms since the tolerance on the set $\mathbb{X}$ is, in some sense, too large.

Example 5.3.7. Three almost aligned points
Let $\mathbb{X}^\varepsilon$ be the set of distinct empirical points having

$$\mathbb{X} = \{(0.1, 0), (0.98, 1), (2.03, 2)\} \subset \mathbb{R}^2$$

as the set of specified values and $\varepsilon = (0.6, 0.6)$ as the tolerance.

Applying the SOI Algorithm to $\mathbb{X}^\varepsilon$ we obtain the order ideal $\mathcal{O} = \{1, y\}$ which is stable; however, this is not a quotient basis, so we cannot obtain the corresponding stable border basis of $\mathcal{I}(\mathbb{X})$. We observe that this result is not due to the inadequacy of the first order approximation approach on which the SOI Algorithm is based, but rather to the fact that $\mathcal{I}(\mathbb{X})$ has no stable quotient bases w.r.t. ε. In fact, it is simple to verify that each of the order ideals of $\mathbb{T}^2$ containing exactly 3 elements, that is:

$$\mathcal{O}_1 = \{1, x, x^2\}, \quad \mathcal{O}_2 = \{1, x, y\}, \quad \mathcal{O}_3 = \{1, y, y^2\}$$

is not a stable quotient basis of $\mathcal{I}(\mathbb{X})$.

The same situation occurs when considering the set of points of Example 5.1.5.

Example 5.3.8. Let $\mathbb{X}^\varepsilon$ be the set of empirical points given in Example 5.1.5. Applying the SOI Algorithm to $\mathbb{X}^\varepsilon$ we obtain the order ideal $\mathcal{O} = \{1\}$, which is trivially stable.

The following example investigates further on the possible causes of the premature termination of the SOI Algorithm. In particular we see that, with a fixed set of specified values, the algorithm produces different results for different values of ε.

Example 5.3.9. Empirical points close to two conics and a cubic
Let $\mathbb{X}^\varepsilon$ be the set of distinct empirical points having

$$\mathbb{X} = \{(1,6),(2,3),(2.449,2.449),(3,2),(6,1)\} \subset \mathbb{R}^2$$

as the set of specified values and $\varepsilon_1 = (0.15, 0.15)$ as the tolerance. Applying the SOI Algorithm to $\mathbb{X}^{\varepsilon_1}$, we obtain the stable order ideal $\mathcal{O}_1 = \{1, y, x, y^2\}$; however, this is not a quotient basis, so we cannot obtain the corresponding stable border basis. This is due to the fact that the points of $\mathbb{X}$ lie close to the hyperbola $\gamma_1 : xy-6 = 0$, the circle $\gamma_2 : (x-6)^2+(y-6)^2-25 = 0$ and the cubic $\gamma_3 : y^3 - 12y^2 + 6x + 47y - 73 = 0$. So, if the tolerance is too big, they "almost satisfy" all of them. In Figure 5.1 we plot with black dots the points of $\mathbb{X}$, we plot in red the hyperbola γ_1 and in blue the circle γ_2. In Figure 5.2 we zoom in around the third point of $\mathbb{X}$, $p_3 = (2.449, 2.449)$. As in the previous figure, the red line refers to γ_1 and the blue line to γ_2; with the black line we plot the circle which bounds the admissible perturbations of $p_3^{\varepsilon_1}$. Note that both γ_1 and γ_2 intersect the neighbourhood of perturbations.

Observe how the problem does not arise if we use a smaller tolerance, *e.g.* $\varepsilon_2 = (0.01, 0.01)$. Applying the SOI Algorithm to $\mathbb{X}^{\epsilon_2}$ we obtain the stable quotient basis $\mathcal{O}_2 = \{1, y, x, y^2, y^3\}$, and its corresponding border basis:

$$\mathcal{B}_2 \approx \begin{cases} xy + 0.0047y^3 - 0.0560y^2 + 0.0280x + 0.2194y - 6.336 \\ x^2 - 0.4265y^3 + 6.118y^2 - 14.559x - 32.047y + 77.711 \\ xy^2 + 0.0114y^3 - 0.1372y^2 + 0.0686x - 5.463y - 0.8231 \\ y^4 - 14.477y^3 + 76.724y^2 - 14.862x - 188.419y + 214.345 \\ xy^3 + 0.0280y^3 - 6.336y^2 + 0.1680x + 1.316y - 2.016 \end{cases}$$

In Figures 5.3 and 5.4 we zoom in around the point p_3; as above, the red line refers to the hyperbola γ_1, the blue line to the circle γ_2, and the black lines to the circles which bound the admissible perturbations of p_3 w.r.t. ε_1 and ε_2. Note that when we consider the tolerance ε_2 only the hyperbola crosses the neighbourhood of perturbations of p_3. We suspect that in many cases the reason for the premature termination of the SOI Algorithm is related to the presence of more varieties which almost satisfy the original points. We plan to investigate the existence and properties of these special varieties in our future research (see Chapter 7).

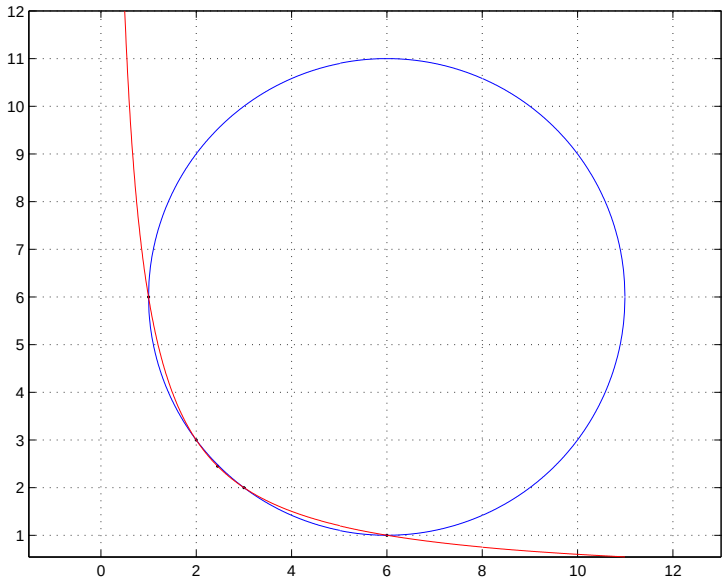


Figure 5.1: The set $\mathbb{X}$ and the curves γ_1 and γ_2

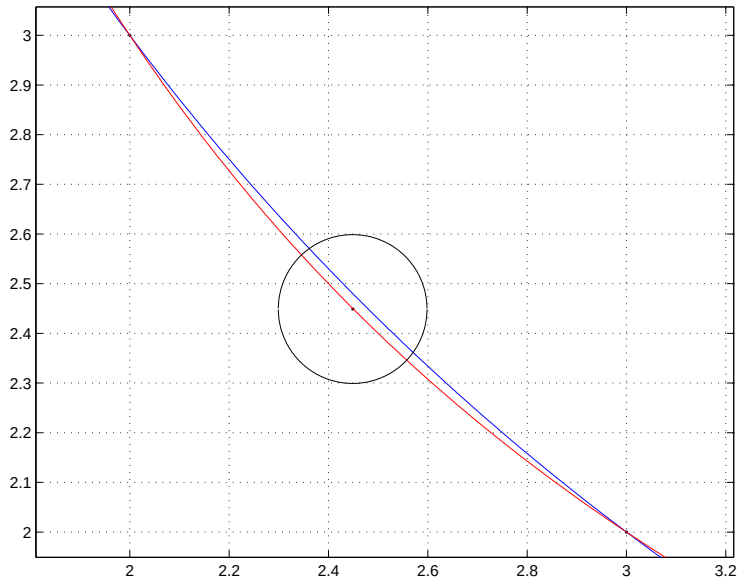


Figure 5.2: The perturbations of p_3 w.r.t. ε_1, the curves γ_1 and γ_2

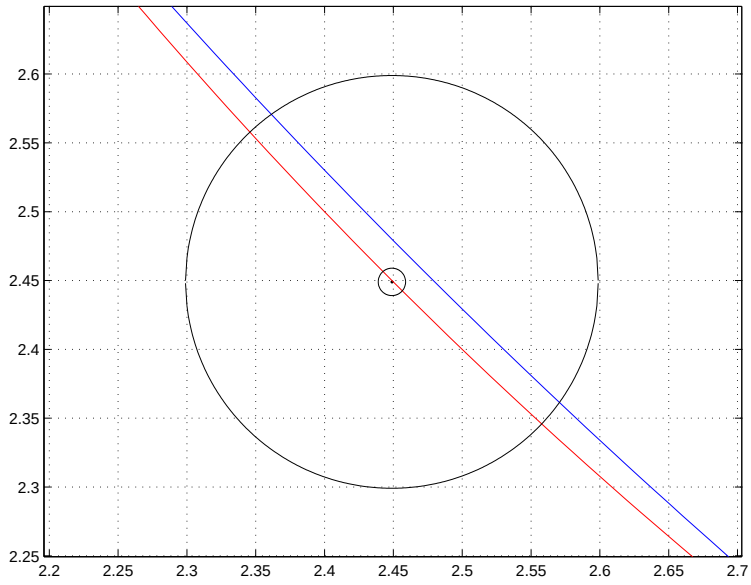


Figure 5.3: The perturbations of p_3 w.r.t. ε_1 and ε_2, the curves γ_1 and γ_2

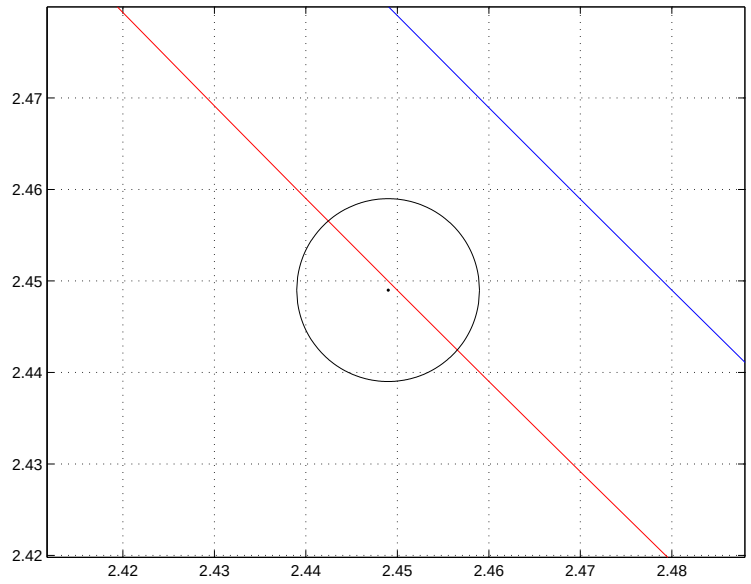


Figure 5.4: The perturbations of p_3 w.r.t. ε_2, the curves γ_1 and γ_2

Chapter 6

Application in the oil industry

In this chapter we show some results of our methods (illustrated in Chapters 4 and 5) when applied to real data coming from oil industry. The addressed problem is the modeling of oil production in the case of a multi-zone well. The traditional modeling techniques assume that equations which describe the flow of the fluids through the reservoir are available. However their limited success suggest that they do not provide a good representation of the interactions occuring during the production phase. Our aim is to find a model for the total production of a group of wells or a collection of zones which describes the production behaviour correctly over longer time scales; the idea is to take into account the interactions between the zones and so to provide a decomposition of the total production as a combination of the separate contributions of the individual wells. Rather than starting from the physical knowledge of the phenomenon, we share the idea that good models for many industrial problems can be constructed using a *bottom-up* process, in which the mathematical model is derived by interpolating the measured values on a finite set of points. For this reason we make the following assumption: we view the oil reservoir as a physical system which can be completely described in terms of measurable physical quantities, one of which being the oil production itself. We represent the set of all measurements associated to the physical quantities as a (mathematical) set of points on top of which we can build the interpolation model. In this way we find an algebraic model (which is polynomial in nature) for the oil production which depends on the relevant physical entities of the oil reservoir. During the modeling process we pay particular attention to the relations which describe the interactions of the different zones in the same production unit, and we use them to provide a decomposition of the total oil production as a combination of the separate contributions of the well.

This chapter is organized as follows. Based on [KPR08] and [HH01], in Section 6.1 we provide some background about the physical nature of an oil

reservoir and describe the problem of the control of the oil production, which is very relevant in the frame of the oil industry. In Section 6.2 we describe in detail the case of a two-zone well, we introduce new indeterminates for representing the physical entities involved in the production process, and state the crucial assumption on the existence of a causal relationship between the new indeterminates and the production. In Section 6.2.1 we describe a set $\mathcal{M}$ of numerical data coming from tests done on a two-zone well; then, in Section 6.2.2 we reduce the redundancy of $\mathcal{M}$ using the techniques introduced in Chapter 4. In Section 6.2.3 we apply the SOI Algorithm on this new set of empirical points and, starting from these results, we compute different polynomials for the production, we test their reliability and compare their prediction skills.

6.1 Oil fields, gas fields and drilling wells

In order to have a commercial deposit of oil or gas, three geological conditions must have been met. In the subsurface of the producing area there must be a source rock, that is a rock which generated the oil or gas at some time in the geological past; there must be a separate reservoir rock to hold the two fluids, and finally a trap in the reservoir rock to concentrate them into commercial quantities. If any one of the above conditions fails to hold, the process of oil/gas field formation and/or exploitation cannot occur.

The uppermost crust of the earth in oil and gas-producing areas is composed of *sedimentary rock* layers. Sedimentary rocks are composed of particles, such as sand grains, seashells, and salt, which are the source and reservoir for oil and gas. Since the densities of oil and gas are lower than the density of water (which also occurs in the subsurface sedimentary rocks), buoyancy forces them to rise through fractures in the subsurface rocks. The rising oil and gas can encounter a layer of *reservoir rock*, a sedimentary rock containing billions of tiny interconnected spaces called *pores*. In this case the oil, gas, and water will flow into the pores and move along the reservoir rock layer, which is the path of least resistance. The movement of these fluids toward the surface is called *migration* (see Figure 6.1).

Because of migration the oil, gas, and water can end up a considerable distance, both vertically and horizontally, from where they were originally formed, to where they encounter a *trap*, that is a high point in the reservoir rock, usually a natural arch, where they stop, concentrate, and separate according to their densities. The gas is the lightest and goes to the top of the trap to form the free *gas cap*; the oil goes in the middle to form the *oil reservoir*; the (salt) water, the heaviest, goes to the bottom. To complete the trap, a *caprock*, that is a seal which does not allow fluids to flow through it (usually shale and salt), must overlie the reservoir rock; without a caprock the oil and gas would leak up to the surface of the ground (see Figure 6.2).

During the early days of drilling, it was erroneously thought that there were large flowing underground rivers and subsurface pools of oil. Early drillers, however, had some success because many subsurface traps were leaky: there

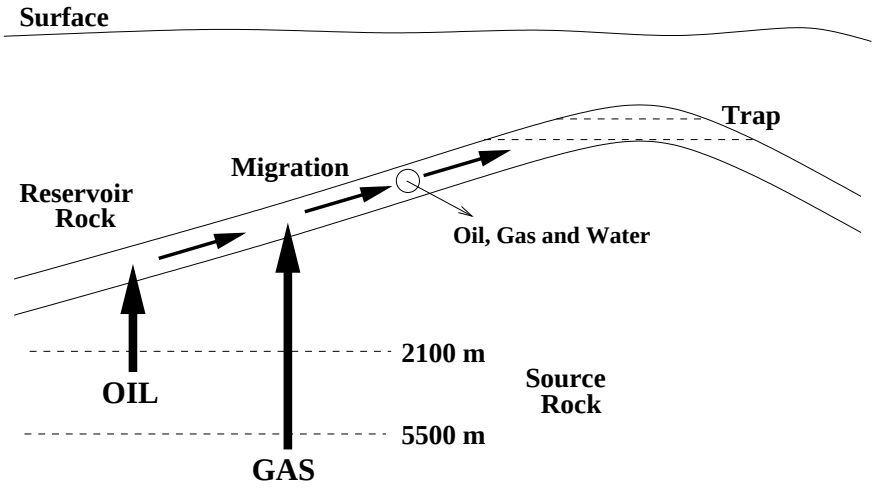


Figure 6.1: Generation and migration of oil and gas

were small fractures in the caprock, and some of the oil and gas leaked up and seeped onto the surface. Only by the early 1900s, the principles of subsurface oil and gas deposits became better known: oil companies realized that by mapping how the sedimentary rock layers crop out on the surface of the ground, the rock layers could be projected into the subsurface, and traps could be located. However, up to now, the only way to know for sure if a trap contains commercial amounts of gas and oil is to drill a well.

A well drilled to find a new oil or gas field is called a *wildcat well*. Unfortunately, most wildcats are dry holes with no commercial amounts of oil or gas. In fact, depending on the test results, the wildcat can be plugged and abandoned or recognized as a producer. In this case, a long length of large diameter steel pipe (called a *casing*), is lowered down the hole to complete the well.

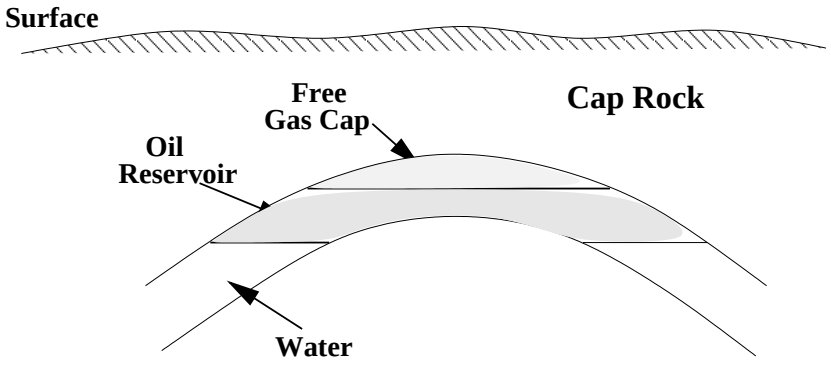
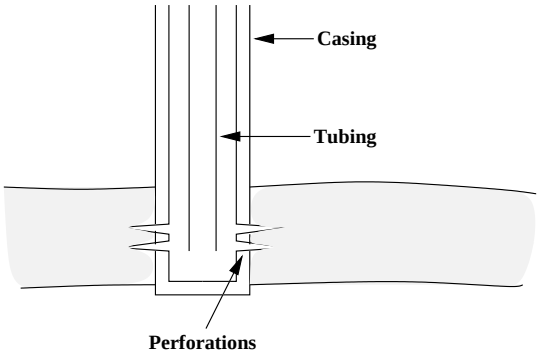


Figure 6.2: Petroleum trap

In order for the oil or gas to flow into the well, the casing is shot with explosives to form holes called *perforations*. A long length of narrow diameter steel pipe (called *tubing*) is then suspended down the center of the well. The produced fluids are brought up the tubing to the surface to prevent them from touching and corroding the casing, which is the hardest part to repair.



Perforations and tubing in a well

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 to the surface by itself. Most oil wells, however, do not have enough pressure and an *artificial lift* must be used: gas is injected into the production tubing of the well, it mixes with the oil and makes it lighter, so that the back pressure of the reservoir is reduced. On the surface, gas is prepared for delivery to a pipeline by gas-conditioning equipment that removes impurities such as water vapor and corrosive gases. For oil, a long steel tank, called a *separator*, is used to separate natural gas and salt water from it; the separated oil is then stored in steel stock tanks.

During the exploitation of a reservoir the pressure on the remaining fluids drops; the production of oil and gas from a field decreases with time, and this decrease is represented by the associated *decline curve*. The shape of the decline curve and the total volume of fluid that can be produced from a reservoir (which is called *ultimate recovery*) depend on the *reservoir drive*, the natural energy that forces the oil or gas through the subsurface reservoir and into 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: they range from 5% to 80% recovery but the average is only 30% of the oil present in the reservoir. This leaves 70% of the oil remaining in the pressure depleted reservoir which cannot be economically extracted anymore.

6.1.1 Multi-zone wells

A well may produce from different parts, called *pockets* or *zones*, of an oil reservoir. Usually each producing zone has its own packer and tubing string, so that the fluids coming from the different formations do not intermingle (*multiple completions*). However, a complete separation among the different zones is often very difficult to achieve. The well is then completed into two or more interacting zones and is called a *multi-zone well*. The total production of a multi-zone well is measured at the surface, and consists of the contributions

of the different pockets interacting with each other in the common production tubing (*commingled* production). The separate contributions can be controlled by valves, called the *down-hole valves*, to determine the rate of flow into the common tubing at the different locations of the reservoir.

As for most physical systems showing interactions, a multi-zone well can not be described by an additive model: the total oil production is not, in general, the sum of the productions of each zone. During the operations of oil extraction, the physical state of the reservoir changes: the production of oil in one zone is usually accompanied by the loss of a certain amount of gas in the reservoir, which might stimulate or inhibit the oil production in the other zones. We believe that the reason of the (usually) low ultimate recovery rate of a multi-zone well is due to the fact that the interactions among the different producing zones are unknown. A deeper understanding of the mentioned interactions would lead to a bigger production rate and therefore to a partial solution of the ultimate recovery problem, which is, up to now, the most challenging problem in oil and gas production operations.

6.2 A two-zone well and its production polynomial

We consider a multi-zone well consisting of **two** producing and interacting zones (see Figure 6.3).

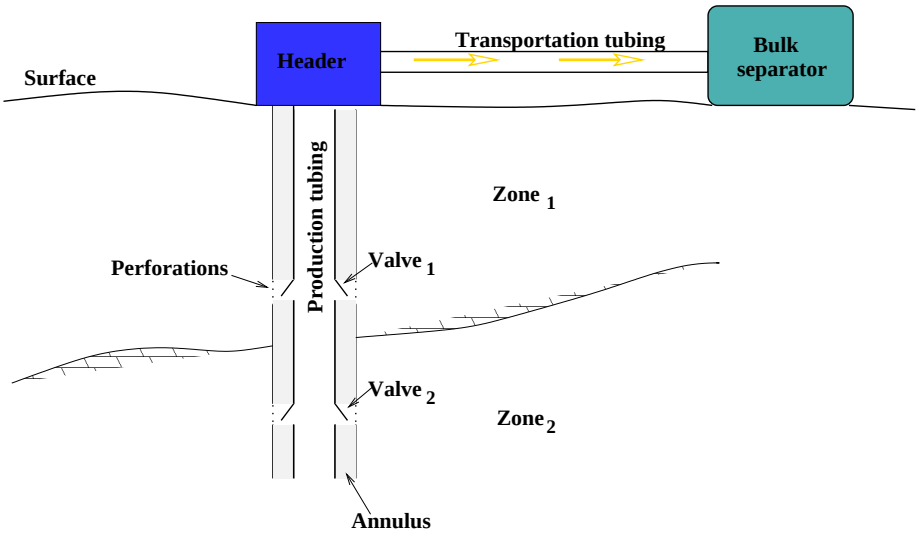


Figure 6.3: Representation of a two-zone well

Like in a single oil well, the common production flows to the bulk separator, where the different phases, namely oil, gas, and water 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 two-zone well is contained in Table 6.1.

Prod vars	**Physical Description**
ICV_1	Valve Position - zone 1
ICV_2	Valve Position - zone 2
P_{ann_1}	Down Hole Pressure Annulus - zone 1
P_{ann_2}	Down Hole Pressure Annulus - zone 2
P_{tub_1}	Down Hole Pressure Tubing - zone 1
P_{tub_2}	Down Hole Pressure Tubing - zone 2
THP	Tubing Head Pressure
FLP	Flow Line Pressure
G	Gas Production
Q	Gross (Oil, Water)

Table 6.1: Production variables in a two-zone well

The physical meaning of the production variables introduced in Table 6.1 is described in the following list (see also Figure 6.4 for a schematic representation):

ICV_i**:** opening of the valve positioned at zone i=1 or i=2, and through which the oil produced in that zone is collected; the opening is measured in percentages: 0% means that the valve is completely closed, 100% means that the valve is completely open;

P_{ann_i}**:** pressure of the oil measured at the annulus at zone i=1 or i=2;

P_{tub_i}**:** pressure of the oil measured at the tubing at zone i=1 or i=2;

THP**:** pressure of the oil measured at the head of the tubing;

FLP**:** pressure of the oil measured at the transportation tubing;

G**:** quantity of gas produced while extracting the oil;

Q**:** quantity of gross (oil and water) produced.

We simplify the set of measurements by using very basic knowledge of the physical system. We define new indeterminates (which will be used for the rest of the experiment) by associating them with physical quantities in the production problem, that means that the evaluation of each indeterminate at the set of data is indeed the measurement of the associated physical quantity. An example could be the indeterminate $x = (y_1 - y_2)y_3$, where, for instance, the original production variables y_1 and y_2 are related to pressures, and y_3 is related to the physical state of a *restriction*, that is y_3 describes the status of any obstacle in the flow like a valve, a piece of tubing, or the inflow opening from the reservoir to the production tubing. The quantity $y_1 - y_2$ is associated to the

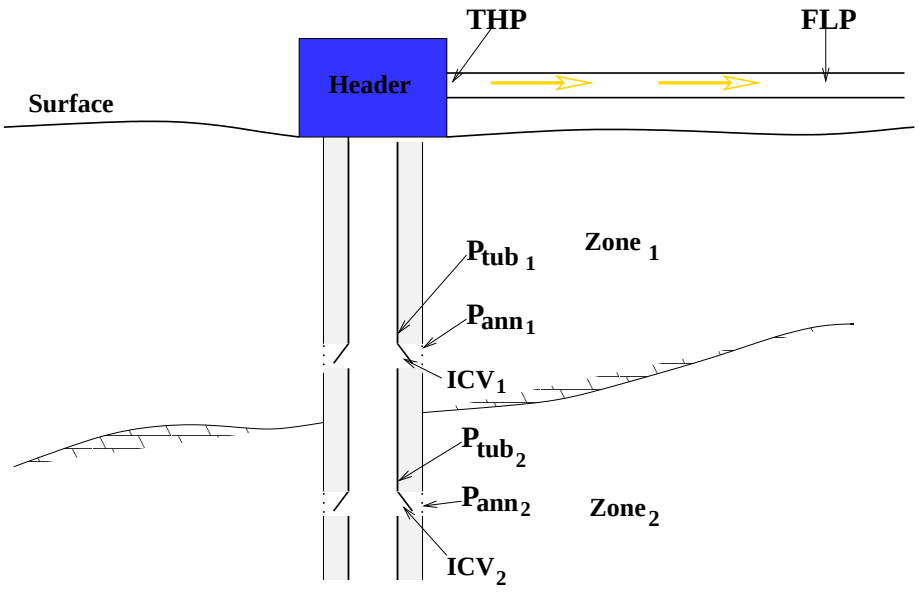


Figure 6.4: Production variables in a two-zone well

pressure drop over the restriction, and thus the new indeterminate x acquires the physical meaning of a driving force over it.

We define the 8 new indeterminates $x_1, \ldots, x_8$ and associate them with functions of the original production variables as described in Table 6.2.

New indets	Physical meaning
x_1	ICV_1
x_2	ICV_2
x_3	$\mathrm{ICV}_1 * \mathrm{ICV}_2$
x_4	$(\mathrm{P}_{\mathrm{ann}_2} - \mathrm{P}_{\mathrm{tub}_2}) * \mathrm{Ind}_2 := \Delta P_{inflow_2}$
x_5	$(\mathrm{P}_{\mathrm{ann}_1} - \mathrm{P}_{\mathrm{tub}_1}) * \mathrm{Ind}_1 := \Delta P_{inflow_1}$
x_6	G
x_7	$(\mathrm{P}_{\mathrm{tub}_2} - \mathrm{P}_{\mathrm{tub}_1}) * \mathrm{Ind}_2 := \Delta P_{tubing}$
x_8	$\mathrm{THP} - \mathrm{FLP} := \Delta P_{transport}$

Table 6.2: New indeterminates and their physical meaning

The functions Ind_1 and Ind_2 used in Table 6.2 are defined by:

$$\mathrm{Ind_i} = \begin{cases} 1 & \text{if } \mathrm{ICV_i} > 0 \iff \mathrm{ICV_i} \text{ is open} \\ 0 & \text{if } \mathrm{ICV_i} = 0 \iff \mathrm{ICV_i} \text{ is closed} \end{cases}$$

Note that they depend on the valve positions ICV_1 and ICV_2, and so give information on the status of the down-hole valves. Furthermore, we have the

following implications:

$$\begin{array}{lll} x_1 = 0 & \Longrightarrow & x_3 = 0, \quad x_5 = 0 \\ x_2 = 0 & \Longrightarrow & x_3 = 0, \quad x_4 = 0, \quad x_7 = 0 \end{array} \tag{6.1}$$

The physical interpretation of the differences of pressures used in Table 6.2 is schematically sketched in Figure 6.5.

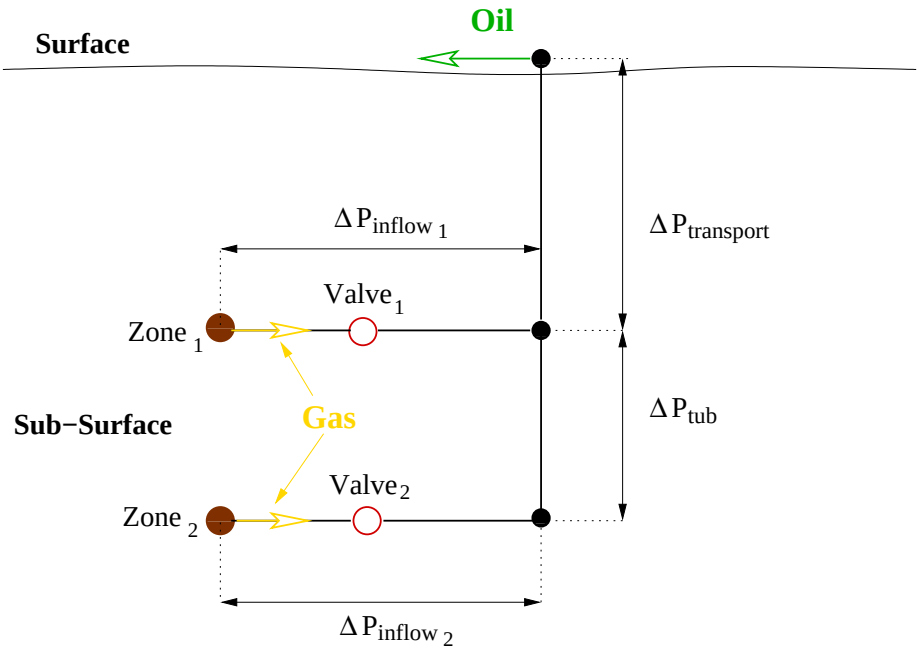


Figure 6.5: Schematic representation of a two-zone well

The indeterminates x_1 and x_2 correspond to the down-hole valve positions; x_3 is associated to the physical condition of the interaction between the two zones. We recall that the only possible inflows from the reservoir into the production tubing are at zone 1 via valve 1, or at zone 2 via valve 2, or both when the two valves are opened simultaneously and the two zones interact, which is the state represented by x_3.

The physical quantities associated to $x_4, \ldots, x_8$ may all be interpreted as driving forces of the oil production: x_4 (x_5 respectively) is the pressure which drives the oil out of the first (second) zone when the valve in that zone is open; if the valve is closed we assume its value to be zero. For this assumption some physical knowledge about the problem has been applied; we explain it briefly by focussing on x_4. If the valve in zone 2 is closed, it may very well be that the pressure difference $\mathrm{P}_{\mathrm{ann}_2} - \mathrm{P}_{\mathrm{tub}_2}$ is not zero, but it does not have the meaning of a driving force over the oil, as there is no flow through the valve. Hence we set x_4 to zero for this situation, and we do exactly the same for x_5 with respect

to the first zone. Also the gas production, represented by the indeterminate x_6, behaves as a driving force. 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 increases the lift on the oil, and consequently the amount of production. The indeterminate x_7 is associated to the pressure which drives out the oil from the first zone when the second zone (assumed to be the lower one, see Figure 6.3) is producing; if the valve at zone 2 is closed, we assume its value to be zero. Also in this case, it may happen that the pressure difference $\mathrm{P}_{\mathrm{tub}_2} - \mathrm{P}_{\mathrm{tub}_1}$ does not vanish, though this does not imply that there is transport of fluids in the lowest part of the tubing. Hence, in order to have the significance of a driving force, x_7 is set equal to zero when the valve at zone 2 is closed. Finally, the indeterminate x_8 is associated to the pressure which drives the oil at the surface and through the transportation tubing.

Being recognized as driving forces, the physical quantities associated to $x_4, \ldots, x_8$ may all be viewed as the *causing* quantities of the oil production, which itself can be viewed as their *effect*; the whole process is regulated by the variables x_1, x_2, x_3, that is by the valve positions. Basically, we make the following crucial assumption.

Assumption 6.2.1. *There exists a causal relationship between the oil production and the driving forces. Using suitable inputs, this causal relationship is polynomial in nature.*

If we denote the oil production by f, the algebraic translation of the above assumption becomes the following: $f \in \mathbb{R}[x_1, \ldots, x_8]$, where the indeterminates $x_1, \ldots, x_8$ have the physical meaning expressed in Table 6.2. Rather than being associated with an indeterminate, the oil production is expressed by a polynomial, its measurements being the evaluation of that polynomial at the set of data collected during the experiments. The problem of modeling the production in terms of the measurable physical quantities can be rephrased as the problem of finding a polynomial $f \in \mathbb{R}[x_1, \ldots, x_8]$ properly fitting the values of the oil production at the measured values of $x_1, \ldots, x_8$.

6.2.1 Description of the data

The numerical data come from tests done on a two-zone well; the information recorded refers to the situation where at most one of the down-hole valves is closed. The possible inflows from the reservoir to the production tubing are at the location of zone 1, or zone 2, or both. In all three situations data have been collected at different valve openings.

Using the simplification on the production variables illustrated in Section 6.2, the numerical data are indeed the measurements of the physical quantities associated to the variables $x_1, \ldots, x_8$, and are available at 7400 time samples. They are organized as a set of points $\mathcal{M} \subseteq \mathbb{R}^8$ which appear noisy and redundant. The production Q is the amount of gross (oil and water) produced, and it also consists of 7400 samples.

The set $\mathcal{M}$ is divided into blocks made up of measurements collected at 36 combinations of the valve openings; note that some combinations are repeated, and so $\mathcal{M}$ corresponds to 31 different physical situations of the two-zone well. Table 6.3 contains the values of the valve openings x_1 and x_2, their combination x_3, and the number of measurements contained in each block (labeled with $i = 1, \ldots, 36$).

To quantify the dispersion of the data we compute the difference between the maximum and the minimum values for each of the variables $x_1, \ldots, x_8$ in each block $i = 1, \ldots, 36$; the differences are labeled with $\Delta x_1, \ldots, \Delta x_8$ and are contained in Table 6.4. Table 6.5 contains the averages $\overline{\Delta x_1}, \ldots, \overline{\Delta x_8}$ of $\Delta x_1, \ldots, \Delta x_8$ computed across the blocks.

6.2.2 Data reduction

In order to test the reliability of our approach, the set $\mathcal{M}$ of measurements is divided into two parts: a *computational part* $\mathcal{M}_1$ of size 6000 and corresponding to the first 31 combinations (with repetition) of the valves openings, and a *testing part* $\mathcal{M}_2$ of size 1400 and corresponding to the last 5 combinations of the valves. The two parts are a partition of $\mathcal{M}$, that is $\mathcal{M}_1 \cup \mathcal{M}_2 = \mathcal{M}$ and $\mathcal{M}_1 \cap \mathcal{M}_2 = \emptyset$.

We observe that the points of the data set $\mathcal{M}_1$ are noisy and redundant. We reduce them using the techniques described in Chapter 4; in particular, we apply the Divisive Algorithm (DA, included in CoCoALib [CoC]) to the empirical set $(\mathcal{M}_1, \varepsilon)$, where

$$\varepsilon = (0.1, 0.1, 0.1, 100, 250, 40, 80, 280) \tag{6.2}$$

The choice of the tolerance vector ε comes from the analysis of the data performed in Section 6.2.1: each component ε_i, with $i \geq 4$, satisfies the relation $\varepsilon_i \approx \overline{\Delta x_i}/2$ (approximated by excess, compare the value of ε with Table 6.5). This means that, from an intuitive point of view, we consider as equivalent the set of points whose componentwise distances are less than the corresponding averages of dispersion. Furthermore, note that the value 0.1 of the first three components of ε is fictitious and it has been chosen to avoid to handle the case $\varepsilon_i = 0$. Nevertheless, it is important to observe that 0.1 plays the same role as 0: it allows to identify only the measurements having the same first three entries, exactly as if there is no error affecting their valus.

After the Divisive Algorithm we obtain the reduced set of points $PPP \subseteq \mathbb{Q}^8$, made up of 77 well separated points. From Chapter 4 we recall that the basic idea of the preprocessing technique is to gather together points which lie close together w.r.t. the tolerance ε, and replace them by a single valid representative. The multiplicity of a representative point has been defined as the number of original points it replaces. Table 6.6 contains the multiplicity of the points in PPP ordered decreasingly. The multiplicities smaller than 15 are typed in red: the corresponding points derive from observations that are numerically distant from the rest of the data, and so most likely encode failures in the measurements, for instance caused by the transit from one valve opening combination to another.

i	$\#i$-th block	x_1	x_2	x_3
1	59	10	0	0
2	47	20	0	0
3	285	30	0	0
4	119	40	0	0
5	120	50	0	0
6	240	60	0	0
7	272	70	0	0
8	118	80	0	0
9	150	90	0	0
10	113	100	0	0
11	151	0	100	0
12	85	0	40	0
13	41	0	70	0
14	631	40	100	4000
15	94	10	100	1000
16	118	20	100	2000
17	119	30	100	3000
18	92	40	100	4000
19	66	50	100	5000
20	693	10	100	1000
21	100	100	10	1000
22	95	100	20	2000
23	157	100	30	3000
24	769	100	40	4000
25	92	0	40	0
26	121	10	40	400
27	120	20	40	800
28	121	30	40	1200
29	105	30	50	1500
30	122	30	60	1800
31	740	30	70	2100
32	202	40	70	2800
33	113	50	70	3500
34	48	0	70	0
35	143	30	70	2100
36	739	40	70	2800

Table 6.3: Valve openings and number of experiments in $\mathcal{M}$

i	Δx_1	Δx_2	Δx_3	Δx_4	Δx_5	Δx_6	Δx_7	Δx_8
1	0	0	0	0	863.29	24.85	0	755.6
2	0	0	0	0	766.4	101.96	0	1031.8
3	0	0	0	0	1476.04	55.16	0	1289.92
4	0	0	0	0	902.17	34.08	0	604.06
5	0	0	0	0	516.21	25.01	0	571.44
6	0	0	0	0	800.4	67.42	0	727.36
7	0	0	0	0	175.17	80.76	0	279.88
8	0	0	0	0	36.4	66.65	0	152.81
9	0	0	0	0	31.28	87.2	0	134.58
10	0	0	0	0	28.53	91.65	0	110.73
11	0	0	0	7.5	0	25.5	24.56	292.64
12	0	0	0	109.13	0	98.2	54.19	375.1
13	0	0	0	88.61	0	41.8	163.84	410.3
14	0	0	0	8.27	178.32	26.33	76.77	514.79
15	0	0	0	15.44	336.75	23.29	172.27	247.44
16	0	0	0	25.88	524.96	25.92	251.06	349.14
17	0	0	0	15.88	436.7	65.91	211.99	271.93
18	0	0	0	9.43	170.16	42.66	94.41	171.81
19	0	0	0	3.99	20.24	17.9	23.2	84.89
20	0	0	0	10.55	90.17	13.79	31.33	208.83
21	0	0	0	359.55	45.12	114.61	84.27	650.39
22	0	0	0	285.46	53.69	49.93	129.42	291.54
23	0	0	0	411.71	53.69	47.6	220.79	316.04
24	0	0	0	2492	4593.37	163.99	1072.54	3359.84
25	0	0	0	539.66	0	98.2	201.95	875.16
26	0	0	0	291.43	595.87	92.17	121.81	521.69
27	0	0	0	478.76	910.62	50.53	194.6	719.6
28	0	0	0	487.7	247.04	47.75	242.62	341.22
29	0	0	0	253.2	127.25	11.76	109.16	198.12
30	0	0	0	190.08	93.32	10.79	95.99	154.51
31	0	0	0	75.24	378.52	56.52	151.41	892.56
32	0	0	0	42.04	220.22	29.65	88.32	212.84
33	0	0	0	469.5	2472.48	76.08	1061.55	1775.99
34	0	0	0	91.08	0	56.77	173.38	410.3
35	0	0	0	68.48	386.4	52.87	151.41	348.49
36	0	0	0	21.61	56.4	799.34	39.21	237.09

Table 6.4: Dispersion of $x_1, \ldots, x_8$ in $\mathcal{M}$

$\overline{\Delta x_1}$	$\overline{\Delta x_2}$	$\overline{\Delta x_3}$	$\overline{\Delta x_4}$	$\overline{\Delta x_5}$	$\overline{\Delta x_6}$	$\overline{\Delta x_7}$	$\overline{\Delta x_8}$
0	0	0	190.34	488.53	77.07	145.61	552.51

Table 6.5: Averages $\overline{\Delta x_1}, \ldots, \overline{\Delta x_8}$ on the blocks of $\mathcal{M}$

Such points can be regarded as *outliers*, and are omitted in the rest of the computation. The multiplicities between 15 and 17 are typed in yellow: with high probability the corresponding points encode some failures which occured when the measurements were collected, though the situation is not as clear as above; these points may or may not be considered in the rest of the computation.

Point	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
Mult	783	621	593	463	237	210	168	153	151	150
Point	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}	p_{17}	p_{18}	p_{19}	p_{20}
Mult	130	120	119	118	117	116	116	116	115	113
Point	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}	p_{26}	p_{27}	p_{28}	p_{29}	p_{30}
Mult	111	110	104	103	102	101	91	73	66	63
Point	p_{31}	p_{32}	p_{33}	p_{34}	p_{35}	p_{36}	p_{37}	p_{38}	p_{39}	p_{40}
Mult	62	41	29	18	17	16	13	13	12	12
Point	p_{41}	p_{42}	p_{43}	p_{44}	p_{45}	p_{46}	p_{47}	p_{48}	p_{49}	p_{50}
Mult	11	8	8	8	8	7	5	5	5	5
Point	p_{51}	p_{52}	p_{53}	p_{54}	p_{55}	p_{56}	p_{57}	p_{58}	p_{59}	p_{60}
Mult	4	4	4	4	3	3	3	3	3	3
Point	p_{61}	p_{62}	p_{63}	p_{64}	p_{65}	p_{66}	p_{67}	p_{68}	p_{69}	p_{70}
Mult	3	3	3	2	2	2	2	2	2	2
Point	p_{71}	p_{72}	p_{73}	p_{74}	p_{75}	p_{76}	p_{77}			
Mult	1	1	1	1	1	1	1			

Table 6.6: Multiplicities of the points in PPP

As an example we plot some elements of $\mathcal{M}$ replaced by some points in PPP having small multiplicity. We consider the points of $\mathcal{M}$ at positions $4948-5184$; in Figures 6.6 and 6.7 we plot the values of their 5-th and 7-th coordinates. In green we plot the points at positions $4948-4960$ replaced by p_{37}, multiplicity 13; in blue we plot the points at positions $4961-5063$ replaced by p_{24}, multiplicity 103; in black the points at positions $5064-5065$ replaced by p_{69}, multiplicity 2; in magenta the points at positions $5066-5067$ replaced by p_{64}, multiplicity 2; in cyan the points at positions $5066-5080$ replaced by p_{38}, multiplicity 13; in red the points at positions $5081-5184$ replaced by p_{23}, multiplicity 104. The points plotted in green, black, magenta and cyan are associated to points whose multiplicity is smaller than 15, and so they are considered as outliers. Indeed they result from measurements occured during the transit from one valve openings combination to another: at positions $4947-4948$ valve 1 passes from 10% to 20% opening, and again at positions $5067-5068$ it goes from 20% to 30%. These points are omitted from any future computation.

We define the sets $Z_1, Z_2 \subseteq PPP$, in which the points having yellow and red multiplicities, or only red multiplicities (see again Table 6.6), are omitted:

$$\begin{aligned} Z_1 &= \{p_i \in PPP \mid \text{multiplicity}(p_i) \geq 18\} \\ Z_2 &= \{p_i \in PPP \mid \text{multiplicity}(p_i) \geq 15\} \end{aligned} \tag{6.3}$$

We observe that $\#Z_1 = 34$ and $\#Z_2 = 36$; further, in both cases the orig-

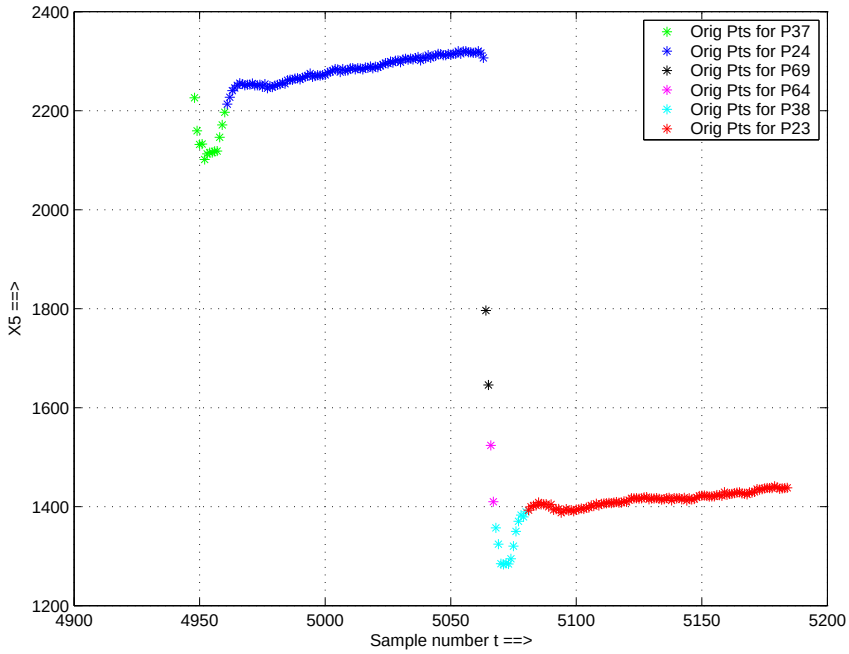


Figure 6.6: Coordinate x_5 of $\mathcal{M}$, positions 4948 – 5184

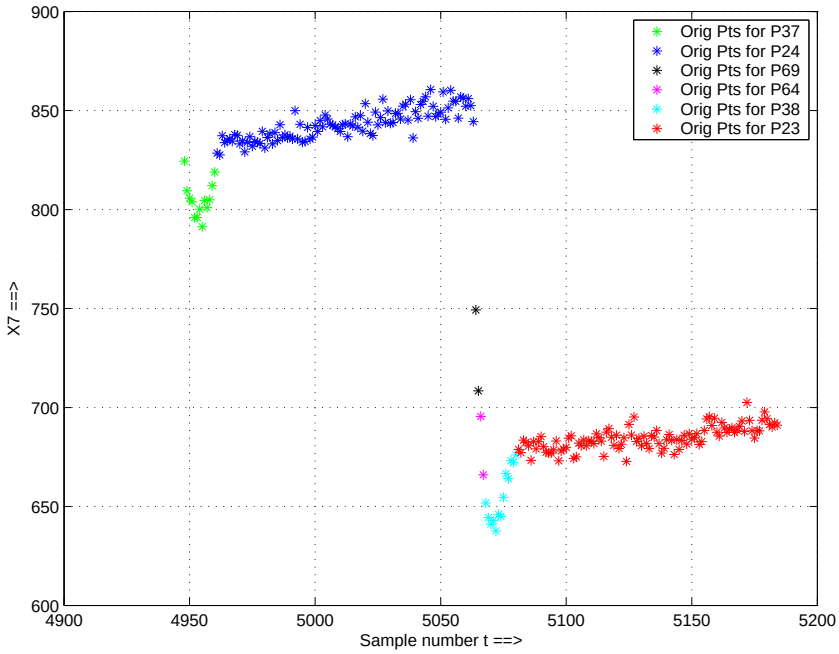


Figure 6.7: Coordinate x_7 of $\mathcal{M}$, positions 4948 – 5184

inal measurements are properly represented since, for instance, all the valve combinations in PPP are also contained in Z_1 and Z_2.

6.2.3 Computation of stable order ideals and production polynomials

In this section we aim at finding a good representation/prediction for the oil production. The following approximation problem is addressed: given the values of the oil production at a sample set we want to find the polynomial of best approximation. As in the multivariate regression theory, our purpose is indeed to find a multivariate polynomial which fits the input data. In our method the model is given by the order ideal $\mathcal{O}$ returned by the SOI Algorithm when applied to a suitable data set. Starting from $\mathcal{O}$, the production polynomial f is computed as a linear combination of its elements. The coefficients of such a linear combination are obtained by projecting the vector of the measured values of the production onto the linear span generated by the evaluation vectors $\{t(\mathcal{M}_1) \mid t \in \mathcal{O}\}$ (and using the least squares technique, see Section 2.3.1).

We consider the sets Z_1 and Z_2 defined in (6.3); we fix a tolerance vector ε and, for each Z_i, we compute a stable order ideal of $\mathcal{I}(Z_i)$ by applying the SOI Algorithm to the empirical set (Z_i, ε). Apart from the first three components ε_1, ε_2, ε_3 which are set equal to zero, the choice of the tolerance vector ε is independent of the error estimates used during the phase of data reduction (see formula (6.2) and Table 6.5). In fact, working with fewer points which represent a much larger and very redundant data set means giving much more importance to the clustered points than to the original ones, and therefore lowering the error allowed on each component. Moreover, we recall that the strategy used in the SOI Algorithm works only for small perturbations of the input data, while it loses its meaning if we consider big changes on them. For all these reasons, we fix a tolerances ε such that $|\varepsilon_i| < 1$.

Note that all the experiments and computations are done in the polynomial ring $\mathbb{Q}[x_1, \ldots, x_8]$; the SOI Algorithm is executed using the degree lexicographic term ordering σ, with $x_1 <_\sigma \ldots <_\sigma x_8$; in addition, the coefficients of the polynomials are displayed as truncated decimals.

First experiment
We apply the SOI Algorithm to (Z_1, ε), with $\varepsilon = (0, 0, 0, 0.8, 0.8, 0.8, 0.8, 0.8)$. The returned stable order ideal is:

$$\begin{aligned}\mathcal{O}_1 \;=\; \{&1,\ x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6,\ x_7,\ x_8,\ x_1^2,\ x_1x_3,\ x_1x_4,\ x_1x_5,\ x_1x_6,\\ &x_1x_7,\ x_1x_8,\ x_2^2,\ x_2x_3,\ x_2x_4,\ x_2x_5,\ x_2x_6,\ x_2x_7,\ x_2x_8,\ x_3^2,\ x_3x_4,\\ &x_3x_5,\ x_3x_6,\ x_3x_7,\ x_3x_8,\ x_4^2,\ x_4x_5,\ x_5^2,\ x_5x_6,\ x_5x_8\}\end{aligned}$$

Note that $\mathcal{O}_1$ contains 34 terms and so it is a stable quotient basis for the vanishing ideal $\mathcal{I}(Z_1)$.

Starting from $\mathcal{O}_1$ we compute the production polynomial f_1:

$$\begin{aligned} f_1 \;=\; & x_1\,(0.001x_8 - 0.004x_7 - 0.032x_6 - 4\cdot 10^{-4}x_5 + 9\cdot 10^{-4}x_4 + 0.011x_1 \\ & -9.205) + x_2\,(0.001x_8 - 0.004x_7 - 0.017x_6 - 2\cdot 10^{-4}x_5 + 3\cdot 10^{-4}x_4 \\ & +0.073x_2 - 12.317) + x_3\,(-3\cdot 10^{-4}x_8 + 2\cdot 10^{-4}x_7 + 6\cdot 10^{-4}x_6 \\ & -3\cdot 10^{-5}x_5 + 10^{-5}x_4 + 3\cdot 10^{-5}x_3 - 0.003x_2 - 0.002x_1 + 0.312) \\ & -2\cdot 10^{-5}x_4^2 + 2\cdot 10^{-5}x_4x_5 - 10^{-5}x_5^2 - 2\cdot 10^{-5}x_5x_6 + 4\cdot 10^{-6}x_5x_8 \\ & -0.016x_4 - 0.004x_5 + 3.547x_6 + 0.206x_7 - 0.063x_8 + 1005.56 \end{aligned}$$

We compare the values of f_1 evaluated at $\mathcal{M}$ with the measured values of the gross production Q. In Figure 6.9 we plot in blue the real values of the production Q, in green the evaluations of f_1 at $\mathcal{M}_1$, and in red the evaluations of f_1 at $\mathcal{M}_2$ (the "predictions"); a zooming of the prediction part is contained in Figure 6.10. We observe that the polynomial f_1 is a good predictor: the absolute values of the differences between the evaluations of f_1 and the measured values of the oil production are plotted in Figure 6.8.

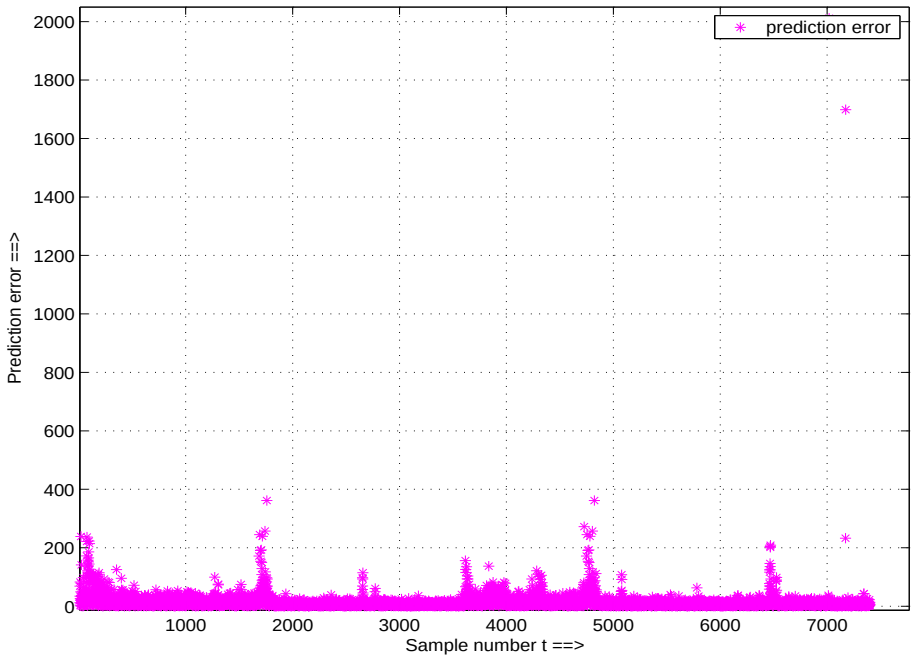


Figure 6.8: Error in the prediction of f_1

It is evident that the polynomial f_1 has extremely high values at two points of $\mathcal{M}$, namely the measurements recorded at positions 7019 and 7173. We look

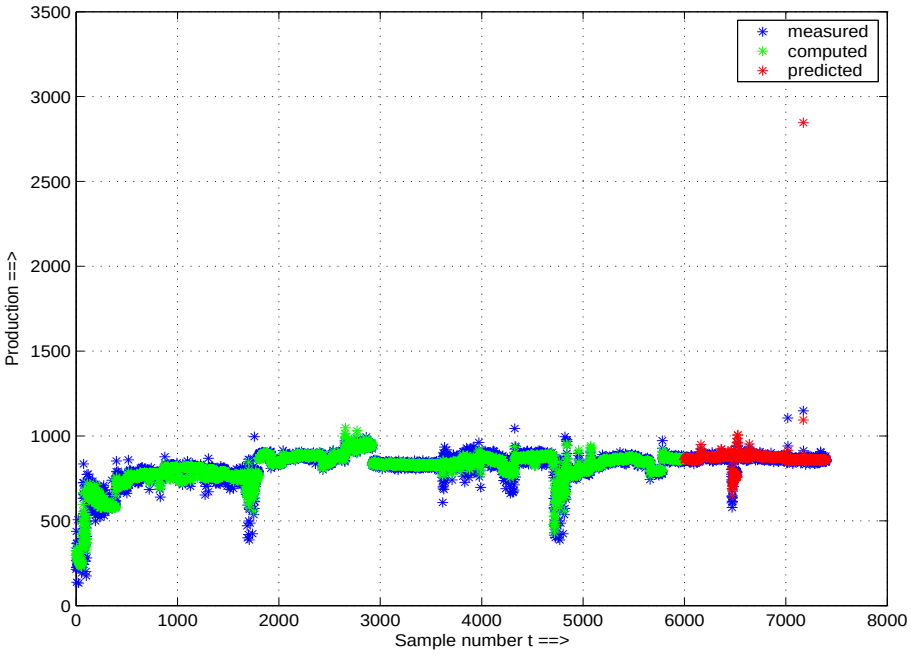


Figure 6.9: Oil production from polynomial f_1

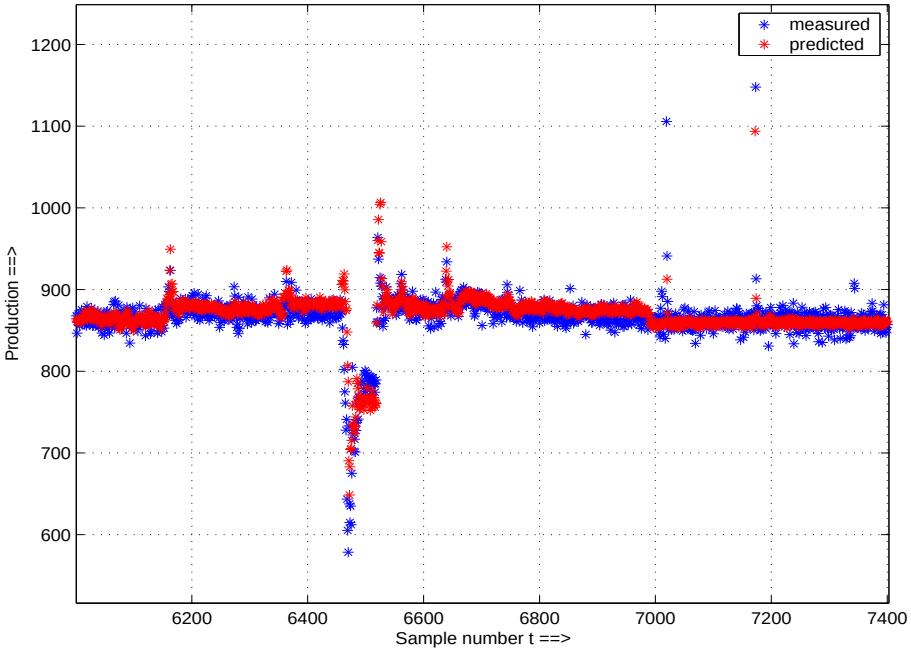


Figure 6.10: Oil production from polynomial f_1: the predicted part

at the measurements registered in $\mathcal{M}$ at the position 7019 and its neighbours:

7018 :	(40,	70,	2800,	223.47,	680.20,	133.54,	791.10,	10132.71)
7019 :	(40,	70,	2800,	220.12,	661.39,	930.22,	770.65,	9961.28)
7020 :	(40,	70,	2800,	215.77,	641.83,	154.18,	761.60,	10151.90)
7021 :	(40,	70,	2800,	224.75,	661.55,	138.52,	772.98,	10124.02)

and at the values of the production Q:

7018 :	867.35
7019 :	1105.71
7020 :	941.12
7021 :	885.54

We notice that indeed a big variation occurs in $\mathcal{M}$, position 7019, at variable x_6, which corresponds to the gas production: the quantity of gas produced during the extraction varies from 133.54 to 930.22 (in red), and then back to 154.18 and 138.52. A similar alteration is registered for the gross production: its values fluctuate from 867.35 to 1105.71 (in red) and then back to 941.12 and 885.54. Since the physical quantities are collected at every minute, it is highly probable that such changes are related to faults in the processing of data acquisition, which can be regarded as false measurements. The same situation occurs at the position 7173; the values of the neighbour points in $\mathcal{M}$ are:

7171 :	(40,	70,	2800,	228.48,	691.24,	134.80,	793.58,	10123.63)
7172 :	(40,	70,	2800,	229.14,	687.34,	216.37,	789.45,	10101.54)
7173 :	(40,	70,	2800,	223.46,	653.33,	835.71,	759.31,	10027.89)
7174 :	(40,	70,	2800,	218.64,	643.32,	145.71,	760.54,	10095.67)

and in the production Q:

7171 :	872.00
7172 :	861.21
7173 :	1147.96
7174 :	913.20

Despite the satisfactory results achieved in the prediction of the oil production, we can't conclude that the polynomial f_1 is the solution we looked for: f_1 does not give information about the unknown interactions occurring in a production unit, and so cannot provide a complete solution to the problem of oil production. From our point of view, a good numerical approximation and prediction of the production values are not enough to deal completely with the production problem. We require stronger results involving the structure of the model for the production in terms of the driving inputs.

Second experiment
We apply the SOI Algortihm to (Z_2, ε), with $\varepsilon = (0, 0, 0, 0.8, 0.8, 0.8, 0.8, 0.8)$.

The returned stable order ideal is:

$$\begin{aligned}\mathcal{O}_2 \;=\; \{&1,\ x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6,\ x_7,\ x_8,\ x_1^2,\ x_1x_3,\ x_1x_4,\ x_1x_5,\ x_1x_6,\\ &x_1x_7,\ x_1x_8,\ x_2^2,\ x_2x_3,\ x_2x_4,\ x_2x_5,\ x_2x_6,\ x_2x_7,\ x_2x_8,\ x_3^2,\ x_3x_4,\\ &x_3x_5,\ x_3x_6,\ x_3x_7,\ x_3x_8,\ x_4^2,\ x_4x_5,\ x_5^2,\ x_5x_6,\ x_5x_8,\ x_1^3,\ x_1x_5^2\}\end{aligned}$$

Note that $\mathcal{O}_2$ contains 36 terms and so it is a stable quotient basis for the vanishing ideal $\mathcal{I}(Z_2)$.

In this case we compute the production polynomial in a slightly different way: we select from $\mathcal{O}_2$ the "maximal" subset S such that $S \in \langle x_1, x_2, x_3\rangle$,

$$\begin{aligned}S \;=\; \{&x_1,\ x_2,\ x_3,\ x_1^2,\ x_1x_3,\ x_1x_4,\ x_1x_5,\ x_1x_6,\ x_1x_7,\ x_1x_8,\ x_2^2,\ x_2x_3,\ x_2x_4,\\ &x_2x_5,\ x_2x_6,\ x_2x_7,\ x_2x_8,\ x_3^2,\ x_3x_4,\ x_3x_5,\ x_3x_6,\ x_3x_7,\ x_3x_8,\ x_1^3,\ x_1x_5^2\}\end{aligned}$$

By applying the least squares method w.r.t. the monomial set S, we obtain the production polynomial f_2:

$$\begin{aligned}f_2 \;=\; &x_1\,(28.523 - 0.530x_1 + 3\cdot 10^{-4}x_4 - 9\cdot 10^{-5}x_5 + 0.010x_6 - 0.002x_7\\ &+7\cdot 10^{-4}x_8 + 0.003x_1^2 - 10^{-9}x_5^2) + x_2\,(10.068 - 0.072x_2 + 0.002x_4\\ &-10^{-4}x_5 + 0.036x_6 - 0.001x_7 + 2\cdot 10^{-4}x_8) + x_3\,(-0.363 + 0.003x_1\\ &+0.002x_2 + 10^{-5}x_3 + 3\cdot 10^{-5}x_4 - 6\cdot 10^{-5}x_5 - 10^{-4}x_6 + 10^{-4}x_7\\ &-2\cdot 10^{-5}x_8)\end{aligned}$$

We compare the values of f_2 evaluated at $\mathcal{M}$ with the measured values of the gross production Q. In Figure 6.11 we plot in blue the real values of the production Q, in green the evaluations of f_2 at $\mathcal{M}_1$, and in red the evaluations of f_2 at $\mathcal{M}_2$ (the "predictions"); a zooming of the prediction part is contained in Figure 6.12. The differences between the evaluations of f_2 and the measured values of the oil production are plotted in Figure 6.13. We observe that f_2 takes extremely high values when evaluated at the measurements recorded at positions 7019 and 7173 of $\mathcal{M}$; as argued for f_1, we conclude that such big variations are related to failures in the measurements rather than to faults in the computation, and so they do not undermine the reliability of the model.

Now, we compare the reliability of the two computed models; in Figure 6.14 we plot the difference between the prediction errors deriving from f_2 and f_1. Since most of the plot lies in the first quadrant, we conclude that the predictions of f_1 are more reliable than those of f_2. Despite this, f_2 gives a more satisfactory answer to the problem of oil production, since it provides more information about the interactions of the two zones. We decompose f_2 as follows:

$$f_2 = f_{21} + f_{22} + f_{23} \tag{6.4}$$

where $f_{21}, f_{22}, f_{23} \in \mathbb{Q}[x_1, \ldots, x_8]$ are given by:

$$\begin{aligned}f_{21} &= x_1\, g_1\\ f_{22} &= x_2\, g_2\\ f_{23} &= x_3\, g_3 + g_4\end{aligned}$$

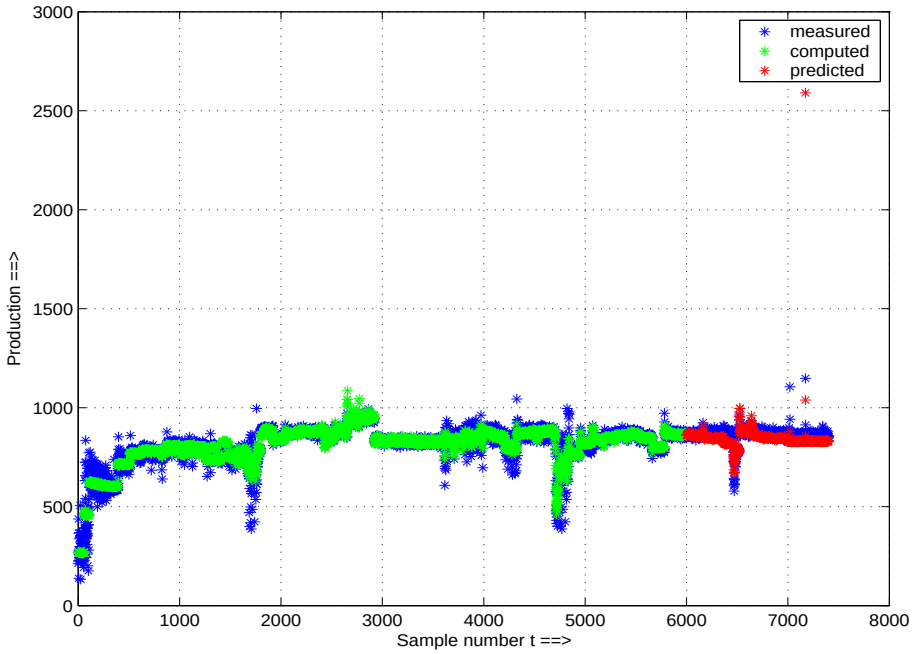


Figure 6.11: Oil production from polynomial f_2

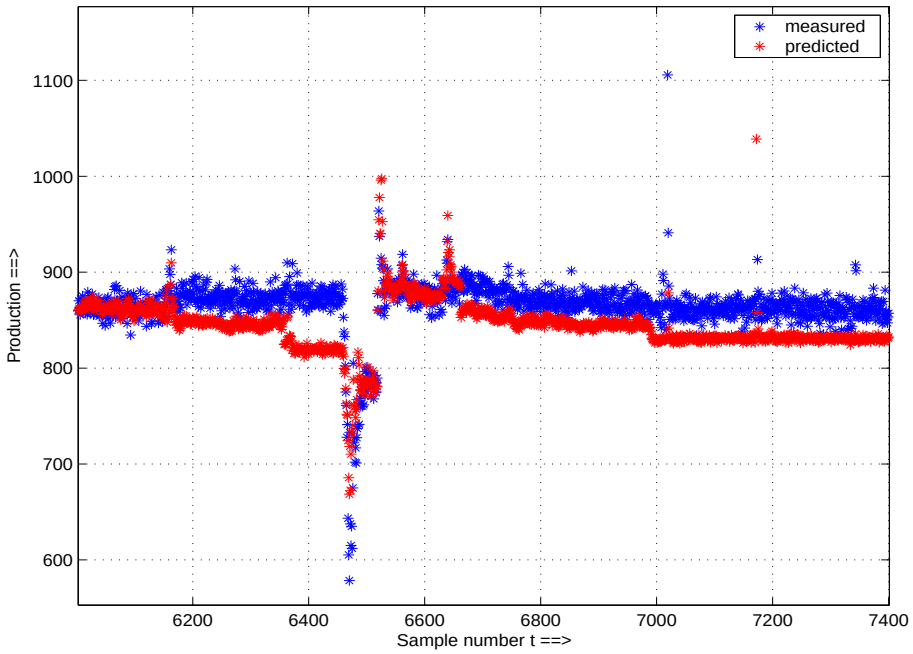


Figure 6.12: Oil production from polynomial f_2: the predicted part

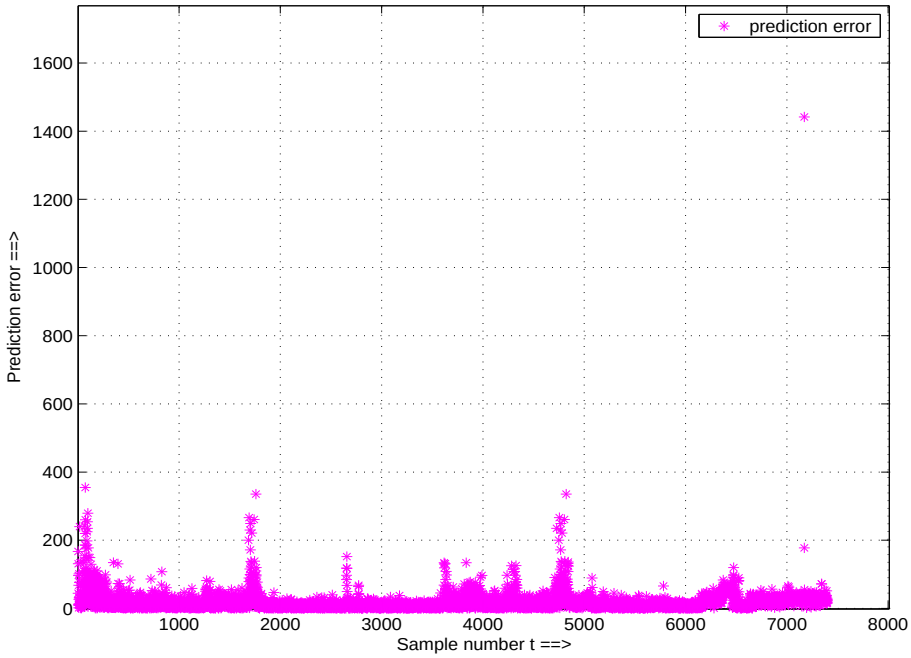


Figure 6.13: Error in the prediction of f_2

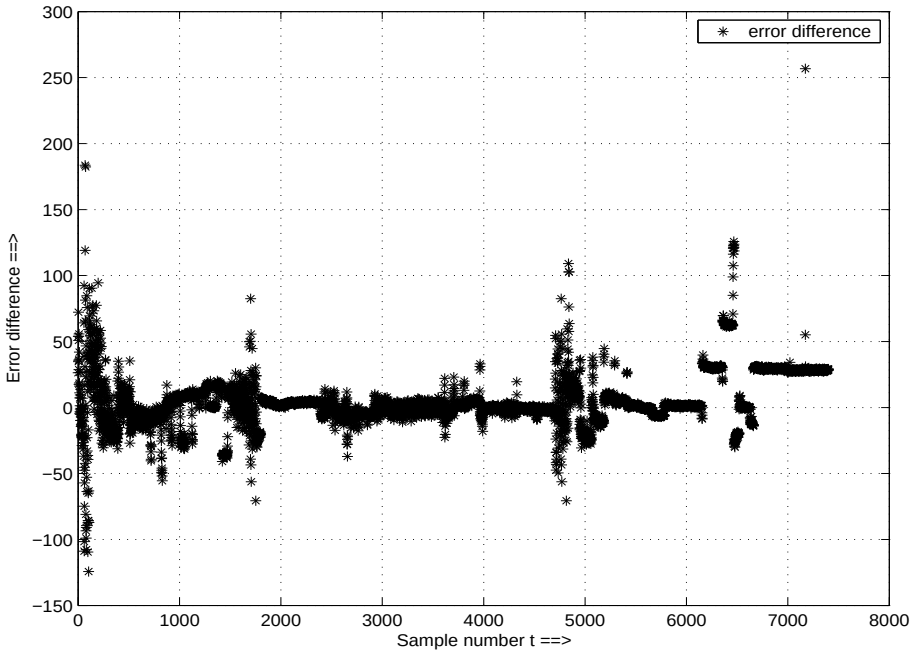


Figure 6.14: Difference between prediction errors committed by f_1 and f_2

and $g_1, g_2, g_3, g_4 \in \mathbb{Q}[x_1, \ldots, x_8]$ are

$$\begin{aligned}
g_1 &= -10^{-9}x_5^2 + 0.003x_1^2 + 7 \cdot 10^{-4}x_8 + 0.010x_6 - 9 \cdot 10^{-5}x_5 - 0.530x_1 + \\
&\quad +28.523 \\
g_2 &= 2 \cdot 10^{-4}x_8 - 0.001x_7 + 0.036x_6 + 0.002x_4 - 0.072x_2 + 10.068 \\
g_3 &= -2 \cdot 10^{-5}x_8 + 10^{-4}x_7 - 10^{-4}x_6 - 6 \cdot 10^{-5}x_5 + 3 \cdot 10^{-5}x_4 + 10^{-5}x_3 + \\
&\quad +0.002x_2 + 0.003x_1 - 0.363 \\
g_4 &= 10^{-4}x_2x_5 - 0.002x_1x_7 + 3 \cdot 10^{-4}x_1x_4
\end{aligned}$$

By using decomposition (6.4) we are able to detect the contributions of each single well to the total oil production, and have a tool to investigate their unknown interactions. In fact, note that

$$\begin{aligned}
\text{supp } (f_{21}) &\subseteq \mathcal{I}(x_1) \\
\text{supp } (f_{22}) &\subseteq \mathcal{I}(x_2) \\
\text{supp } (f_{23}) &\subseteq \mathcal{I}(x_1x_4, x_1x_7, x_2x_5, x_3)
\end{aligned}$$

and so, using (6.1), we have

$$f_2 \mid_{x_2=0} = f_{21} \quad \text{and} \quad f_2 \mid_{x_1=0} = f_{22}$$

As a consequence, f_{21} is a model for the production of zone 1, and f_{22} represents the production of zone 2. We consider the polynomials g_1 and g_2 as describing the *nature* of the inflow at zone 1 and zone 2, in the sense that for all points of $\mathcal{M}$ at which the evaluation of a monomial ct, where t is a term in the support of g_1 or g_2, and c is the corresponding coefficient, is positive the inflow is stimulated, whereas if the evaluation is negative the inflow is inhibited. The same holds for the polynomials g_3 and g_4, which are used to give a description of the nature of the interactions occuring inside the production tubing. The polynomial f_{23} provides with information on the state of the two (interacting) zones, since it is non-zero when zone 1 and zone 2 are simultaneously producing.

We now give a more detailed interpretation of the physical meaning of the polynomials f_{21}, f_{22} and f_{23}. Firstly we consider f_{21} and f_{22}. The contribution to the production represented by each polynomial is naturally adjusted via the variables x_1 and x_2, that is via the openings of the valves. Then we focus our attention on g_1 and g_2. Note that g_1 and g_2 can be further decomposed into two parts: a part which is related to the pushing from the sub-surface and which represents the driving of the fluids through the valve openings, and a part which represents the influence on the individual productions due to "external" factors. In our particular case we decompose g_1 and g_2 in the following way:

$$\begin{aligned}
g_1 &= g_{11} + g_{12} \\
g_2 &= g_{21} + g_{22}
\end{aligned}$$

where

$$\begin{aligned} g_{11} &= -10^{-9}x_5^2 + 0.003x_1^2 + -9 \cdot 10^{-5}x_5 - 0.530x_1 + 28.523 \\ g_{12} &= 7 \cdot 10^{-4}x_8 + 0.010x_6 \\ g_{21} &= 0.002x_4 - 0.072x_2 + 10.068 \\ g_{22} &= 2 \cdot 10^{-4}x_8 - 0.001x_7 + 0.036x_6 \end{aligned}$$

In the polynomial g_{11} the coefficients of the terms x_5 and x_5^2, which represent the pressure that drives the oil out of the first zone, see Table 6.2, are small and negative: under the tacit assumption of a (relatively) high and constant reservoir pressure this means that a (relatively) large volume of fluids per unit of time is flowing from the reservoir into the well's tubing system. For the polynomial g_{21} the situation is the opposite: the coefficient of the term x_4, representing the pressure that drives the oil out of the second zone, is relatively big and positive, and so the production is inhibited.

The "external" influence which comes into play in the polynomials g_{12} and g_{22} is due to two main factors: the presence of oil and gas in the production tubing, and the interaction between surface and sub-surface events. In both polynomials g_{12} and g_{22} the coefficient of the term x_6 is positive: the gas present inside the production tubing mixes with the oil, makes it lighter, so that the production rate increases. The term x_7 appears only in g_{22} and with a negative coefficient. In fact zone 1 is above zone 2, and so the fluid already present in the production tubing does not affect the inflow of oil occuring through valve 1. On the contrary the contribution of zone 2 is inhibited by the pressure of the fluid which occurs on valve 2. Finally in both polynomials g_{12} and g_{22} the coefficient of the term x_8 is small (order of magnitude 10^{-4}) and positive, which means that a slight pressure in the transportation tubing stimulates the production of the two zones.

Before we perform a similar analysis for the polnomial f_{13}, we compare the values of the polynomials f_{21} and f_{22} evaluated at subsets of $\mathcal{M}$ with the measured values of the gross production Q. In the following figures we plot in blue the real values of the production Q, in red and in green the evaluations of f_{21} and f_{22} at the points of the chosen subset of $\mathcal{M}$. Figure 6.15 contains the first 10 experiments consisting of the first 1523 measurements of $\mathcal{M}$: the valve positioned at zone 2 is kept closed, and only zone 1 is producing (see Table 6.3), so no interaction between the two zones can occur (that is $f_{23} = 0$). In Figures 6.16, 6.17, and 6.18 we analyze a similar situation: in $\mathcal{M}$ we select the experiments in which valve 1 is kept closed, and only zone 2 is producing. Again no interaction between the two zones can occur. The three plots correspond to the three blocks in $\mathcal{M}$ made up by the measurements $1524 - 1800$, $4735 - 4826$, and $6471 - 6518$. Note that in these cases the prediction polynomial f_2 does an excellent job.

We consider the polynomial f_{23}. The amount of interaction occuring between the two different zones while they are simultaneously producing is adjusted via the variables x_1, x_2, and x_3, that is via the openings of the valves. Then we

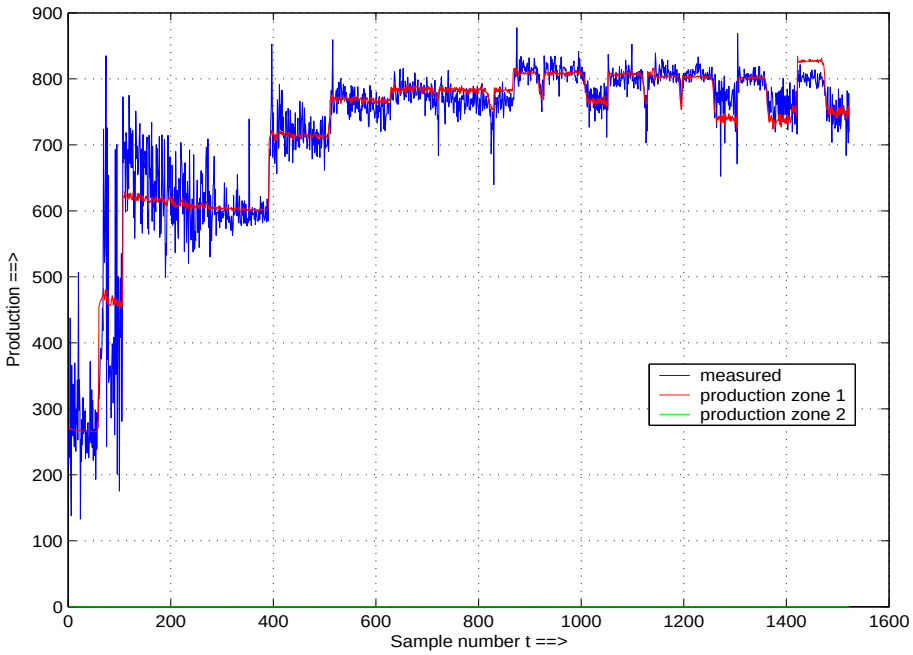


Figure 6.15: Evaluation of f_{21} and f_{22} at points $1 - 1523$ of $\mathcal{M}$

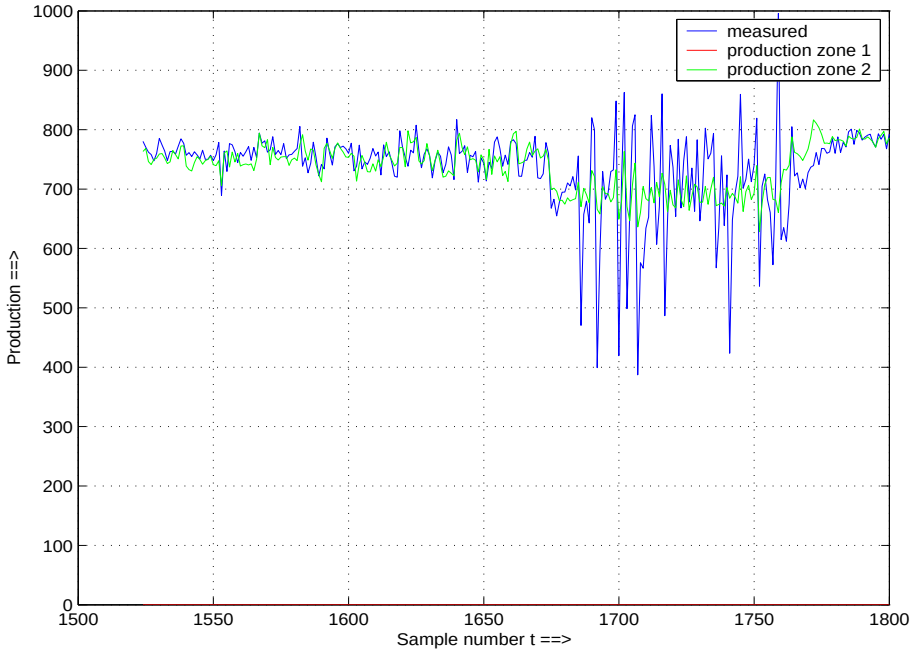


Figure 6.16: Evaluation of f_{21} and f_{22} at points $1524 - 1800$ of $\mathcal{M}$

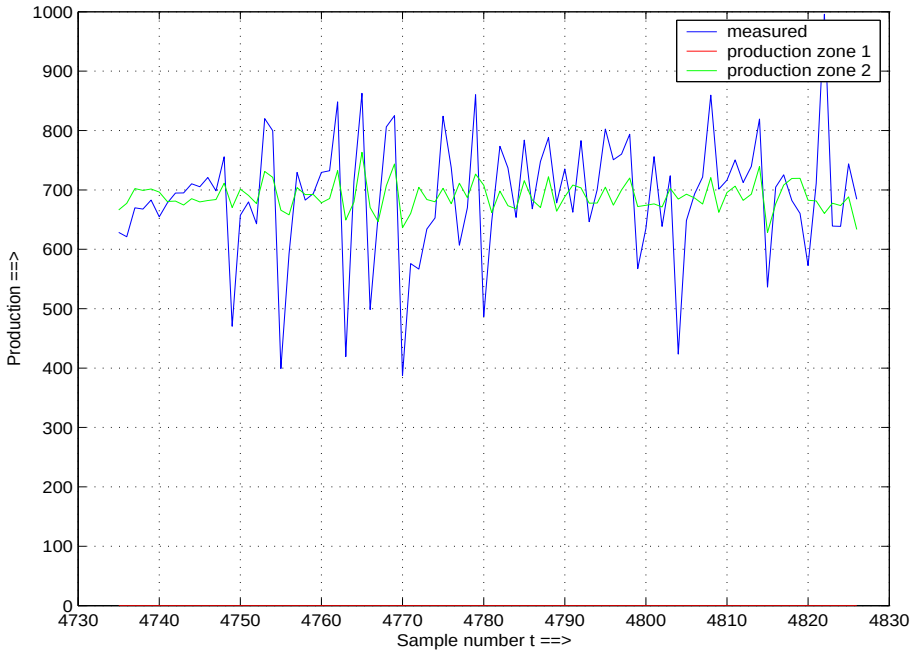


Figure 6.17: Evaluation of f_{21} and f_{22} at points $4735 - 4826$ of $\mathcal{M}$

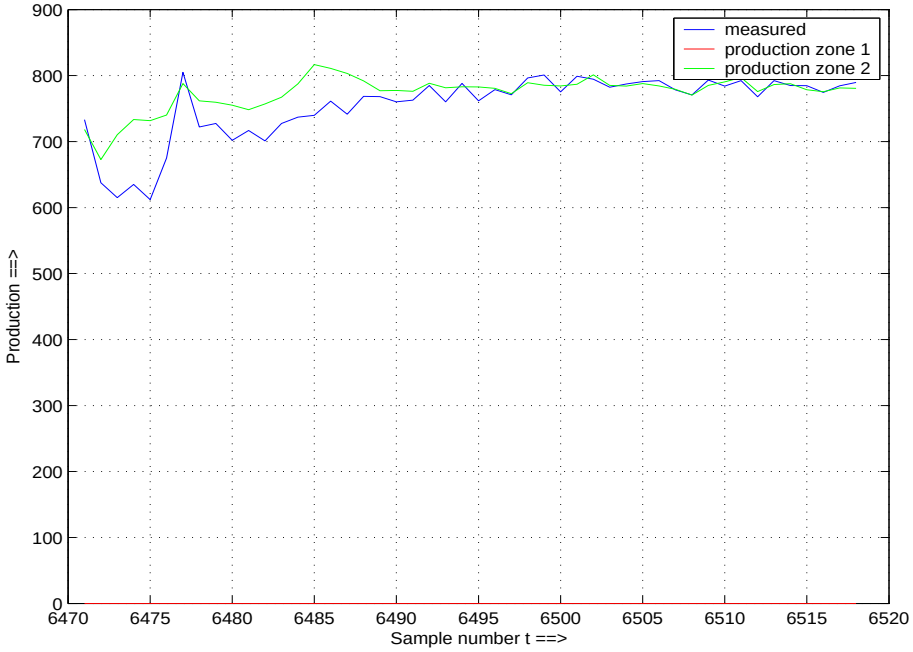


Figure 6.18: Evaluation of f_{21} and f_{22} at points $6471 - 6518$ of $\mathcal{M}$

focus our attention on the polynomials g_3 and g_4 (we recall that $f_{23} = x_3 g_3 + g_4$). In g_3 the coefficients of the terms x_1 and x_2 are approximately the same, which suggests that the two zones exhibit the same behaviour. Nevertheless, if we consider the coefficients of the terms x_4 and x_5 we conclude that the contribution from zone 1 stimulates the production, while the simultaneous opening of valve 2 produces an inhibitory effect. We characterize further the polynomial f_{23} by analyzing its evaluations at subsets of $\mathcal{M}$ corresponding to the case of the two zones producing together; the corresponding blocks of measurements inside $\mathcal{M}$ are: block $1801-4734$, block $4827-6470$, and block $6519-7400$. In Figures 6.19, 6.20 and 6.21 we plot in cyan the evaluations of f_{23} at the points of $\mathcal{M}$; as for the separate productions, we plot in blue the real values of the production Q, in red and in green the evaluations of f_{21} and f_{22} at the points of $\mathcal{M}$. Note that in some cases the evaluation of f_{21} is nearly constant, despite changes in the evaluations of f_{22} and f_{23} over the same subset. An example is given by the measurements $5067 - 6155$ (see again Figure 6.20): in this case the valve positioned at zone 1 is constantly open for 30%, the opening of the valve positioned at zone 2 varies from 40% to 70%. We notice that the contribution to the total production measured by the polynomial f_{21} is nearly constant, while the evaluations of f_{22} and f_{23} at the selected points varies with the opening of valve 2. Our claim is that the multi-phase flow of the fluids from the reservoir at valve 1 is not influenced by changes which occur in its down-stream path. In oil production operations this type of information is crucial.

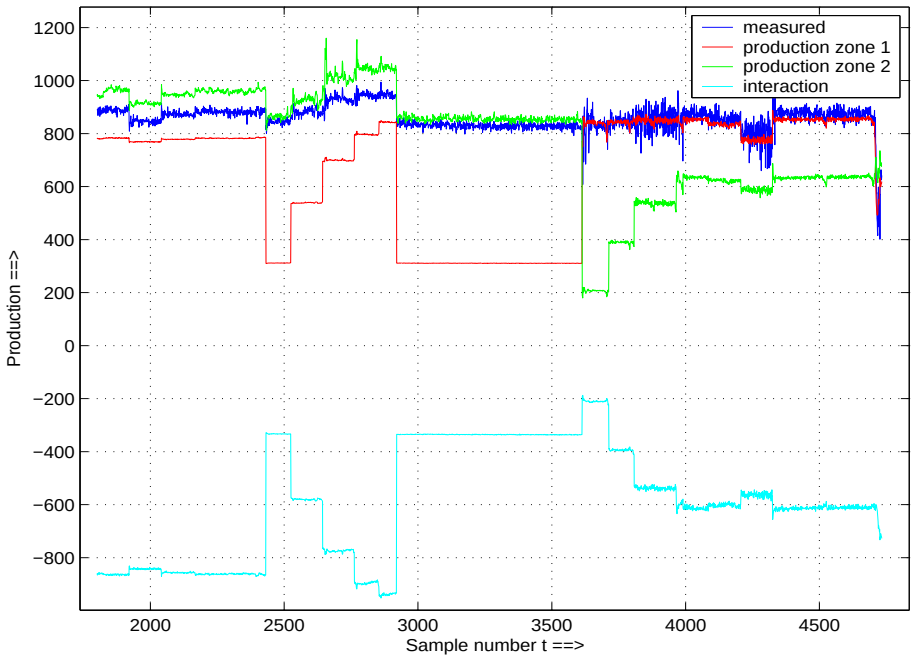


Figure 6.19: Evaluation of f_{21}, f_{22} and f_{23} at points $1801 - 4734$ of $\mathcal{M}$

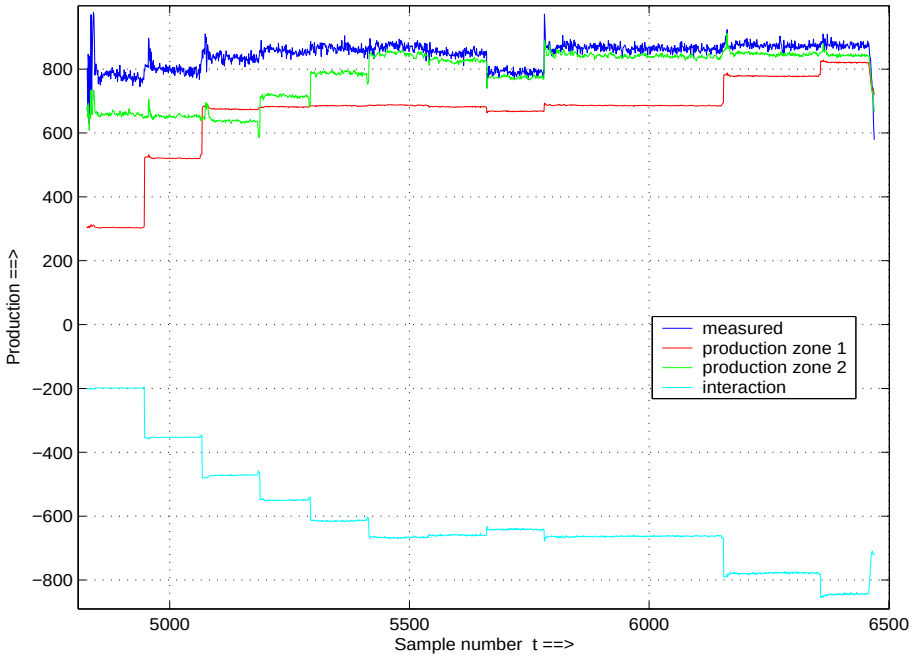


Figure 6.20: Evaluation of f_{21}, f_{22} and f_{23} at points $4827 - 6470$ of $\mathcal{M}$

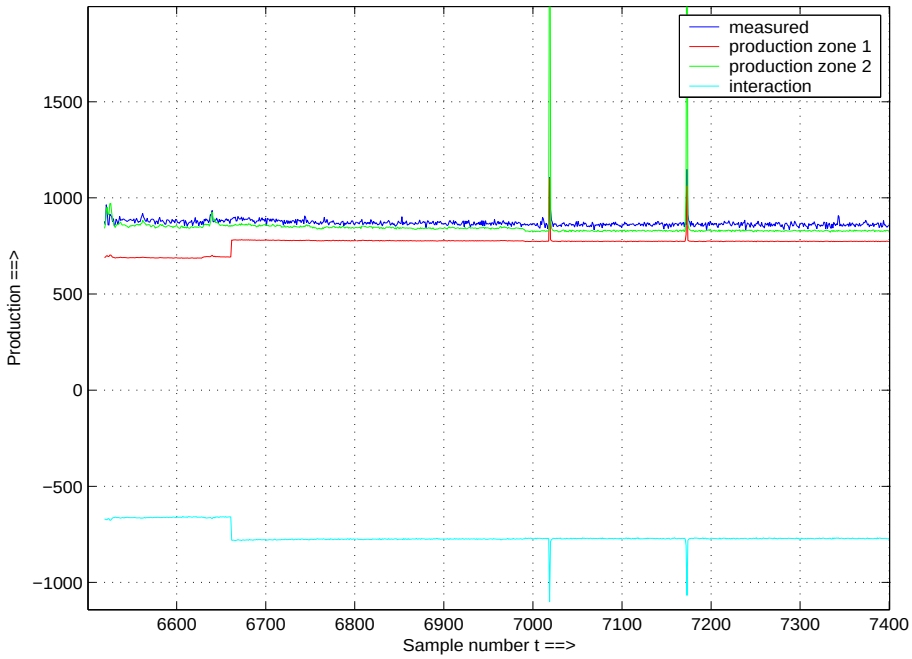


Figure 6.21: Evaluation of f_{21}, f_{22} and f_{23} at points $6519 - 7400$ of $\mathcal{M}$

Chapter 7

Future works

Following the investigations decribed in this thesis we make a list of problems that are left open to be further developed.

- The study of the properties of the polynomials "almost vanishing" at the input points.
- The study of the relationship between the SOI Algorithm and the other "numerical" algorithms available (*e.g.* the AVI Algorithm [HKPP06], the NBM Algorithm [Fas08]).
- A variation to the SOI Algorithm which is invariant under the scaling of the input empirical points.
- The construction of a set of polynomials really vanishing at a set of admissible points (for instance using the theory of the complete intersections).

Notation

Sets and Special Sets

$A \subseteq B$	set A is a (not necessarily proper) subset of B
$A \subset B$	set A is a proper subset of B
$A \setminus B$	set difference of A and B
$\#A$	number of elements of a finite set A
$\mathbb{N}$	set of natural numbers, $\mathbb{N} = \{0, 1, 2, \ldots\}$
$\mathbb{Z}$	set of integers
$\mathbb{Q}$	set of rational numbers
$\mathbb{R}$	set of real numbers
$\mathbb{R}^+$	set of strictly positive real numbers
$\mathbb{T}^n$	set of terms in the indeterminates $x_1, \ldots, x_n$
$\mathcal{O}$	factor closed set of terms of $\mathbb{T}^n$
$\mathbb{X}$	affine point set
$\mathbb{X}^\varepsilon$	finite set of empirical points
$\widetilde{\mathbb{X}}$	admissible perturbation of $\mathbb{X}^\varepsilon$
$N^\alpha(p^\varepsilon)$	neighbourhood of perturbations of an empirical point

Rings and Fields

R	ring
K	field
$Q(R)$	field of fractions of an integral domain R
$K[x_1, \ldots, x_n]$	polynomial ring in the indeterminates $x_1, \ldots, x_n$ over K
$K(x_1, \ldots, x_n)$	field of rational functions in the indeterminates $x_1, \ldots, x_n$ over K

Vectors and Matrices

$\|v\|_\alpha$	α-norm of a vector
$\|v\|$	Euclidean norm of a vector ($\alpha = 2$)
$\|v\|_{\alpha,E}$	E-weighted α-norm of a vector
$\|v\|_E$	E-weighted Euclidean norm of a vector
$\mathrm{Mat}_{m,n}(R)$	set of $m \times n$ matrices over R
M^{-1}	inverse of a matrix
$M^\dagger$	Moore-Penrose pseudoinverse of a matrix
$\det(M)$	determinant of a matrix
$\mathrm{rank}(M)$	rank of a matrix
$M_G(\mathbb{X})$	evaluation matrix of G at $\mathbb{X}$

Orderings

$\geq_\sigma$	monoid ordering
`Lex`	lexicographic term ordering
`DegLex`	degree-lexicographic term ordering
`DegRevLex`	degree-reverse-lexicographic term ordering

Polynomials

$\deg(f)$	degree of a polynomial
$\mathrm{Supp}(f)$	support of a polynomial
$\mathrm{LT}_\sigma(f)$	leading term of a polynomial
$\mathrm{LC}_\sigma(f)$	leading coefficient of a polynomial
$\mathrm{LM}_\sigma(f)$	leading monomial of a polynomial
$\mathrm{LT}_\sigma(f)$	leading term of a polynomial

Ideals

$\langle f_\lambda \mid \lambda \in \Lambda \rangle$	ideal generated by the set $\{f_\lambda \mid \lambda \in \Lambda\}$
$\mathrm{LT}_\sigma(I)$	leading term ideal of I
$\mathrm{LT}_\sigma\{I\}$	monoideal of terms in I
$\mathcal{O}_\sigma(I)$	the set $\mathbb{T}^n \setminus \mathrm{LT}_\sigma\{I\}$
$\mathcal{I}(\mathbb{X})$	vanishing ideal of a set of points

Bibliography

[Abb06] J. Abbott, *The design of CoCoALib*, ICMS, 2006, pp. 205–215.

[Abb07] ———, *Twin-float arithmetic*, submitted.

[ABKR00] J. Abbott, A. Bigatti, M. Kreuzer, and L. Robbiano, *Computing ideals of points*, J. Symb. Comput. **30** (2000), no. 4, 341–356.

[AFG+00] A. M. Anile, B. Falcidieno, G. Gallo, M. Spagnuolo, and S. Spinello, *Modeling uncertain data with fuzzy b-splines*, Fuzzy Sets and Systems **113** (2000), no. 3, 397–410.

[AFT07] J. Abbott, C. Fassino, and M. Torrente, *Thinning out redundant empirical data*, Mathematics in Computer Science **1** (2007), no. 2, 375–392.

[AFT08] ———, *Stable border bases for ideals of points*, J. Symb. Comput. (2008), article in press, available online.

[AKR05] J. Abbott, M. Kreuzer, and L. Robbiano, *Computing zero-dimensional schemes*, J. Symb. Comput. **39** (2005), no. 1, 31–49.

[AS88] W. Auzinger and H. J. Stetter, *An elimination algorithm for the computation of all zeros of a system of multivariate polynomial equations*, Conference in Numerical Analysis, Internat. Series on Numer. Math., vol. 86, Birkhäuser Verlag, 1988, pp. 11–30.

[BM82] B. Buchberger and H. M. Möller, *The construction of multivariate polynomials with preassigned zeros*, EUROCAM '82, Lecture Notes in Comp.Sci., vol. 144, Springer, Marseille, 1982, pp. 24–31.

[CLO92] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms : An introduction to computational algebraic geometry and commutative algebra*, Springer-Verlag, 1992.

[CoC] CoCoATeam, CoCoA*: a system for doing Computations in Commutative Algebra*, Available at `http://cocoa.dima.unige.it`.

[DBA74] G. Dahlquist, Å. Björck, and N. Anderson, *Numerical methods*, Prentice-Hall, 1974.

[DH93] J. W. Demmel and N. J. Higham, *Improved error bounds for underdetermined system solvers*, SIAM J. Math. Anal. **14** (1993), no. 1, 1–14.

[Fas08] C. Fassino, *Vanishing ideal of limited precision points*, J. Symb. Comput. (2008), submitted.

[GL89] G. H. Golub and C. F. Van Loan, *Matrix computations*, second ed., The Johns Hopkins University Press, Baltimore, 1989.

[HH01] N. J. Hyne and N. J. P. Hyne, *Nontechnical guide to petroleum geology, exploration, drilling, and production*, second ed., Pennwell Books, 2001.

[HKPP06] D. Heldt, M. Kreuzer, S. Pokutta, and H. Poulisse, *Approximate computation of zero-dimensional polynomial ideals*, J. Symb. Comput. (2006), submitted.

[HKY99] L. J. Heyer, S. Kruglyak, and S. Yooseph, *Exploring expression data: identification and analysis of coexpressed genes*, Genome Res. **9** (1999), 1106–1115.

[JMK91] A. T. Jessup, W. K. Melville, and W. C. Keller, *Breaking waves affecting microwave backscatter: 1. detection and verification*, Journal of Geophysical Research **96** (1991), no. 20, 547–559.

[KKR05] A. Keherein, M. Kreuzer, and L. Robbiano, *An algebraist's view on border bases*, Solving Polynomial Equations: Foundations, Algorithms, and Applications (Heidelberg) (A. Dickenstein and I. Emiris, eds.), vol. 14, Springer Verlag, 2005, pp. 160–202.

[KM00] W. J. Krzanowski and F. H. C. Marriott, *Multivariate analysis*, second ed., Springer, Berlin, 2000.

[KPR08] M. Kreuzer, H. Poulisse, and L. Robbiano, *From oil fields to hilbert schemes*, in preparation.

[KR00] M. Kreuzer and L. Robbiano, *Computational commutative algebra. 1*, Springer-Verlag, Berlin, 2000.

[KR05] ———, *Computational commutative algebra. 2*, Springer-Verlag, Berlin, 2005.

[KR08] ———, *Deformations of border bases*, Collectanea Mathematica **59** (2008), no. 3, 275–297.

[Lak91] Y. Lakshman, *A single exponential bound on the complexity of computing gröbner bases of zero dimensional ideals*, MEGA, Progress in Math., vol. 94, Birkhäuser, 1991, pp. 227–234.

[Li06] B. Li, *A new approach to cluster analysis: the clustering-function-based method*, J. Roy. Statist. Soc. Ser. B **3** (2006), no. 68, 457–476.

[Möl93] H. M. Möller, *Systems of algebraic equations solved by means of endomorphisms*, AAECC **673** (1993), 43–56.

[Mou99] B. Mourrain, *A new criterion for normal form algorithms*, AAECC, 1999, pp. 430–443.

[MR02] B. Mourrain and O. Ruatta, *Relation between roots and coefficients, interpolation and application to system solving*, J. Symb. Comput. **33** (2002), no. 5, 679–699.

[MS95] H. M. Möller and H. J. Stetter, *Multivariate polynomial equations with multiple zeros solved by matrix eigenproblems*, Numer. Math. **70** (1995), 311–329.

[MS00] H. M. Möller and T. Sauer, *H-bases ii: applications to numerical problems*, Curve and surface fitting, Vanderbilt Univ. Press, Nashville, 2000, pp. 1–10.

[Rob08] L. Robbiano, *On border basis and gröbner basis schemes*, Collectanea Mathematica (2008), to appear, available at `arXiv:0802.2793`.

[RR03] F. Rapallo and M. P. Rogantin, *Statistica descrittiva multivariata*, second ed., C.L.U.T., Torino, 2003.

[Sau07] T. Sauer, *Approximate varieties, approximate ideals and dimension reductions*, Numerical Algorithms **45** (2007), no. 1–4, 295–313.

[Ste04] H. J. Stetter, *Numerical polynomial algebra*, SIAM, Philadelphia, PA, USA, 2004.