Exact algorithms for determinantal varieties and semidefinite programming

Simone Naldi

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
l'Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Présentée et soutenue le *24/09/2015* **par :**

Simone NALDI

Exact algorithms for determinantal varieties and semidefinite programming

---

JURY

Jean-Charles FAUGÈRE	DR INRIA Paris-Rocquencourt	Examinateur
Stéphane GAUBERT	DR INRIA Saclay-Île-de-France	Rapporteur
Didier HENRION	DR CNRS LAAS, Toulouse	Directeur de thèse
Bernard MOURRAIN	DR INRIA Sophia Antipolis	Examinateur
Giorgio OTTAVIANI	PU Université de Florence	Examinateur
Mohab SAFEY EL DIN	PU Un. P. et M. Curie, Paris	Directeur de thèse
Bruno SALVY	DR INRIA, ENS Lyon	Examinateur
Markus SCHWEIGHOFER	PU Université de Constance	Examinateur
Bernd STURMFELS	PU Un. de Californie, Berkeley	Rapporteur

École doctorale et spécialité :
EDSYS : Automatique et Informatique
Unité de Recherche :
Laboratoire d'Analyse et d'Architecture des Systèmes (UPR8001)
Directeur(s) de Thèse :
Didier HENRION et *Mohab SAFEY EL DIN*
Rapporteurs :
Stéphane GAUBERT et *Bernd STURMFELS*

Abstract

In this thesis we focus on the study of determinantal structures arising in semidefinite programming (SDP), the natural extension of linear programming to the cone of symmetric positive semidefinite matrices. While the approximation of a solution of a semidefinite program can be computed efficiently by interior-point algorithms, neither efficient exact algorithms for SDP are available, nor a complete understanding of its theoretical complexity has been achieved. In order to contribute to this central question in convex optimization, we design an exact algorithm for deciding the feasibility of a linear matrix inequality (LMI) $A(x) \succeq 0$. When the spectrahedron $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$ is not empty, the output of this algorithm is an algebraic representation of a finite set meeting $\mathscr{S}$ in at least one point x^*: in this case, the point x^* minimizes the rank of the pencil on the spectrahedron. The complexity is essentially quadratic in the degree of the output representation, which meets, experimentally, the algebraic degree of semidefinite programs associated to $A(x)$. This is a guarantee of optimality of this approach in the context of exact algorithms for LMI and SDP. Remarkably, the algorithm does not assume the presence of an interior point in the spectrahedron, and it takes advantage of the existence of low rank solutions of the LMI.

In order to reach this main goal, we develop a systematic approach to determinantal varieties associated to linear matrices. Indeed, we prove that deciding the feasibility of a LMI can be performed by computing a sample set of real solutions of determinantal polynomial systems. We solve this problem by designing an exact algorithm for computing at least one point in each real connected component of the locus of rank defects of a pencil $A(x)$. This algorithm admits as input generic linear matrices but takes also advantage of additional structures, and its complexity improves the state of the art in computational real algebraic geometry. Finally, the algorithms developed in this thesis are implemented in a new Maple library called Spectra, and results of experiments highlighting the complexity gain are provided.

Keywords

Semidefinite programming, linear matrix inequalities, exact algorithms, complexity, determinantal varieties, polynomial system solving, real root finding, semialgebraic optimization.

Qu'il est loin mon pays, qu'il est loin
Parfois au fond de moi se raniment
L'eau verte du canal du Midi
Et la brique rouge des Minimes

Ô mon pays, ô Toulouse, ô Toulouse

Toulouse, 1967
CLAUDE NOUGARO

Non domandarci la formula che mondi possa aprirti,
sì qualche storta sillaba e secca come un ramo.
Codesto solo oggi possiamo dirti,
ciò che non siamo, ciò che non vogliamo.

Ossi di seppia, 1925
EUGENIO MONTALE

Contents

Part I

Preliminaries

Introduction

Many problems in the applications boil down to computing a sample set of real solutions of a system of polynomial equations, or to deciding if a system of polynomial inequalities has a solution or not. Moreover, the input polynomials are typically characterized by a particular structure.

Numerical algorithms can often provide the floating point approximation of a solution, but they cannot in general certify that such a solution is unique, if this is the case, or provide guarantees of convergence. In particular, for some decision problems (such as deciding the emptiness of feasible sets in optimization) it is not always possible to obtain a decision criterion, that is an algorithm taking as input the polynomials defining a real algebraic or semi-algebraic set S, and returning in output the empty list if and only if S is empty, and otherwise one point in S. Moreover, frequently the set S is not finite and one has to extract a finite sample subset. Especially when dealing with algebraic structures, ubiquitous in many research areas, this is a hard computational challenge.

This has motivated, in the last decades, the growth of reciprocal exchanges between different disciplines and aspects of mathematics and computer science, such as real algebraic geometry, symbolic computation, convex and polynomial optimization. More specifically, in the context of optimization or systems control theory, many problems boil down to deciding whether the feasible set defined by a linear matrix inequality, called a spectrahedron, is empty or not. This set is convex and basic semi-algebraic, and the defining matrix is typically characterized by an additional structure.

Symbolic computation techniques have been developed in last decades to solve systems of polynomial equations via exact algorithms. The common feature of these algorithms is that the quantities appearing during the computation are not represented by floating point numbers but by polynomial equations with integer or rational coefficients. Typically, the goal is to reduce the hardness of the input problem by substituting it with a simpler low-dimensional one. Finally, it is extremely important to deeply exploit structures when designing dedicated techniques and algorithms for special subclasses of the general problem.

This thesis follows this path by addressing various problems in computational real algebraic geometry involving linear matrices and determinantal varieties, that are extremely interesting objects in many contexts, such as control theory, statistics, polynomial optimization or algebraic geometry. The global objective is to propose an approach via effective techniques to deal with real algebraic or semi-algebraic sets characterized by determinantal structure and arising especially in semidefinite optimization.

Goal of the thesis

The goal of this thesis, whose main results are summarized in the next section, is to develop different techniques for designing exact algorithms for solving systems of polynomial equations with determinantal structure over the reals. In particular, these algorithms will be used for solving exactly linear matrix inequalities.

Problem statements

We address two distinct problems. Given square matrices $A_0, A_1, \ldots, A_n$ of size m, with entries in $\mathbb{Q}$, the associated linear matrix or pencil is the matrix

$$A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n.$$

We define as follows the collection of low rank loci associated to the pencil. For $r \leq m-1$, let

$$\mathcal{D}_r = \{x \in \mathbb{C}^n \,:\, \operatorname{rank} A(x) \leq r\}.$$

The set $\mathcal{D}_r$ encodes any linear map from $\mathbb{C}^m$ to itself, whose associated matrix is described by the pencil $A(x)$ and has at least $m-r$ rank defects. It is the algebraic set defined by the minors of order $r+1$ of $A(x)$. It is called a determinantal variety.

Problem (A): Given $A_0, A_1, \ldots, A_n$ with rational coefficients, compute at least one point in each connected component of the real algebraic set $\mathcal{D}_r \cap \mathbb{R}^n$, or establish that this set is empty.

We will contextualize Problem (A) in the historical approaches based on symbolic computation for solving polynomial equations, in Section 1.2.2 of Chapter 1. Suppose, now, that matrices A_i are all symmetric, and let

$$\mathscr{S} = \{x \in \mathbb{R}^n \,:\, A(x) \succeq 0\}.$$

The set $\mathscr{S}$ is convex basic semi-algebraic, since it is defined by sign conditions on the coefficients of the characteristic polynomial of $A(x)$. Indeed, if

$$\det(A(x) + t\,\mathbb{I}_m) = f_m(x) + f_{m-1}(x)t + \cdots + f_1(x)t^{m-1} + t^m,$$

is the characteristic polynomial of $A(x)$, then $\mathscr{S} = \{x \in \mathbb{R}^n : f_i(x) \geq 0,\ i = 1, \ldots, m\}$. Remark that, with the previous definition, $f_m = \det A$ and, in general, $f_j(x)$ is the $j-$th elementary symmetric polynomial on the eigenvalues of A at x. Finally, $\mathscr{S}$ consists of real points x such that all eigenvalues of $A(x)$ are nonnegative, it is called a spectrahedron, and the optimization of a linear function on $\mathscr{S}$ is called a semidefinite program.

Problem (B): Given $A_0, A_1, \ldots, A_n$ symmetric, with rational coefficients, compute at least one point in the spectrahedron $\mathscr{S}$, or establish that this set is empty.

Similarly, Problem (B) is a particular instance of the emptiness problem for semi-algebraic sets (*cf.* Section 1.2.2). These two problems are intimately related since both the algebraic sets $\mathcal{D}_r$ and the semi-algebraic set $\mathscr{S}$ have a special hidden determinantal structure. Indeed, the closure of the Euclidean boundary of $\mathscr{S}$ in the Zariski topology is the hypersurface defined by the vanishing of the determinant of the generating linear pencil $A(x)$. Hence, a first step towards solving Problem (B) is to design a dedicated variant of Problem (A) when the matrix is symmetric. In this thesis, we prove that the geometry of spectrahedra implies that this step is also sufficient to compute one point in $\mathscr{S}$, and to solve the associated emptiness decision problem.

Motivations

Both Problem (A) and (B) are motivated by the necessity of obtaining certified results for linear matrix inequalities (LMI) and semidefinite programming (SDP). Indeed, the feasibility problem for SDP, that is deciding whether a spectrahedron is empty or not (Problem (B)), can be considered as a semidefinite program with constant objective function, and where the goal is to exhibit, if it exists, at least one minimizer. Semidefinite programming, as the natural extension of linear programming to the cone of symmetric positive semidefinite matrices

Further, deciding the feasibility of a LMI is a basic subroutine of numerical algorithms used in optimization or control theory. For example, computing the analytic center of an LMI $A(x) \succeq 0$, that is the point x^* in the interior of $\mathscr{S}$, that maximizes the multivariate polynomial function $x \mapsto \det A(x)$, requests in input a feasible point $x \in \mathscr{S}$. Such a point could be computed with LMI solvers by solving random generated semidefinite programs. Moreover, the general polynomial optimization problem admits a hierarchy of semidefinite relaxations that provide an approximation of the solution. Finally, deciding whether, for a given even-degree polynomial f, there exists a positive semidefinite Gram matrix, is related to the sum-of-squares decomposability of f, and hence to the problem of giving a certificate of nonnegativity. Hence, a purely symbolic approach able to certify the correctness of the feasible point $x^* \in \mathscr{S}$, or to establish that $\mathscr{S}$ is empty, or more deeply to solve a semidefinite program, would lead to obtain efficient hybrid approaches to a larger class of interesting problems.

Let us clarify that the intent of this thesis is not to prove that exact and numerical algorithms can be competitive in terms of admissible size of input problems; indeed, LMI solvers can nowadays treat inputs with a high number of variables that are out of reach for our algorithms. The expected contribution is to provide an exact proof for the feasibility of semidefinite programs, and to show that also exact algorithms can handle extremely degenerate examples that typically cause errors in numerical contexts. Here the word certificate means that the output of the algorithm solves a decision problem involving the formula $A(x) \succeq 0$: the algorithm is expected to return in output either the empty list, if and only if there are no solutions to this formula; otherwise, a rational parametrization of a finite set whose solutions meet the feasible set where $A(x)$ is positive semidefinite.

Similarly to the case of symmetric pencils, generic linear matrices and the related determinantal varieties model a large class of structures appearing in contexts such as statistics, combinatorics, polynomial optimization, or such as multilinear algebra and classical algebraic geometry. For example, the joint probability matrices for couples of discrete random variables have linear entries and the loci of rank defects correspond to mixtures of independence models; the Sylvester matrix of two univariate polynomials f, g is linear in the coefficients of f and g and its rank defect is the degree of the g.c.d. of f and g; in systems control theory, the Hurwitz stability locus for univariate complex polynomials is the determinantal semi-algebraic set defined by the leading principal minors of a given linear matrix. Solving Problem (A) efficiently and in exact arithmetic, exploiting further relations among the entries of the input linear matrix, is a first step for tackling associated questions arising in these contexts.

From a theoretical viewpoint, a motivation inspiring Problem (A) is, finally, also to compute complexity estimates for exact algorithms dealing with determinantal real algebraic sets. As detailed in the next section, for general inputs there exist different algorithms for solving this task, with a worst-case complexity which is singly exponential in the number of variables n.

Geometrical and algebraic properties of determinantal varieties actually motivate the development of a dedicated approach that can strongly exploit the hidden structure. Moreover, recent results [149, 44, 45, 120] show that structures in polynomial systems can be exploited for efficient Gröbner bases computations in the zero-dimensional case. Our goal in this context is to investigate complexity estimates in the real and positive-dimensional situations, designing algorithms taking into account and preserving the determinantal structure step-by-step.

State of the art

Problem (A) is a particular and structured instance of the question of how to compute a sample set of real solutions of a system of polynomial equations. In particular, if the set of complex solutions is non-empty and finite, solving Problem (A) means enumerating all real solutions.

The first algorithm addressing this question is due to Tarski [156], in the more general setting of quantifier elimination as decision problem of the first-order theory of the reals. The complexity of Tarski's algorithm is not elementary recursive, hence no finite towers of exponentials can express its value, as a function of input data parameters. The first finite bound for the complexity, at least doubly exponential in the number of variables, is due to Collins and to its Cylindrical Algebraic Decomposition algorithm [33]. Since the number of connected components of a real algebraic set defined by polynomials of degree at most d in n variables, is upper-bounded by the Thom-Milnor bound (*cf.* [20, Th. 7.23]) which is in $\mathcal{O}(d)^n$, further efforts had the main goal to fill this exponential gap between the output magnitude and the effective number of operations.

The first singly exponential algorithm (that is with complexity in $d^{\mathcal{O}(n)}$) is due to Grigoriev and Vorobjov [55] and is based on the critical points method. The idea is to compute algebraic subsets of the input set containing the local extrema of the restrictions of algebraic functions. For example, if the input set is non-empty and compact, computing the local minima of the Euclidean distance to a fixed point $x \in \mathbb{R}^n$, yields, generically, a finite set meeting each connected component. Moreover, when the input system is quadratic, it can be solved in polynomial time with respect to n, and exponentially with respect to the codimension of the associated complex algebraic set [56]. Further works of Renegar [133], Heintz, Roy and Solernó [65, 67], Basu, Pollack and Roy [20] have improved these algorithms for general real algebraic sets using the critical points method. The main idea mainly developed in the book [20] and in the aforementioned papers, is that computing critical points yields better complexity bounds that those obtained by Collins' algorithm (and by its improvements) for different classical questions in real algebraic geometry (quantifier elimination, sampling algebraic sets, deciding connectivity queries ...).

To control the degree of the intermediate sets produced in the course of such algorithms, one can prefer to use linear projections over generic subspaces, and hence to encode the critical points in the so-called polar varieties [11]. This idea has been developed in last years, mainly in the work of Bank, Giusti, Heintz, Mbakop, Pardo, Safey El Din and Schost. In [13, 12, 15, 14] the theory of polar varieties is developed and applied to the specific case of solving polynomial equations over the reals.

Since the constant hidden in $d^{\mathcal{O}(n)}$ is not fixed *a priori* by these algorithms, further results have tried to obtain uniform bounds on such exponents. In [137, 138] the authors showed how a dedicated variant of the critical point method allows to solve Problem (A) for smooth and equidimensional algebraic sets, under the assumption that the input polynomials generate a radical ideal: the complexity is essentially cubic in d^n. Further, specific techniques to deal with singular situations have also been developed [140], leading to complexities essentially quartic in d^n. Recent works have also reached better complexity results for the polynomial optimization problem over real algebraic sets, via the theory of polar varieties [58, 57].

Problem (B) is also a particular case of a more general question in semi-algebraic geometry, that of deciding the emptiness of semi-algebraic sets. Recall, indeed, that $\mathscr{S}$ is the semi-algebraic set defined by sign conditions on the m coefficients of the characteristic polynomial of the pencil. The algorithms in [66] or in [20, Ch. 13] compute a description of the connected components of the input semi-algebraic set, and run within $m^{\mathcal{O}(n)}$ arithmetic operations. Clearly, such algorithms do not exploit the particular structure of spectrahedra. In [91], the authors showed that deciding emptiness of $\mathscr{S}$ can be done in time $\mathcal{O}(m^{\min(n,m^2)})$, that is in polynomial time if either n or m is fixed. The main drawback of this algorithm is that it is based on Renegar's quantifier elimination [133], and hence it does not lead to efficient practical implementations.

To get a purely algebraic certificate of emptiness for $\mathscr{S}$, one could use the classical approach by Positivstellensatz ([102], *cf.* Section 1.3.2). This leads to a hierarchy of semidefinite programs. Indeed, by fixing an upper bound for the degrees of the sum-of-squares multipliers, the resulting problem is semidefinite in their unknown coefficients. Bounds for the degree of Positivstellensatz representations are exponential in the number of variables and have been computed in [147, 119]. In the recent remarkable result of Lombardi, Perrucci and Roy described in [106], a uniform $5-$fold exponential bound for the degree of the Hilbert 17th problem, which asks for similar certificates for nonnegative polynomials as sums of squares of rational functions. Klep and Schweighofer recently obtained an emptiness certificate dedicated to the spectrahedral case, by means of special quadratic modules associated to these sets [96]. It is shown there that deciding emptiness of $\mathscr{S}$ amounts to solving a sufficiently large SDP problem (whose size is exponential in either n or m), but for this latter task one has to use floating point implementations of interior-point algorithms.

Main results and organization

The main drawback of the general algorithms described in the previous section for computing sample subsets of the real solutions of polynomial equations or inequalities, is that their complexity and also their practical efficiency do not take advantage of possible structures of the input. Conversely, they typically break these structures. This thesis presents dedicated algorithms for determinantal real algebraic and semi-algebraic sets, deeply exploiting such a special structure and additional relations among the entries of the defining linear matrix.

Let us first present the data structures used as input and output of our algorithms. The input is given by vectors of rational numbers encoding the entries of a linear matrix $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$, and possibly integer numbers encoding the expected rank. The output will encode a finite set, represented as follows. Given a finite algebraic set $\mathcal{Z} \subset \mathbb{C}^n$, a rational parametrization representing $\mathcal{Z}$ is a vector $(q_0, q_1, \ldots, q_n, q_{n+1}) \subset \mathbb{Q}[t]$ such that:

- q_0 and q_{n+1} are coprime (*i.e.*, they do not share complex roots);
- $\mathcal{Z} = \left\{ \left(\frac{q_1(t)}{q_0(t)}, \ldots, \frac{q_n(t)}{q_0(t)} \right) \in \mathbb{C}^n : q_{n+1}(t) = 0 \right\}$;
- the degree of q_{n+1} equals the cardinality of $\mathcal{Z}$;

- $\mathcal{Z}$ is in one-to-one correspondence with the roots of q_{n+1}.

This representation is particularly useful since one can use it to compute a rational approximation of the coordinates of the solution by solving the univariate equation $q_{n+1}(t) = 0$. A given point in $\mathcal{Z}$ is represented by an isolating interval of rational numbers. The expected output for the exact algorithm solving Problem (A) will be a rational parametrization whose set of solutions meet each connected component of $\mathcal{D}_r$; similarly, for Problem (B) we will return a parametrization meeting the set $\mathscr{S}$.

In the next two sections we present the strategy for addressing Problems (A) and (B) and the main results of this thesis.

Strategy

Problem (A) asks to compute a sample set of real solutions of the polynomial system generated by the $(r+1) \times (r+1)$ minors of the linear matrix $A(x)$, hence a system of $\binom{m}{r+1}^2$ polynomials of degree at most $r+1$ in n variables. The associated algebraic set is typically singular, and for generic parameters these singularities are given by matrices with $m-r+1$ or more rank defects. Hence one could use the algorithm in [140], whose complexity is essentially quartic in $(r+1)^n$.

The first step of our strategy is to avoid the minor formulation modeling the determinantal algebraic set. Indeed, we introduce slack variables $y = (y_{i,j})_{1 \leq i \leq m, 1 \leq j \leq m-r}$ stored in a $m \times (m-r)$ matrix $Y(y)$, and we define the incidence variety

$$\mathcal{V}_r = \{(x, y) \in \mathbb{C}^n \times \mathbb{C}^{m(m-r)} : A(x)Y(y) = 0, UY(y) = S\},$$

for some full rank matrices $U \in \mathbb{M}_{m-r,m}(\mathbb{Q})$ and $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$. This technique is classically known as Room-Kempf desingularization for determinantal varieties (*cf.* for example [11]). The projection of $\mathcal{V}_r$ onto the $x-$space is included in $\mathcal{D}_r$ directly by definition.

Following this new model, the determinantal structure which appears directly in the minor formulation is now hidden into the equations $A(x)Y(y) = 0$, but the new algebraic set $\mathcal{V}_r$ gains two decisive properties:

- it is defined by bilinear equations;

- up to genericity assumptions on the input A, it is smooth and equidimensional.

The new regularity properties allow to design a dedicated critical point method to extract a finite set of sample points on the set $\mathcal{V}_r$. This is done by building algebraic sets encoding the critical points of the restriction to $\mathcal{V}_r$ of projections π on generic lines lying in the $n-$dimensional space containing the determinantal variety $\mathcal{D}_r$.

To ensure that this method is successful for computing sample points in each connected component $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$, we prove that when the linear projection $\pi \colon \mathbb{R}^n \to \mathbb{R}$ is generic enough, the image of each component $\mathcal{C}$ by π is closed. This fact enables to conclude the following dichotomy: given a connected component $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$, either one of the critical points of π is contained in $\mathcal{C}$, or π is surjective when restricted to $\mathcal{C}$ and $\pi(\mathcal{C})$ is the whole line. Remark that the genericity of the projection is equivalent to the composition of a generic change of variables $x \mapsto M\,x$ with the projection over the first variable x_1.

The previous reasoning allows to conclude that in order to compute at least one point per connected component of $\mathcal{D}_r$ it is sufficient to iterate this variant of the critical point method by progressively eliminating one variable x (for example instantiating x_1 to 0). Geometrically, this corresponds to the intersection of the determinantal variety $\mathcal{D}_r$ with the fiber $\pi^{-1}(0)$. Once that the first variable is eliminated, the critical points computation is applied on the fiber $\mathcal{D}_r \cap \mathbb{R}^n \cap \pi^{-1}(0)$. Remark, finally, that at each step both the determinantal structure of $\mathcal{D}_r$ and the bilinear structure and regularity of $\mathcal{V}_r$ are preserved, since the change of variables does not involve the slack variables y.

To design an exact algorithm dealing with Problem (B), we also take advantage of the determinantal structure of the input spectrahedron $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$. Indeed, every point x^* on the boundary of $\mathscr{S}$ generates a singular matrix $A(x^*)$ in the pencil $A(x)$. In this thesis we prove that, if $\mathscr{S}$ is not empty, and denoted by r the minimum rank attained by $A(x)$ on $\mathscr{S}$, then at least one of the connected components of $\mathcal{D}_r \cap \mathbb{R}^n$ is entirely contained in the boundary of $\mathscr{S}$, hence in $\mathscr{S}$.

The strategy for Problem (B) essentially relies on the previous geometrical result. Indeed, this implies that in order to decide the emptiness of $\mathscr{S}$, one has just to decide the emptiness of the associated low rank loci of the symmetric pencil $A(x)$. This can be done by performing a dedicated variant of the algorithm for Problem (A), when the linear matrix is symmetric. Moreover, a stronger result is provided in this thesis: when $\mathscr{S}$ is not empty, a smallest-rank point x^* is provided, that is with this procedure we are able to compute a solution x^* of a linear matrix inequality $A(x) \succeq 0$ where $A(x^*)$ has the smallest possible rank.

The previous fact is particularly remarkable since, for example, interior-point algorithms for semidefinite programming typically compute solutions with maximal possible rank (that is, points x^* lying in the relative interior of the optimal face) while computing low-rank solutions is considered a hard challenge.

Our method also takes into account possible additional structures arising in the linear pencil. As an example, we design a dedicated algorithm for solving the emptiness problem for Hankel spectrahedra, that is convex sets defined by a LMI $A(x) \succeq 0$ where A is generated by Hankel matrices. The interesting fact involving Hankel matrices is that their structure is inherited by their kernel, which can be defined by the column of matrices with a particular sparsity structure (*cf.* [64, Th. 5.1]). We recall here that the Hankel structure in linear matrices arises in different contexts, for example the catalecticant (or moment) matrix associated to a binary form is a Hankel matrix which is linear in the coefficients of the binary form.

Main theorems

We present here a concise summary of the main results. As announced in the previous section, to establish correctness of the algorithms for solving Problems (A) and (B), we need to prove that the images of all connected components of the real determinantal variety $\mathcal{D}_r \cap \mathbb{R}^n$ by a generic linear projection are closed sets.

To do that, given any algebraic set $\mathcal{Z} \subset \mathbb{C}^n$, of dimension d, we propose in Chapter 2, Section 2.1.2, a decomposition of $\mathcal{Z}$ of recursive type. We define a collection $\{\mathcal{O}_i(\mathcal{Z})\}_{1 \leq i \leq d}$ of algebraic subsets of $\mathcal{Z}$ by recursively considering union of equidimensional components of low dimension, singular loci and critical loci of linear projections. In particular, no regularity assumptions are requested for the input set $\mathcal{Z}$ to get this decomposition.

Denote by $M^{-1}\mathcal{Z} = \{x \in \mathbb{C}^n : M\,x \in \mathcal{Z}\}$ the image of $\mathcal{Z}$ via the linear action of a given $M \in \mathrm{GL}_n(\mathbb{Q})$. We prove in our first main result that the algebraic set $\mathcal{O}_i(\mathcal{Z})$ has dimension bounded by i. We finally prove closure properties of projection maps $\pi_i(x) = (x_1, \ldots, x_i)$ restricted to $\mathcal{Z} \cap \mathbb{R}^n$. We use the definition of Noether position, which will be recalled in Section 1.1.3. This theorem appeared in the paper [68], accepted for publication in Journal of Symbolic Computation.

Main Theorem 1. *Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d. There exists a non-empty Zariski open set $\mathscr{M} \subset \mathrm{GL}_n(\mathbb{C})$ such that, if $M \in \mathscr{M} \cap \mathbb{M}_{n,n}(\mathbb{Q})$, the following holds:*

1. *for $0 \leq i \leq d$, the set $\mathcal{O}_i(M^{-1}\mathcal{Z})$ has dimension $\leq i$ and is in Noether position*
2. *for any connected component $\mathcal{C} \subset \mathcal{Z} \cap \mathbb{R}^n$, for $0 \leq i \leq d$ and for $w \in \mathbb{R}^i$ on the boundary of $\pi_i(M^{-1}\mathcal{C})$, $\pi_i^{-1}(w) \cap M^{-1}\mathcal{C}$ is non-empty and contained in $\mathcal{O}_{i-1}(M^{-1}\mathcal{Z}) \cap M^{-1}\mathcal{C}$.*

Main Theorem 1 allows to deduce that via generic changes of variables one can retrieve expected behaviors, such as closure of projection maps.

We use this result to design, in Chapter 3, a probabilistic exact algorithm solving Problem (A). We prove that building the incidence variety $\mathcal{V}_r$ (as introduced in the previous section) and computing critical points of the restriction of generic projections to $\mathcal{V}_r$ allows to extract the requested sample subset of $\mathcal{D}_r$. The probabilistic feature of the algorithm is due to the random choice of linear changes of variables. These changes are sufficient to ensure that the genericity hypotheses of Main Theorem 1 above are satisfied, for $\mathcal{Z} = \mathcal{D}_r$, and hence to conclude closure properties of the projection maps.

The main result of Chapter 3 is the following one. This theorem appeared in the work [70], under review in Journal of Complexity. The case $r = m - 1$ is accepted for publication in Journal of Symbolic Computation [68].

Main Theorem 2. *Suppose that for* $0 \leq p \leq r$*, the incidence variety* $\mathcal{V}_p$ *is smooth and equidimensional and that its defining polynomial system generates a radical ideal. Suppose also that, for* $0 \leq p \leq r$*,* $\mathcal{D}_p$ *is empty or has the expected codimension* $(m-p)^2$*.*

There exists a probabilistic algorithm such that, with input (A, r)*, returns in output a rational parametrization* $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \subset \mathbb{Q}[t]$ *of degree in*

$$\mathcal{O}\left((n-(m-r)^2)\binom{n+m(m-r)}{n}^3\right)$$

within

$$\mathcal{O}\left((n+m^2-r^2)^7\binom{n+m(m-r)}{n}^6\right)$$

arithmetic operations over $\mathbb{Q}$*. The set defined by* q *contains at least one point in each connected component of the real algebraic set* $\mathcal{D}_r \cap \mathbb{R}^n$*. The set of input data for which the genericity assumptions are not satisfied is included in a proper algebraic subset of its parameter space; similarly, the set of parameters for which the algorithm is not correct lies in a proper algebraic subset of its parameter space.*

We finally address Problem (B) in Chapter 4. The goal is to obtain an algorithm detecting the emptiness of spectrahedra. The key idea originates from the following geometric fact that relates the loci of rank defects of symmetric pencils to the geometry of spectrahedra. When $\mathscr{S}$ is not empty, we denote by $r(A)$ the minimum rank attained by $A(x)$ on $\mathscr{S}$. The next two theorems are part of the paper [71], under review in SIAM Journal on Optimization.

Main Theorem 3. *Suppose that* $\mathscr{S} \neq \emptyset$*. Let* $\mathcal{C}$ *be a connected component of* $\mathcal{D}_{r(A)} \cap \mathbb{R}^n$ *such that* $\mathcal{C} \cap \mathscr{S} \neq \emptyset$*. Then* $\mathcal{C} \subset \mathscr{S}$*, and hence* $\mathcal{C} \subset (\mathcal{D}_{r(A)} \setminus \mathcal{D}_{r(A)-1}) \cap \mathbb{R}^n$*.*

As already detailed in the previous section, by Main Theorem 3 one can reduce the emptiness problem for spectrahedra to the problem of sampling real determinantal algebraic sets defined by symmetric pencils. The main results for Problem (B) follows. Remarkably, the complexity of the algorithm presented in Chapter 4 is strictly related to $r(A)$.

Main Theorem 4. *Suppose that for* $0 \leq r \leq m-1$*, the incidence variety* $\mathcal{V}_r$ *is smooth and equidimensional and that its defining polynomial system generates a radical ideal. There is a probabilistic algorithm that takes* A *as input and returns:*

1. *either the empty list, if and only if* $\mathscr{S} = \emptyset$*, or*
2. *the coordinates of a vector* x^* *such that* $A(x^*) = 0$*, if and only if the linear system* $A(x) = 0$ *has a solution, or*
3. *a rational parametrization* $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \in \mathbb{Q}[t]^{n+2}$ *such that there exists* $t^* \in \mathbb{R}$ *with* $q_{n+1}(t^*) = 0$ *and:*

- $A(q_1(t^*)/q_0(t^*), \ldots, q_n(t^*)/q_0(t^*)) \succeq 0$ *and*
- $\operatorname{rank} A(q_1(t^*)/q_0(t^*), \ldots, q_n(t^*)/q_0(t^*)) = r(A)$.

The number of arithmetic operations performed are in

$$\tilde{\mathcal{O}}\left(n \sum_{r \leq m-1} \binom{m}{r} (n + p_r + r(m-r))^7 \binom{p_r + n}{n}^6\right) \quad \text{if } \mathscr{S} \text{ is empty}$$
$$\tilde{\mathcal{O}}\left(n \sum_{r \leq r(A)} \binom{m}{r} (n + p_r + r(m-r))^7 \binom{p_r + n}{n}^6\right) \quad \text{if } \mathscr{S} \text{ is not empty,}$$

with $p_r = (m-r)(m+r+1)/2$. *If* $\mathscr{S}$ *is not empty, the degree of* q *is in*

$$\mathcal{O}\left(\binom{m}{r(A)}\binom{p_{r(A)} + n}{n}^3\right).$$

We finally present at the end of Chapter 4 a dedicated variant for the algorithm solving Problem (A) when the input linear matrix is Hankel, that is generated by Hankel matrices $A_0, A_1, \ldots, A_n$. This variant can be used for solving the emptiness problem for Hankel spectrahedra. The idea is to exploit the additional structure of Hankel matrices which induces relations among the slack variables that are added to define the incidence varieties. We present next the dedicated complexity result for Hankel linear matrices, which appeared in the Proceedings of ISSAC 2015 [69], Bath (UK).

Main Theorem 5. *Let* A *be a* $n-$*variate linear Hankel matrix of size* m, *and let* $r \leq m-1$. *Suppose that for* $0 \leq p \leq r$, *the incidence variety* $\mathcal{V}_p$ *is smooth and equidimensional and that its defining polynomial system generates a radical ideal. Suppose that, for* $0 \leq p \leq r$, $\mathcal{D}_p$ *is empty or has the expected dimension* $n - 2m + 2p + 1$.

There exists a probabilistic algorithm such that, with input (A, r), *returns in output a rational parametrization* $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \subset \mathbb{Q}[t]$ *of degree at most*

$$\binom{2m-r-1}{r} + \sum_{k=2m-2r}^{n} \sum_{p=0}^{r} d(m,k,p)$$

with $d(m,k,p) \in \mathcal{O}(\binom{k+2m-p-1}{k}^3)$, *within*

$$\mathcal{O}\left(\sum_{p=0}^{r} pn(2m-p)(pn(2m-p)(n+2m)^2 + (n+2m)^4)d(m,n,p)^2\right)$$

arithmetic operations over $\mathbb{Q}$. *The set defined by* q *contains at least one point in each connected component of the real algebraic set* $\mathcal{D}_r \cap \mathbb{R}^n$.

The gain in terms of complexity of these algorithms with respect to the state of the art in real algebraic geometry is explicit, for example, when the size of the matrix is fixed. In this case, the number of arithmetic operations is polynomial in the number of variables. This polynomial dependency does not appear in general algorithms and is special for the determinantal algebraic sets. We conclude that determinantal systems generated by rank defects in linear matrices are a new class of polynomial systems that can be solved in polynomial time when the size of the defining matrix is fixed.

The degree estimates are computed via multilinear Bézout bounds, since the polynomial systems encoding the critical points have a bilinear structure. We are not able to give an exact formula for the output degree, but we conjecture (*cf.* Conjecture 5.1 page 154) that they are given, for symmetric linear matrices, as the sum of algebraic degrees of intermediate semidefinite programs.

The complexity estimates are finally put into practice via several experiments. Chapter 5 contains results and discussions about the results. To test the performance on average data, we present timings for generic input linear matrices (the rationals are chosen randomly according to a given distribution). Also, different interesting examples of the literature on linear matrices are analyzed.

The tests are performed in MAPLE, via an new original library called SPECTRA. This library will be freely released in September 2015, and depends on the software FGB developed by J.-C. Faugère, for computations with Gröbner bases. We show with computer experiments two important features of our algorithms.

The first aspect is that it is able to run over a large set of input instances with remarkable size with respect to the typical admissible range of exact algorithms. Indeed, for dense linear matrices, it can handle matrices of size 4×4 with 11 variables, returning its output in less than 7 minutes, or matrices of size 5×5 with 9 variables and expected rank 2 in 15 minutes. The same examples cannot be handled by standard exact algorithms.

The second aspect is that we can give a practical solution to Problem (B) for extremely degenerate spectrahedra. We report on results of tests on a spectrahedron $\mathscr{S} = \{x = (x_1, \ldots, x_6) \in \mathbb{R}^6 : A(x) \succeq 0\}$ defined by the Gram matrix $A(x)$ of a nonnegative ternary quartic $f \in \mathbb{Q}[u_1, u_2, u_3]$ (a homogeneous polynomial of degree 4 in 3 variables) with rational coefficients. In other words, f is equal to $v'A(x)v$ for all $x \in \mathbb{R}^6$, where v is the vector of the monomials of degree 2 in u_1, u_2, u_3. The set $\mathscr{S}$ parametrizes the decompositions of f as a sum of squares. It turns out by a result of Scheiderer [142] that f is a sum of squares of polynomial with real coefficients but does not admit a similar decomposition with polynomials with rational coefficients. Applying our algorithm, we succeeded in finding a rational parametrization of two rank 2 points lying in the boundary of $\mathscr{S}$. These two points correspond to the unique 2 sums of squares decompositions of f as a sum of 2 squares.

Outline of the manuscript

The manuscript is organized as follows.

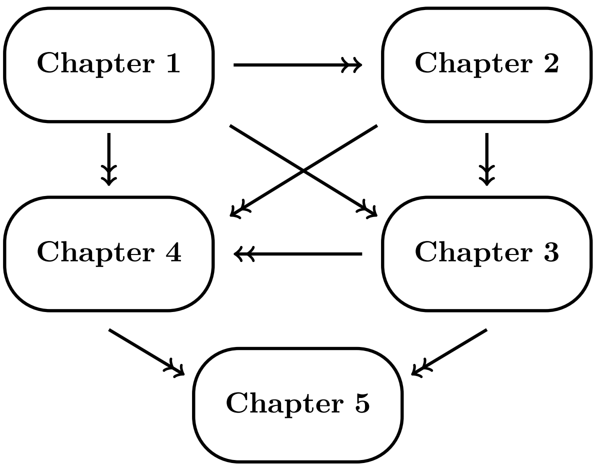
Chapter 1: It contains basic prerequisites of algebraic geometry, commutative algebra and polynomial systems solving. Moreover, we give an overview of the possible applications of semidefinite optimization. We also report on results in algebraic transversality theory (Sard's Lemma and Thom's Weak Transversality Theorem) that are deeply exploited for proving the aforementioned results.

Chapter 2: The goal of this chapter is to prove Main Theorem 1, which is done on page 64. To do that we propose a recursive decomposition of any algebraic set $\mathcal{Z} \subset \mathbb{C}^n$ of dimension d into a collection of subsets $\{\mathcal{O}_i(\mathcal{Z})\}_{1 \leq i \leq d}$. This allows to deduce closure properties of the restriction of the map $\pi_i(x) = (x_1, \ldots, x_i)$ to the real algebraic set $\mathcal{Z} \cap \mathbb{R}^n$, for $i = 1, \ldots, d$.

Chapter 3: We describe an algorithm for solving Problem (A) in exact arithmetic, that is we prove Main Theorem 2, (*cf.* page 73). In particular, we develop a local analysis of determinantal varieties and of the algebraic sets of critical points that are built during the procedure.

Chapter 4: We prove Main Theorem 4 (*cf.* page 112), yielding an algorithm for solving the emptiness problem for spectrahedra, that is for Problem (B). As already explained, the first step is addressing Main Theorem 3, whose proof is given on page 114.

Chapter 5: At the end of the manuscript we present numerical experiments of an implementation of the algorithm. We report on results of tests on randomly generated linear matrices (generic, symmetric or with Hankel or other additional structures). We also discuss some simple illustrative example and finally the interesting case of Scheiderer's spectrahedron.



The previous diagram pictures the dependencies of the different chapters of the manuscript. Indeed, Chapter 1 contains standard definitions to be understood as a background of the whole thesis. The results of Chapters 3 and 4 depend on those on Chapter 2, as already explained. A subroutine of the algorithm in Chapter 4 depends on a variant of that of Chapter 3, and we use the local description of determinantal varieties contained in this chapter. Finally, Chapter 5 reports on experiments for the algorithms in Chapters 3 and 4.

Perspectives

The techniques developed in this thesis give rise to a series of potential perspectives in convex algebraic geometry and optimization. We give below a brief description of some problems that can be addressed via exact approaches and for which the techniques developed in this thesis could be adapted.

What is the affine dimension of the input spectrahedron?

Given a spectrahedron $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$, the algorithm in [60] can decide whether $\mathscr{S} \cap \mathbb{Q}^n$ is empty or not, but *a priori* it cannot decide the emptiness of $\mathscr{S}$: actually, for degenerate but extremely interesting examples, it may happen that $\mathscr{S}$ is included in some hyperplane of $\mathbb{R}^n$ and it does not contain rational points.

The algorithm described in Chapter 4 of this manuscript can decide the emptiness of $\mathscr{S}$ and exhibit a sample point if one exists (*cf.* also [71]). Unfortunately, it cannot detect the full-dimensionality of $\mathscr{S}$, that is it is not able to decide whether its interior is empty or not. Hence, two questions arise:

- compute the affine dimension of $\mathscr{S}$;
- compute one sample point in its relative interior.

Algorithms for computing the affine dimension of semi-algebraic sets typically do not exploit the particular structure of these special semi-algebraic sets. The strong structure of spectrahedra understood as convex sets determined by determinantal inequalities plays a significant role for answering such questions.

How to decide the emptiness of spectrahedral shadows?

One of the results of this thesis implies that one can design efficient algorithms for deciding whether a spectrahedron is empty or not. It turns out that the class of spectrahedra is strictly included in the class of convex closed basic semi-algebraic sets, and that the former is not closed under linear projections.

Semidefinite representable sets (also called spectrahedral shadows or simply projections of spectrahedra) have consequently attracted a lot of attention during the last years. For example, it is currently conjectured (by Helton and Nie [72]) that every convex semi-algebraic set is semidefinite representable. Scheiderer has proved that the convex hull of any one-dimensional semi-algebraic subset of $\mathbb{R}^n$ is semidefinite representable [143], proving the Helton-Nie conjecture in dimension 2, however without an estimate of the number of lifting variables.

As an example, the (basic) closed convex planar semi-algebraic set

$$\{(x_1, x_2) \in \mathbb{R}^2 : 1 - x_1^4 - x_2^4 \geq 0\},$$

the *TV-screen set*, is not a spectrahedron but it is semidefinite representable (it has a representation as a projection of a $4-$dimensional spectrahedron). Spectrahedral shadows are described by positivity conditions on symmetric linear matrices subject to existential quantifiers, that is:

$$S = \left\{ x \in \mathbb{R}^n : \exists\, y \in \mathbb{R}^d, A(x, y) \succeq 0 \right\},$$

where $A(x, y) = A_0 + \sum x_i B_i + \sum y_j C_j$ is a symmetric linear matrix defined on variables $(x, y) = (x_1, \ldots, x_n, y_1, \ldots, y_d)$. The set S is the projection of the lifted spectrahedron

$$\mathscr{S} = \{(x, y) \in \mathbb{R}^{n+d} : A(x, y) \succeq 0\}$$

on the space of x variables. We can deduce that an algorithm that solves the emptiness decision problem for $\mathscr{S}$ also solves the same problem for S. In particular, one can compute sample points on S by computing sample points on $\mathscr{S}$ and by eliminating the slack variables.

Nevertheless, complexity obstacles arise by the following fact: the number d of slack variables $y \in \mathbb{R}^d$ and hence the size of the semidefinite representation can be superpolynomial in the dimension of S. Consequently, one deduces that the aforementioned algorithms can be inefficient when applied to the lifted sets $\mathscr{S}$. Since spectrahedral shadows have a nice determinantal description, one can hope to adapt the techniques described in this thesis, to more general situations. This research direction is also motivated by recent results about spectrahedral shadows, that link the boundary structure of S to the rank stratification of the defining matrix of $\mathscr{S}$, see *e.g.* [148].

How much does it cost to solve exactly a semidefinite program?

Exact algorithms for spectrahedral sets constitute the first step towards a more general symbolic approach to optimization on semi-algebraic sets, and in particular to semidefinite programming. Actually, the previous results, described above, strongly rely on optimization techniques adapted to real algebraic geometry such as the construction of polar varieties (varieties defining critical points). Hence one can strongly hope that similar techniques can be exploited to solve semidefinite programs in exact arithmetic. Let

$$\begin{aligned} \inf_{x \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} c_i x_i \\ s.t. \quad & A(x) \succeq 0, \end{aligned}$$

be a given semidefinite program, with exact data $c \in \mathbb{Q}^n$ and $A_i = A_i' \in \mathbb{Q}^{m \times m}$, and with $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$. Numerically, this semidefinite program can be solved in polynomial time, when the desired accuracy is fixed. This means that one can produce a numerical proof of feasibility and compute approximations of a solution of this program (*cf.* [9, Ch. 1, Sec. 4.1]). Indeed, there are numerical algorithms (implemented in various solvers) for computing efficiently approximations of a solution in floating-point arithmetic.

So far, exact algorithms for computing solutions to semidefinite programs, and a precise complexity analysis of semidefinite programming in the exact model, are missing. In the context of a possible symbolic approach that exploits the determinantal structure of spectrahedra, one could provide answers to the following questions:

- design an exact algorithm whose input is (c, A) and, if the infimum of $\sum_{i=1}^{n} c_i x_i$ on the feasible set is attained at x^*, whose output is a rational parametrization q whose solutions contain x^*;
- given such an algorithm, give an estimate of the degree of the rational parametrization q;
- how to compare such degrees with the algebraic degree [117] of the associated semidefinite program?
- what is the complexity of solving a semidefinite program in exact arithmetic?

In this thesis we design an efficient algorithm for the computation of minimal rank solutions of linear matrix inequalities. Indeed, if the spectrahedron $\mathscr{S}$ is not empty, the output is a rational parametrization whose solutions contain one point $x^* \in \mathscr{S}$ which minimizes the function $x \mapsto \operatorname{rank} A(x)$ over $\mathscr{S}$. Consequently, one could adapt the described techniques to compute minimal rank solutions to any semidefinite program. This is, in general, a hard problem, since typically interior-point algorithms returns in output points lying in the relative interior of the optimal face, where the rank is maximized (*cf.* [102, Lem. 1.4]).

A symbolic approach to nonnegative polynomials

The techniques described in this thesis can be useful also for related problems involving nonnegative polynomials and sums-of-squares representations. The importance of developing such theory is that approaches via sums of squares relaxations have been developed in the last decade to address general optimization problems over semi-algebraic sets (*cf.* [99, 102, 144, 94]).

Suppose, first, that $f \in \mathbb{Q}[y]$ is a polynomial of degree $2d$ which is a sum of squares of polynomials of degree d. One can write such a SOS decomposition via Gram matrices, that is if v is the vector of monomials in y of degree at most d, one writes $f = v'Gv$ for some positive semidefinite $G \succeq 0$. Factorizing, *e.g.*, $G = V'V$ where V is a $r \times m$ matrix (G is square of size m and of rank r) gives the requested SOS decomposition $f = (Vv)'(Vv)$. When G depends linearly on parameters $x = (x_1, \ldots, x_p)$, the minimum rank r attained on the spectrahedron $\{x \in \mathbb{R}^p : G(x) \succeq 0\}$ corresponds to minimal SOS decompositions. This spectrahedron parametrizes the SOS-decompositions of the given polynomial f, and is called the Gram spectrahedron of f (*cf.*, *e.g.*, [126, Sec. 6]).

So one can use such techniques to compute minimal SOS-decompositions of nonnegative polynomials in exact arithmetic. On the dual side, the same method can analogously be adapted to compute minimal sum of powers decompositions of nonnegative polynomials (*cf.* [134]), minimal sums of squares of matrix polynomials (*cf.* [95]) or extremal SOS-decompositions in the sense of Carathéodory (*cf.*, *e.g.*, [114]).

Résumé en français

De nombreux problèmes dans les applications se réduisent à calculer l'ensemble des solutions réelles d'un système d'équations algébriques, ou à décider si un système d'inégalités polynomiales a une solution ou non. En outre, les polynômes en entrée sont fréquemment caractérisés par une structure particulière.

Des algorithmes de nature numérique peuvent souvent fournir l'approximation en virgule flottante d'une solution, mais ils ne peuvent pas en général certifier qu'une telle solution est unique, si tel est le cas, ou donner une garantie de convergence ou d'exhaustivité (par exemple, l'obtention d'un point échantillon dans chaque composante connexe de l'ensemble des solutions). En particulier, pour certains problèmes de décision (par exemple, établir si l'ensemble des contraintes d'un problème d'optimisation est vide), il n'est pas toujours possible d'obtenir un critère de décision en s'appuyant sur du calcul numérique. En outre, fréquemment l'ensemble des solutions a dimension positive ce qui rend souvent le problème encore plus difficile.

Ceci a motivé, dans les dernières années, la croissance des échanges entre différents disciplines et aspects des mathématiques et de l'informatique théorique, comme la géométrie algébrique réelle, le calcul formel, l'optimisation convexe et polynomiale. Plus précisément, dans le cadre de l'optimisation ou la théorie du contrôle, de nombreux problèmes se résument à décider si l'ensemble défini par une inégalité matricielle linéaire, appelé spectraèdre, est vide. Cet ensemble est convexe et basique semi-algébrique, et la matrice linéaire est typiquement caractérisée par une structure supplémentaire.

Des techniques de calcul formel ont été développées dans les dernières décennies pour résoudre des systèmes d'équations polynomiales avec des algorithmes exacts. La caractéristique commune à ces algorithmes est que les données apparaissant pendant le calcul ne sont pas représentées par des nombres à virgule flottante, mais plutôt par des équations polynomiales à coefficients entiers ou rationnels. L'objectif est de réduire le problème de départ à un problème plus simple et en dimension petite. Par exemple, si l'ensemble des solutions réelles du système polynomial donné est de cardinalité infinie, on pourrait demander de calculer un ensemble fini de points intersectant chaque composante connexe sur l'ensemble algébrique réel associé. Enfin, il est extrêmement important d'exploiter les structures des objets en entrée, pour concevoir des techniques et des algorithmes dédiés.

Cette thèse suit cette voie, en abordant des problèmes en géométrie algébrique réelle concernant les variétés déterminantielles, qui sont des objets apparaissant dans de nombreux contextes, tels que la théorie du contrôle, les statistiques, l'optimisation polynomiale ou la géométrie algébrique. Le but principal est de proposer une approche par des techniques exactes pour résoudre ces types de problèmes apparaissant en particulier en optimisation semi-définie, c'est-à-dire la programmation linéaire sur le cône convexe des matrices semi-définies positives.

But de la thèse

L'objectif de cette thèse, dont les résultats principaux sont résumés dans la section suivante, est de développer différentes techniques et algorithmes exacts pour la résolution réelle des systèmes d'équations algébriques avec structure déterminantielle. En particulier, ces algorithmes sont utilisés pour résoudre les inégalités matricielles linéaires en arithmétique exacte.

Énoncé du problème

Nous nous intéressons à la résolution de deux problèmes. Étant données n matrices carrées $A_0, A_1, \ldots, A_n$ de taille $m \times m$, à coefficients dans $\mathbb{Q}$, on considère la matrice linéaire (ou faisceau) associée

$$A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n.$$

Nous allons définir la famille des ensembles algébriques des matrices du faisceau $A(x)$ de rang borné. Soit $r \leq m-1$, et soit

$$\mathcal{D}_r = \{x \in \mathbb{C}^n : \operatorname{rank} A(x) \leq r\}.$$

L'ensemble $\mathcal{D}_r$ contient les applications linéaires de $\mathbb{C}^m$ dans lui-même, dont la matrice associée appartient au faisceau $A(x)$ et a au moins $m-r$ chutes de rang. Il s'agit de l'ensemble algébrique défini par les mineurs de taille $r+1$ de $A(x)$. Cet ensemble est appelé varieté déterminantielle.

Problème (A) : Étant données $A_0, A_1, \ldots, A_n$ et $r \leq m$, calculer au moins un point par composante connexe de l'ensemble algébrique réel $\mathcal{D}_r \cap \mathbb{R}^n$, ou établir qu'il est vide.

On précisera le contexte dans le cadre des approches historiques en calcul formel pour la résolution des systèmes polynomiaux en Section 1.2.2 du Chapitre

1. Supposons, maintenant, que les matrices A_i soient symétriques, et soit

$$\mathscr{S} = \{x \in \mathbb{R}^n \,:\, A(x) \succeq 0\}\,.$$

L'ensemble $\mathscr{S}$ est convexe et basique semi-algébrique, puisqu'il est défini par des conditions de signe sur les coefficients du polynôme caractéristique de $A(x)$. En effet, soit

$$\det(A(x) + t\,\mathbb{I}_m) = f_m(x) + f_{m-1}(x)t + \cdots + f_1(x)t^{m-1} + t^m$$

le polynôme caractéristique de $A(x)$. On en déduit que $\mathscr{S} = \{x \in \mathbb{R}^n : f_i(x) \geq 0,\ i = 1, \ldots, m\}$. Remarquons ici qu'avec les notations introduites $f_m = \det A$ et, en général, $f_j(x)$ est le $j-$ème polynôme symétrique élémentaire evalué en les valeurs propres de A en x. Finalement, $\mathscr{S}$ contient tous les points réels $x \in \mathbb{R}^n$ tels que les valeurs propres de $A(x)$ soient positives ou nulles. L'ensemble $\mathscr{S}$ est un spectraèdre, et optimiser une fonction linéaire sur $\mathscr{S}$ est un problème d'optimisation semi-définie.

Problème (B) : Étant donnée $A_0, A_1, \ldots, A_n$ symétriques, calculer au moins un point dans $\mathscr{S}$, ou établir que cet ensemble est vide.

De la même manière, le Problème (B) est un cas particulier du problème de décider le vide d'un ensemble semi-algébrique (*cf.* Section 1.2.2). Ces deux problèmes sont intimement liés puisque les ensembles $\mathcal{D}_r$ et $\mathscr{S}$ ont une structure déterminantielle. De plus, la clôture (pour la topologie de Zariski) de la frontière Euclidienne de $\mathscr{S}$ est l'hypersurface définie par l'annulation du determinant de $A(x)$. Ainsi, une première étape nécessaire vers la résolution du Problème (B) est de résoudre une variante du Problème (A) lorsque la matrice est symétrique. Dans cette thèse, nous montrons que la géométrie des spectraèdres implique que cette étape est aussi suffisante pour calculer un point dans $\mathscr{S}$, et donc pour résoudre le problème de décision.

Motivation

Les Problèmes (A) et (B) sont motivés par la nécessité d'obtenir des résultats certifiés pour les inégalités matricielles linéaires (LMI) et la programmation semi-définie (SDP). En effet, le Problème (B) coïncide avec le problème de faisabilité pour la SDP, et peut être considéré lui-même comme un programme semi-défini, avec fonction objectif constante, et où le but est de retourner, s'il existe, au moins un minimiseur.

En outre, décider de la faisabilité d'une LMI est une sous-routine de base pour des algorithmes numériques utilisés en optimisation ou en théorie du contrôle. Par exemple, le calcul du centre analytique d'une LMI $A(x) \succeq 0$, c'est-à-dire le point $x^* \in \mathscr{S}$ qui maximise le polynôme multivarié $x \mapsto \det A(x)$, demande en entrée un point $x \in \mathscr{S}$. Un tel point pourrait être calculé avec des solvers LMI en résolvant des problèmes SDP avec fonctions objectifs générées aléatoirement.

De plus, le problème général de l'optimisation polynomiale

$$\begin{aligned} \inf_{x\in\mathbb{R}^n} \ & f(x) \\ s.t. \ & f_i(x) \geq 0, \, i = 1, \ldots, s \end{aligned}$$

(où f, f_i sont des polynômes en plusieurs variables) admet une hiérarchie de relaxations semi-définies qui peuvent fournir une approximation de la solution en virgule flottante. Enfin, décider si, pour un certain polynôme f, il existe une matrice de Gram semi-définie positive, est équivalent à construire des décompositions en sommes de carrées pour f, c'est-à-dire un certificat de positivité. Par conséquent, un algorithme exact pouvant certifier l'exactitude du point $x^* \in \mathscr{S}$, ou pouvant établir que $\mathscr{S}$ est vide, ou encore résoudre un programme semi-défini, pourrait conduire à des approches hybrides efficaces pour une grande classe de problèmes intéressants.

Nous précisons que l'objectif de cette thèse n'est pas de prouver que les algorithmes exacts sont compétitifs avec les algorithmes numériques ou peuvent traiter le même type de problèmes ; en effet, les solveurs LMI peuvent aujourd'hui traiter des entrées avec un nombre très élevé de variables, et les mêmes problèmes sont hors de portée des algorithmes exacts. La contribution principale de cette thèse est un algorithme en arithmétique exacte pour décider la faisabilité d'un programme semi-défini ; nous montrons que les algorithmes exacts peuvent gérer des exemples extrêmement dégénérés qui peuvent provoquer des erreurs dans des contextes numériques.

Comme dans le cas symétrique, les matrices linéaires génériques (sans structure particulière) et leurs variétés déterminantielles modélisent une grande classe d'exemples dans des contextes tels que les statistiques, la combinatoire, l'optimisation polynomiale, ou tels que l'algèbre multilinéaire et la géométrie algébrique classique. Par exemple, les matrices de probabilité conjointe pour un couple de variables aléatoires discrètes ont des entrées linéaires et les lieux de chute de rang correspondent à des mélanges de modèles de probabilité ; la matrice de Sylvester de deux polynômes univariés f, g est linéaire en les coefficients de f et g et sa chute de rang correspond au degré du p.g.c.d. de f et g ; en théorie du contrôle, le lieu de stabilité pour les polynômes complexes univariés est l'ensemble semi-algébrique défini par les mineurs principaux d'une matrice linéaire donnée. Résoudre le Problème (A) efficacement et en précision infinie (c'est-à-dire, en arithmétique exacte), en exploitant la structure des données en entrée, est une première étape pour aborder ces questions plus profondes.

D'un point de vue purement calculatoire, une question importante pour le Problème (A) est également d'établir des bornes de complexité théorique issues de l'exploitation des structures déterminantielles. Comme détaillé dans la section suivante, il existe plusieurs algorithmes pour résoudre cette tâche, avec une complexité, dans le cas pire, qui est simplement exponentielle en le nombre de variables n. Les propriétés géométriques et algébriques des variétés déterminantielles motivent le développement d'une approche dédiée. En outre, des résultats récents [149, 44, 45, 120] montrent que ces structures peuvent être exploitées pour le calcul rapide d'une base de Gröbner dans le cas de la dimension zéro. Notre objectif dans ce contexte est d'obtenir des bornes de complexité dans le cas réel et de dimension positive.

État de l'art

Le Problème (A) est un cas particulier et structuré du problème consistant à trouver des points témoins dans chaque composante connexe d'un ensemble réel défini par des equations algébriques. En particulier, si l'ensemble des solutions complexes est non vide et fini, résoudre le Problème (A) signifie énumérer toutes les solutions réelles.

Le premier algorithme répondant à cette question est dû à Tarski [156], dans le cadre plus général de l'élimination des quantificateurs sur les réels comme un problème de décision de la théorie du premier ordre des nombres réels. La complexité de l'algorithme de Tarski n'est pas élémentairement récursive, donc aucune tour finie d'exponentielles (en fonction de la taille des données en entrée) ne peut la borner. La première complexité finie est celle de l'algorithme de la Décomposition Cylindrique Algébrique (CAD) de Collins, de type récursif sur les variables. Sa complexité est doublement exponentielle en le nombre de variables [33]. Étant donné que le nombre de composantes connexes d'un ensemble algébrique réel défini par des polynômes de degré au plus d en n variables est borné par la borne de Thom-Milnor (*cf.* [20, Th. 7,23]) qui est en $\mathcal{O}(d)^n$, des efforts supplémentaires ont amené à eliminer cette différence exponentielle entre la taille attendue de la sortie et le nombre effectif d'opérations.

Le premier algorithme simplement exponentiel (c'est-à-dire de complexité en $d^{\mathcal{O}(n)}$) est dû à Grigoriev et Vorobjov [55], basé sur la méthode des points critiques. L'idée de base est de calculer des sous-ensembles algébriques du lieu-solution défini par l'entrée, contenant les minima/maxima locaux des restrictions de certaines fonctions algébriques. Par exemple, si l'ensemble d'entrée est non vide et compact, le calcul des minima locaux de la distance Euclidienne d'un point donné $x \in \mathbb{R}^n$ engendre un ensemble fini qui intersecte chaque composante connexe. En outre, lorsque le système en entrée est quadratique, la complexité de la méthode est polynomiale en n, et exponentielle par rapport à la co-dimension de l'ensemble algébrique complexe associé [56]. D'autres travaux successifs de Renegar [133], Heintz, Roy et Solernó [65, 67], Basu, Pollack et Roy [20] ont amélioré ces algorithmes pour des ensembles algébriques réels généraux, en utilisant la méthode des points critiques. L'idée principale (développée dans le livre [20] et dans les papiers mentionnés ci-dessus) est que le calcul des points critiques donne une meilleure complexité que celle obtenue par l'algorithme de Collins (et par ses améliorations) pour différentes questions classiques de la géométrie algébrique réelle (élimination des quantificateurs, échantillonnage d'ensembles algébriques, problèmes de connectivité ...). En effet, il permet de se ramener à l'étude d'un ensemble algébrique complexe fini sans traitement récursif sur les variables.

Pour mieux contrôler le degré des ensembles apparaissant au cours de ces algorithmes, et donc leurs complexités, on peut préférer utiliser des projections linéaires sur des sous-espaces génériques, et donc représenter les points critiques dans des ensembles algébriques appelés variétés polaires [11]. Cette idée a été développée ces dernières années, principalement dans les travaux de Bank, Giusti, Heintz, Mbakop, Pardo, Safey El Din et Schost. Dans [13, 12, 15, 14], la théorie des variétés polaires est développée et appliquée au cas spécifique de la résolution d'équations polynomiales sur les réels.

Puisque la constante cachée dans l'exposant de $d^{\mathcal{O}(n)}$ n'est pas explicitée par les algorithmes précurseurs de la méthode des points critiques, d'autres travaux ont essayé d'obtenir des bornes uniformes sur ces exposants. Dans [137] les auteurs ont montré comment une variante dédiée de la méthode des points critiques permet de résoudre le Problème (A) pour des ensembles algébriques lisses et équidimensionnels, sous l'hypothèse que les polynômes d'entrée engendrent un idéal radical : la complexité est essentiellement cubique en d^n. En outre, des techniques spécifiques pour le cas singulier ont également été développées [140], conduisant à une complexité essentiellement en $\mathcal{O}(d^{4n})$. Des travaux récents ont également obtenu des résultats similaires de complexité pour le problème d'optimisation polynomiale sur des ensembles algébriques réels, en s'appuyant sur la théorie des variétés polaires [58, 57].

Le Problème (B) est également un cas particulier d'une question plus générale en géométrie semi-algébrique, celle de décider la vacuité d'un ensemble semi-algébrique. Nous rappelons à ce propos que $\mathscr{S}$ est l'ensemble semi-algébrique défini par des conditions de signe sur les m coefficients du polynôme caractéristique de A. Les algorithmes dans [66] ou [20, Ch. 13] calculent une description des composantes connexes de l'ensemble semi-algébrique en entrée, avec $m^{\mathcal{O}(n)}$ opérations arithmétiques élémentaires. Évidemment, ces algorithmes n'exploitent pas la structure particulière des spectraèdres. Dans [91], les auteurs ont montré que le Problème (B) peut être résolu en temps $\mathcal{O}(m^{\min(n,m^2)})$, c'est-à-dire en temps polynomial si n ou m est fixé. L'inconvénient principal de cet algorithme est qu'il est basé sur l'élimination des quantificateurs de Renegar, et donc il ne conduit pas à des implantations efficaces en pratique, la constante de complexité située en exposant étant élevée.

Pour obtenir un certificat purement algébrique du vide pour $\mathscr{S}$, on pourrait utiliser l'approche classique par Positivstellensatz ([102], *cf.* Section 1.3.2). Cela conduit à une hiérarchie de programmes semi-définis. En effet, en fixant les degrés de multiplicateurs de sommes de carrés dans ces certificats, le problème associé devient un problème de programmation semi-définie. les bornes pour le degré de représentations Positivstellensatz sont, en géneral, exponentielles en le nombre de variables et ont été calculées dans [147, 119]. Le résultat récent de Lombardi, Perrucci et Roy décrit dans [106] montre une borne uniforme quintuplement exponentielle pour le degré du 17-ème problème de Hilbert, qui demande des certificats similaires pour les polynômes non négatifs comme sommes de carrés de fonctions rationnelles. Klep et Schweighofer ont récemment obtenu un certificat de vide dédié au cas spectraèdral, au moyen de certains modules quadratiques spéciaux associés à $\mathscr{S}$ [96]. Ils montrent que décider le vide de $\mathscr{S}$ est equivalent à résoudre un problème SDP d'une certaine taille (exponentielle en n ou m), mais pour cette dernière tâche, on doit utiliser des implantations en virgule flottante des algorithmes de points intérieurs.

Stratégie

Le principal inconvénient des algorithmes généraux décrits dans la section précédente, pour calculer des points témoins dans les composantes connexes d'un ensemble algébrique ou semi-algébrique, est que leur complexité et aussi leur efficacité en pratique n'exploitent pas les possibles structures qui caractérisent les données en entrée. Au contraire, ils cassent généralement ces structures.

Nous présentons d'abord les structures des données utilisées en entrée de nos algorithmes. L'entrée est représentée par un vecteur de nombres rationnels contenant les coefficients d'une matrice linéaire $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$, et éventuellement un entier r. La sortie est un ensemble fini, représentée de la façon suivante. Étant donné un ensemble algébrique fini $\mathcal{Z} \subset \mathbb{C}^n$, un paramétrage rationnel représentant $\mathcal{Z}$ est un vecteur $(q_0, q_1, \ldots, q_n, q_{n+1}) \subset \mathbb{Q}[t]$ tel que :

- q_0 et q_{n+1} sont premiers entre eux (*i.e.*, ils n'ont pas de racines en commun) ;
- $\mathcal{Z} = \left\{ \left(\frac{q_1(t)}{q_0(t)}, \ldots, \frac{q_n(t)}{q_0(t)} \right) \in \mathbb{C}^n : q_{n+1}(t) = 0 \right\}$;
- le degré de q_{n+1} est égal à la cardinalité de $\mathcal{Z}$;
- $\mathcal{Z}$ est en bijection avec l'ensemble des racines de q_{n+1}.

Cette représentation est particulièrement utile puisqu'on peut l'utiliser pour calculer une approximation rationnelle des coordonnées de la solution en résolvant l'équation univariée $q_{n+1}(t) = 0$. Chaque point $x \in \mathcal{Z}$ est représenté par q_{n+1} et par un intervalle de nombres rationnels.

Le Problème (A) revient à calculer un ensemble fini de solutions réelles du système polynomiale constitué par les mineurs de taille $r+1$ de la matrice linéaire $A(x)$, donc un système de $\binom{m}{r+1}^2$ polynômes de degré au plus $r+1$ en n variables. L'ensemble algébrique associé est génériquement singulier, et ces singularités sont (génériquement) définies par des chutes de rang plus élevées. Par conséquent, on pourrait utiliser l'algorithme décrit dans [140], dont la complexité est essentiellement quartique en $(r+1)^n$.

La première étape de notre stratégie est d'éviter la formulation par mineurs pour modéliser la varieté déterminantielle. Nous introduisons de nouvelles variables

$$y = (y_{i,j})_{1 \leq i \leq m, 1 \leq j \leq m-r}$$

stockées dans une matrice $Y(y)$ de taille $m \times (m-r)$, et nous définissons la variété d'incidence

$$\mathcal{V}_r = \{(x, y) \in \mathbb{C}^n \times \mathbb{C}^{m(m-r)} : A(x)Y(y) = 0, UY(y) = S\},$$

pour certains matrices de rang plein $U \in \mathbb{M}_{m-r,m}(\mathbb{Q})$ et $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$. Cette technique est classiquement connue comme désingularisation *à la Room-Kempf* pour les variétés déterminantielles (*cf.* par exemple [11]). La projection de $\mathcal{V}_r$ sur l'espace des x est contenue dans $\mathcal{D}_r$, par définition.

Suite à cette réécriture, la structure déterminantielle qui apparaît directement dans la formulation par mineurs est maintenant cachée dans les équations $A(x)Y(y) = 0$. Le nouvel ensemble algébrique $\mathcal{V}_r$ gagne deux propriétés décisives :

- il est défini par des équations bilinéaires ;
- avec des hypothèses de généricité sur la matrice A, il est lisse et équidimensionnel.

Les nouvelles propriétés de régularité permettent de concevoir une méthode de points critiques dédiée. Le but est d'extraire un ensemble fini de points sur l'ensemble $\mathcal{V}_r$. Cela se fait par la construction d'ensembles algébriques contenant les points critiques de la restriction à $\mathcal{V}_r$ de projections π sur des droites génériques de l'espace $n-$dimensionel contenant l'ensemble $\mathcal{D}_r$.

Pour que cette réduction permette de calculer un point dans chacque composante connexe $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$, nous prouvons que lorsque la projection linéaire $\pi \colon \mathbb{R}^n \to \mathbb{R}$ est générique, l'image de chaque composante $\mathcal{C}$ par π est un fermé Euclidien. Cela permet de conclure le fait suivant : étant donnée une composante connexe $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$, soit l'un des points critiques de π est contenu dans $\mathcal{C}$, soit π est surjective si restreinte à $\mathcal{C}$ et $\pi(\mathcal{C}) = \mathbb{R}$. Remarquons que la généricité de la projection est équivalente à la composition d'un changement générique de variables $x \mapsto M\,x$ avec la projection sur la première variable x_1.

Le raisonnement précédent permet de conclure que, pour calculer au moins un point par composante connexe de $\mathcal{D}_r$, il suffit d'appliquer cette variante de la méthode des points critiques en éliminant, récursivement, une variable par étape (par exemple, en instanciant x_1 à 0). Géométriquement, cela correspond à l'intersection de la variété déterminantale $\mathcal{D}_r$ avec la fibre $\pi^{-1}(0)$ de π en 0. Une fois que la première variable est éliminée, le calcul des points critiques est appliqué sur la fibre $\mathcal{D}_r \cap \mathbb{R}^n \cap \pi^{-1}(0)$. Nous remarquons enfin, que, à chaque étape, soit la structure déterminantale de $\mathcal{D}_r$, soit la structure bilinéaire et la régularité de $\mathcal{V}_r$, sont conservées.

Pour concevoir un algorithme exact pour le Problème (B), nous profitons également de la structure déterminantielle du spectraèdre $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$. En effet, chaque point x^* sur la frontière de $\mathscr{S}$ correspond à une matrice singulière $A(x^*)$ dans $A(x)$. Dans cette thèse, nous montrons que, si $\mathscr{S}$ n'est pas vide, et en notant r le rang minimum atteint par $A(x)$ sur $\mathscr{S}$, alors au moins une des composants connexes de $\mathcal{D}_r \cap \mathbb{R}^n$ est entièrement contenue dans la frontière de $\mathscr{S}$, donc dans $\mathscr{S}$.

La stratégie du problème (B) repose essentiellement sur ce résultat géométrique, dont on donnera une preuve explicite dans le Chapitre 4. En effet, cela implique que, pour décider de la vacuité de $\mathscr{S}$, il suffit de décider de la vacuité des variétés déterminantielles associées à $A(x)$. Cela peut être effectué avec une variante de l'algorithme dédié au Problème (A), lorsque la matrice linéaire est symétrique. En outre, un résultat plus fort est contenu dans cette thèse : si $\mathscr{S}$ n'est pas vide, nous sommes en mesure de calculer une solution x^* de l'inegalité matricielle linéaire $A(x) \succeq 0$ où $A(x^*)$ minimise le rang parmi les points dans $\mathscr{S}$. Le fait précédent est particulièrement remarquable puisque, par exemple, les algorithmes de points intérieurs pour la programmation semi-définie calculent généralement des solutions avec rang maximal (ce qui correspond à des points dans l'intérieur relatif de la face optimale), alors que le calcul des solutions de rang faible est considéré comme un problème difficile et important.

La méthode développée prend également en compte de possibles structures additionnelles de la matrice linéaire $A(x)$. A titre d'exemple, nous considérons le cas des matrices linéaires Hankel, et les spectraèdres associés. Le fait intéressant est que la structure Hankel de $A(x)$ induit une structure sur leur noyau (*cf.* [64, Th. 5.1]), du coup sur les polynômes qui définissent la variété d'incidence. Les matrices linéaires avec structure Hankel apparaissent dans des contextes différents : par exemple, la matrice catalectique (ou matrice des moments) associée à une forme binaire est une matrice Hankel qui est linéaire en les coefficients de la forme binaire.

Perspectives

Les techniques développées dans cette thèse donnent lieu à une série de perspectives potentielles en géométrie algébrique convexe et en optimisation. Nous donnons ci-dessous une brève description de certaines de ces perspectives.

Calcul de la dimension d'un spectraèdre

Étant donné un spectraèdre $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$, l'algorithme dans [60] peut décider si $\mathscr{S} \cap \mathbb{Q}^n$ est vide ou pas, mais *a priori* il n'est pas capable de décider si $\mathscr{S}$ est vide : en effet, dans des exemples particulièrement dégénérés mais très intéressants dans la littérature, il peut arriver que $\mathscr{S}$ soit inclus dans un certain hyperplan de $\mathbb{R}^n$ et qu'il ne contienne pas de points à coordonnées rationnelles.

L'algorithme décrit dans le chapitre 4 de ce manuscrit peut décider la vacuité de $\mathscr{S}$ et calculer un point témoin dans $\mathscr{S}$ s'il existe (*cf.* également [71]). Malheureusement, il ne peut pas détecter si l'interieur de $\mathscr{S}$ est vide ou non. Par conséquent, deux questions se posent :

- calculer la dimension affine de $\mathscr{S}$ (c'est-à-dire, la dimension du plus petit espace affine contenant $\mathscr{S}$) ;
- calculer un point dans son intérieur relatif.

Des algorithmes pour le calcul de la dimension affine d'ensembles semi-algébriques génériques existent mais ils ont une complexité exponentielle en le nombre des variables. La forte structure des spectraèdres comme ensembles convexes définis par des inégalités déterminantielles joue un rôle important pour répondre à ces questions.

Comment décider la vacuité des projections des spectraèdres ?

La classe des spectraèdres est strictement contenue dans la classe des ensembles basiques semi-algébriques convexes fermés, mais l'image par une projection linéaire d'un spectraèdre n'est pas nécessairement un spectraèdre. Les ensembles représentables comme des projections de spectraèdres (ou SDP-représentables) ont par conséquent attiré beaucoup d'attention pendant les dernières années. Par exemple, il est actuellement conjecturé (par Helton et Nie [72]) que tout ensemble semi-algébrique convexe est SDP-représentable. Scheiderer a prouvé que si $\mathcal{C} \subset \mathbb{R}^n$ est un ensemble semi-algébrique de dimension 1, l'enveloppe convexe de $\mathcal{C}$ est SDP-représentable [143], en deduisant la conjecture de Helton et Nie en dimension 2 (cependant sans estimation du nombre de variables d'écart nécessaire pour la représentation).

A titre d'exemple, l'ensemble semi-algébrique fermé et convexe du plan

$$\{(x_1, x_2) \in \mathbb{R}^2 : 1 - x_1^4 - x_2^4 \geq 0\},$$

n'est pas un spectraèdre mais il est SDP-représentable (il a une représentation comme projection d'un spectraèdre de dimension 4). Les ensembles SDP-représentables sont décrits par une condition de positivité sur une matrice linéaire symétrique soumise à des quantificateurs existentiels, c'est-à-dire :

$$S = \left\{x \in \mathbb{R}^n : \exists\, y \in \mathbb{R}^d, A(x,y) \succeq 0\right\},$$

où $A(x,y) = A_0 + \sum x_i B_i + \sum y_j C_j$ est une matrice symétrique linéaire en les deux groupes de variables $(x,y) = (x_1, \ldots, x_n, y_1, \ldots, y_d)$. L'ensemble S est la projection du spectraèdre

$$\mathscr{S} = \{(x,y) \in \mathbb{R}^{n+d} : A(x,y) \succeq 0\}$$

sur l'espace des variables x. Nous pouvons en déduire qu'un algorithme qui résout le problème du vide pour $\mathscr{S}$ résout également le même problème pour S. En particulier, on peut calculer des points témoins dans S en calculant des points témoins dans $\mathscr{S}$ et en éliminant les variables d'écart.

Néanmoins, des obstacles de complexité surviennent car le nombre d de variables $y \in \mathbb{R}^d$ et donc la taille de la représentation peut être *a priori* superpolynomiale en la dimension de S. Par conséquent, les algorithmes ci-dessus peuvent être inefficaces lorsqu'ils sont appliqués à $\mathscr{S}$ (le nombre de variables étant très élevé). Puisque les ensembles SDP-représentables ont une description de nature déterminantielle, on peut espérer adapter les techniques décrites dans cette thèse à des situations plus générales. Cette perspective est aussi motivée par des résultats récents (voir *e.g.* [148]) qui s'intéressent à la structure de la frontière algébrique de S et son lien avec la stratification du rang de la matrice définissant $\mathscr{S}$.

Quelle est la complexité de la programmation semi-définie ?

Les algorithmes exacts pour les spectraèdres constituent la première étape vers une approche symbolique plus générale à l'optimisation semi-algébrique, et en particulier à la programmation semi-définie. En fait, les résultats précédemment décrits s'appuient fortement sur des techniques classiques en optimisation, adaptées à la géométrie algébrique réelle, tels que la construction des variétés polaires (variétés définissant des points critiques de projections linéaires). Ainsi, on peut fortement espérer que des techniques similaires peuvent être exploitées pour résoudre des programmes semi-définis en arithmétique exacte. Soit

$$\begin{aligned} \inf_{x\in\mathbb{R}^n} \quad & \sum_{i=1}^{n} c_i x_i \\ s.t. \quad & A(x) \succeq 0, \end{aligned}$$

un programme semi-défini donné, où $c \in \mathbb{Q}^n$ et $A_i = A_i' \in \mathbb{Q}^{m\times m}$, et $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$. En arithmétique approchée, ce programme semi-défini peut être résolu en temps polynomial, lorsque la précision des calculs effectués est fixée, avec l'algorithme de l'ellipsoïde (*cf.* [59] ou [9, Ch. 1, Sec. 4.1]). En effet, il existe des algorithmes de type point interieur, implanté dans divers solveurs, pour calculer efficacement des approximations d'une solution en virgule flottante.

Jusqu'à ce moment, il n'existe pas d'algorithme exact efficacement implanté pour la résolution d'un programme semi-défini. Dans le contexte d'une possible approche symbolique qui exploite la structure déterminantielle des spectraèdres, on pourrait donner des réponses aux questions suivantes :

- comment obtenir un algorithme exact dont l'entrée est (c, A) et, si l'infimum de $\sum_{i=1}^{n} c_i x_i$ sur $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$ est atteint en x^*, dont la sortie est un paramétrage rationnel q dont les solutions contiennent x^* ?
- étant donné un tel algorithme, quel est le degré du paramétrage rationnel q dans le cas pire ?
- comment comparer ces degrés avec le degré algébrique [117] de ce programme semi-défini ?
- quelle est la complexité de la résolution d'un programme semi-défini en arithmétique exacte ?

Dans cette thèse, nous concevons un algorithme efficace pour le calcul de solutions de rang minimal d'inégalités matricielles linéaires. En effet, si le spectraèdre $\mathscr{S}$ n'est pas vide, la sortie est un paramétrage rationnel dont les solutions contiennent un point $x^* \in \mathscr{S}$ qui minimise la fonction $x \mapsto \operatorname{rang} A(x)$ sur $\mathscr{S}$. Par conséquent, on pourrait adapter ces techniques pour calculer des solutions de rang minimal au problème précédent. Ceci est, en général, un problème difficile, puisque les algorithmes de point intérieur calculent des solutions dans l'intérieur relatif de la face optimale, où le rang est maximisé (*cf.* [102, Lem. 1.4]).

Une approche exact aux polynômes non négatifs

L'importance du développement de la théorie des polynomes positifs et leurs decompositions en somme de carrées (SOS), est que les approches par les sommes des carrés ont été utilisées dans la dernière décennie pour résoudre le problème général de l'optimisation polynomiale (*cf.* [99, 102, 144, 94]).

Supposons qu'un polynôme $f \in \mathbb{Q}[y_1, \ldots, y_n]$ de degré $2d$ puisse s'exprimer comme somme de carrés de polynômes de degré d. On peut écrire une telle décomposition SOS via la matrice de Gram de f : si v est le vecteur des monômes en $y = (y_1, \ldots, y_n)$ de degré au plus d, on écrit $f = v'Gv$ pour une certaine matrice $G \succeq 0$. En factorisant, *e.g.*, $G = V'V$ où V est une matrice $r \times m$ (G est carrée de taille m et de rang r) on produit la décomposition SOS recherchée $f = (Vv)'(Vv)$. Lorsque G dépend linéairement de paramètres $x = (x_1, \ldots, x_p)$, le rang minimum r atteint sur le spectraèdre $\{x \in \mathbb{R}^p : G(x) \succeq 0\}$ correspond à la longueur minimale d'une décomposition SOS de f. Aussi, ce spectraèdre paramétrise toutes les décompositions de f, et est appelé le spectraèdre de Gram de f (*cf.*, *g.e.*, [126, Sec. 6]).

Donc, on peut utiliser de telles techniques pour calculer des décompositions SOS minimales de polynômes non négatifs en arithmétique exacte. Par dualité, le même algorithme peut être adapté pour calculer des décompositions en sommes de puissances de formes linéaires (*cf.* [134]), ou des sommes de carrés de matrices polynomiales (*cf.* [95]).

Chapter 1

Prerequisites

This is a bibliographical chapter and does not contain original contributions. We first provide in Section 1.1 basic definitions of commutative algebra and algebraic geometry. Then we report on classical results in algebraic transversality theory and on the relation between Noether position and properness properties of linear projections, discussing how these results can be used in effective real algebraic geometry. Section 1.2 contains a description of exact algorithms for solving systems of polynomial equations. We also report on results in real algebraic geometry and symbolic computation for the sampling and emptiness problems for algebraic and semi-algebraic sets. The main subjects of Section 1.3 are spectrahedra and their projections, semidefinite programming, the polynomial optimization and related problems.

1.1 Algebra and geometry

1.1.1 Basics

We denote by $\mathbb{N} = \{0, 1, 2 \ldots\}$ the set of natural numbers, and by $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ respectively the ring of integer numbers and the fields of rational, real and complex numbers. We also introduce the notation $\mathbb{N}_* = \mathbb{N} \setminus \{0\}$, and similarly the notations $\mathbb{Z}_*, \mathbb{Q}_*, \mathbb{R}_*$ and $\mathbb{C}_*$.

If $\mathbb{F}$ is a symbol in $\{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$, we denote the $\mathbb{F}-$vector space of $p \times q$ matrices with entries in $\mathbb{F}$ by $\mathbb{M}_{p,q}(\mathbb{F})$, or by $\mathbb{F}^{p \times q}$ when a choice of a basis is explicited. For $M \in \mathbb{M}_{p,q}(\mathbb{F})$ and $r \leq \min\{p, q\}$, we denote by $\text{minors}\,(r, M)$ the list of $r \times r$ subminors of M. The general linear group of non-singular $p \times p$ matrices with entries in $\mathbb{F}$ is denoted by $\mathrm{GL}_p(\mathbb{F})$, and its unit element by $\mathbb{I}_p$. The determinant of $M \in \mathbb{M}_{p,p}(\mathbb{F})$ is denoted by $\det M$. The rank of $M \in \mathbb{M}_{p,q}(\mathbb{F})$ is denoted by $\operatorname{rank} M$. The vector space of symmetric $p \times p$ matrices is denoted by $\mathbb{S}_p(\mathbb{F})$.

For sets S and T, the expression $S \subset T$ means that S is included in T, while $S \subsetneq T$ means that $S \subset T$ and $S \neq T$ (in this case we say that S is a proper subset of T). If S is not included in T we write $S \not\subset T$. The number of elements of a finite set S, *i.e.* its cardinality, is denoted by $\#S$. Finally, we make frequently use of first order logical symbols such as $\wedge, \vee, \Longrightarrow, \neg$ and of quantifiers such as $\forall, \exists$.

Throughout the chapter, unless otherwise stated, x refers to the vector $(x_1, \ldots, x_n)$ of unknowns.

Basic commutative algebra

For this section, we refer for completeness to [39, 35]. Let $\mathbb{A}$ be a commutative ring. An ideal I of $\mathbb{A}$ is a subset $\{0\} \subset I \subset \mathbb{A}$, such that, for all $f_1, f_2 \in I$ and $g \in \mathbb{A}$, $f_1 - f_2 \in I$ and $gf_1 \in I$. If all elements of I can be written as a finite linear combination of elements of a set $S \subset \mathbb{A}$, with coefficients in $\mathbb{A}$, we say that S is a set of generators for I, and we write $I = \langle S \rangle$.

The ideal $I \subset \mathbb{A}$ is called

- prime, if $I \subsetneq \mathbb{A}$ and for all $a, b \in \mathbb{A}$, $ab \in I$ implies $a \in I$ or $b \in I$;
- primary, if for all $a, b \in \mathbb{A}$, $ab \in I$ implies $a \in I$ or $b^k \in I$ for some $k \in \mathbb{N}$;
- radical, if for all $a \in \mathbb{A}$ and $k \in \mathbb{N}$, $a^k \in I$ implies $a \in I$.

The ideal of elements $a \in \mathbb{A}$ satisfying $a^k \in I$ for some $k \in \mathbb{N}$, $k \geq 1$, is called the radical of I, and denoted by $\sqrt{I}$ (I is radical if and only if $I = \sqrt{I}$). In particular, any prime ideal is radical and the radical ideal of a primary ideal is prime.

Suppose that $\mathbb{A}$ is a Noetherian ring, *i.e.* any ideal $I \subset \mathbb{A}$ is finitely generated. Then any ideal I can be decomposed as the finite intersection of primary ideals, *i.e.* $I = P_1 \cap \cdots \cap P_s$, with $P_i \subset \mathbb{A}$ primary. This is called a primary decomposition of the ideal I, and is, in general, not unique. A minimal primary decomposition is obtained by asking $P_j \not\supset \cap_{i \neq j} P_i$ for all i [35]. Given a minimal primary decomposition, one deduces the related decomposition of the radical of I into prime ideals, *i.e.* $\sqrt{I} = \sqrt{P_1} \cap \cdots \cap \sqrt{P_s}$. The prime ideals $\sqrt{P_i}$ are uniquely determined (*cf.* [35, Th. 9, Ch. 4, § 7]). Given two ideals $I, J \subset \mathbb{A}$, the saturation $I \colon J^\infty$ is defined as the ideal

$$I \colon J^\infty = \{f \in \mathbb{A} : \exists\, k \in \mathbb{N}, J^k f \subset I\}.$$

Polynomial ideals and algebraic sets

Let $\mathbb{C}[x]$ be the ring of polynomials on x with complex coefficients. By Hilbert's Basissatz any ideal $I \subset \mathbb{C}[x]$ is generated by a finite set of elements of $\mathbb{C}[x]$, *i.e.* $\mathbb{C}[x]$ is a Noetherian ring.

Theorem 1.1 (Basissatz, Hilbert, 1890, [81])**.** *Let $I \subset \mathbb{C}[x]$. There exist $f_1, \ldots, f_s \in \mathbb{C}[x]$ such that, for all $g \in I$, there are $g_1, \ldots, g_s \in \mathbb{C}[x]$ with $g = g_1 f_1 + \ldots + g_s f_s$.*

With the previous notation, we write $I = \langle f_1, \ldots, f_s \rangle$ and we say that $f_1, \ldots, f_s$ generate the ideal I. We also refer to the vector $f = (f_1, \ldots, f_s)$ with the following two denominations:

- as a polynomial system, with implicit reference to the associated computational problem: Compute x such that $f_1(x) = 0, \ldots, f_s(x) = 0$;
- as an algebraic map, with explicit reference to the function: $f\colon \mathbb{C}^n \to \mathbb{C}^s$.

Definition 1.2 (Algebraic Set)**.** *Let $\mathbb{F}$ and $\mathbb{K}$ be subfields of $\mathbb{C}$. Let $I = \langle f_1, \ldots, f_s \rangle \subset \mathbb{K}[x]$ be any ideal, with $x = (x_1, \ldots, x_n)$. The associated locus of zeroes in $\mathbb{F}^n$*

$$\begin{aligned} Z_{\mathbb{F}}(I) &= \{x \in \mathbb{F}^n : \forall\, f \in I \;\; f(x) = 0\} \\ &= \{x \in \mathbb{F}^n : f_i(x) = 0,\, i = 1, \ldots, s\} \end{aligned}$$

is called a $\mathbb{K}-$definable $\mathbb{F}-$algebraic set.

We will be mostly concerned with $\mathbb{Q}-$definable $\mathbb{R}-$algebraic sets or $\mathbb{C}-$algebraic sets. In the whole text, whenever the ideal I is explicit, we denote the associated complex algebraic set using the calligraphic style $\mathcal{Z} = Z_{\mathbb{C}}(I)$. By definition, the algebraic subsets of $\mathbb{C}^n$ define the family of closed sets of a topology, called the Zariski topology. Indeed, the class of algebraic subsets of $\mathbb{C}^n$ is closed under finite union and arbitrary intersection. We denote the closure of a set $S \subset \mathbb{C}^n$ with respect to the Zariski topology by $\overline{S}$. If polynomials $f_1, \ldots, f_s \in \mathbb{C}[x]$ are homogeneous, we say that the algebraic set $\mathcal{Z} = Z_{\mathbb{C}}(\langle f_1, \ldots, f_s \rangle)$ is an affine cone (indeed, for $x \in \mathcal{Z}$ and $c \in \mathbb{C}$ then $cx \in \mathcal{Z}$).

Remark that, by definition, any non-empty Zariski open subset of $\mathbb{C}^n$ is dense in $\mathbb{C}^n$ with respect to the euclidean topology. For example, if $\mathbb{M}_{p,q}(\mathbb{C})$ is equipped with the Zariski topology, the subset of full-rank matrices (*e.g.* $\mathrm{GL}_n(\mathbb{C}) \subset \mathbb{M}_{n,n}(\mathbb{C})$) is open and dense, since it is the union of the complements of the algebraic sets defined by the vanishing of maximal minors.

We call the $\mathbb{R}-$algebraic set $Z_{\mathbb{R}}(I)$ a real algebraic set. Suppose that $Z_{\mathbb{R}}(I)$ is $\mathbb{Q}-$definable. Observe that up to doubling the maximum degree of polynomials defining $Z_{\mathbb{R}}(I)$, one can describe it by a single equation:

$$\forall\, x \in \mathbb{R}^n \;\; \Big(\forall\, i \;\; f_i(x) = 0 \iff f_1^2(x) + \ldots + f_s^2(x) = 0\Big)\,.$$

Definition 1.3 (Vanishing Ideal)**.** *Let $S \subset \mathbb{C}^n$ be any set. The set of polynomials vanishing on S*

$$I(S) = \{f \in \mathbb{C}[x] : \forall\, x \in S,\;\; f(x) = 0\},$$

is an ideal of $\mathbb{C}[x]$ and is called the vanishing ideal of S.

For $M \in \mathrm{GL}_n(\mathbb{C})$, and $f \in \mathbb{C}[x]$, we denote by $f \circ M$ the polynomial such that $(f \circ M)(x) = f(M\,x)$ for all $x \in \mathbb{C}^n$. If $f = (f_1, \ldots, f_s) \subset \mathbb{C}[x]$, and $\mathcal{Z} = Z_{\mathbb{C}}(f)$, we denote by $M^{-1}\mathcal{Z}$ the set defined by $\langle f_1 \circ M, \ldots, f_s \circ M \rangle$, that is $M^{-1}\mathcal{Z} = \{x \in \mathbb{C}^n : M\,x \in \mathcal{Z}\}$.

It is easy to check that vanishing ideals are radical ideals. Hilbert's Nullstellensatz asserts that all radical ideals can be described as the vanishing ideal of an algebraic set.

Theorem 1.4 (Nullstellensatz, Hilbert, 1893, [82])**.** *Let $I = \langle f_1, \ldots, f_s \rangle \subset \mathbb{C}[x]$. Then f vanishes over $Z_{\mathbb{C}}(I)$ if and only if there exist $g_1, \ldots, g_s$ and $k \geq 1$ such that $f^k = g_1 f_1 + \cdots + g_s f_s$. In symbols: $I(Z_{\mathbb{C}}(I)) = \sqrt{I}$.*

Hence, by Hilbert's Nullstellensatz, operators $Z_{\mathbb{C}}(\cdot)$ and $I(\cdot)$ define a bijective correspondence between radical ideals and algebraic sets:

$$\begin{tikzcd} \boxed{\begin{array}{c}\textbf{radical ideals}\\ I = \sqrt{I}\end{array}} \arrow[r, shift left] & \boxed{\begin{array}{c}\textbf{algebraic sets}\\ \mathcal{Z} = \overline{\mathcal{Z}}\end{array}} \arrow[l, shift left] \end{tikzcd}$$

The quotient $\mathbb{C}[x] \big/ I(Z)$ is called the coordinate ring of the algebraic set $\mathcal{Z}$: it is the ring of polynomial functions defined over $\mathcal{Z}$. Also, for any set $S \subset \mathbb{C}^n$, the set $Z_{\mathbb{C}}(I(S)) \subset \mathbb{C}^n$ is an algebraic set and it is the smallest one containing S, hence it equals $\overline{S}$.

There exists an equivalent form of Hilbert's Nullstellensatz: let $I \subset \mathbb{C}[x]$, then $1 \in I$ if and only if $Z_{\mathbb{C}}(I) = \emptyset$. Remark that the weak Nullstellensatz states that one can *a priori* produce a certificate of emptiness of the algebraic set $Z_{\mathbb{C}}(\langle f_1, \ldots, f_s \rangle)$ (in other words, an exact proof that the polynomial system $f_1(x) = 0, \ldots, f_s(x) = 0$ has no complex solutions) of type $1 = g_1 f_1 + \ldots + g_s f_s$ for some polynomials g_i (cf. Section 1.2.2 for computational aspects).

Observe that this statement is false for polynomial rings over non algebraically closed fields: $\langle x^2 + 1 \rangle \subsetneq \mathbb{R}[x]$ but $Z_{\mathbb{R}}(\langle x^2 + 1 \rangle) = \emptyset$. Always following the meaning in terms of algebraic certificates of Hilbert's Nullstellensatz, we give below the real version of this result.

Theorem 1.5 (Real Nullstellensatz, Th. 4.1.4 in [24])**.** *Let $I = \langle f_1, \ldots, f_s \rangle \subset \mathbb{R}[x]$. Then $Z_{\mathbb{R}}(I) = \emptyset$ if and only if there exists $g_1, \ldots, g_s, h \in \mathbb{C}[x]$ such that h is a sum of squares and $-1 = h + g_1 f_1 + \ldots + g_s f_s$.*

Concerning the previous example, a Real Nullstellensatz certificate for the emptiness of $Z_{\mathbb{R}}(\langle x^2 + 1 \rangle)$ is given by the equality $-1 = x^2 + (-1)(x^2 + 1)$.

Definitions for algebraic sets

We set some standard definition on the geometry of algebraic sets. For further details we refer to classical textbooks [145, 62, 63].

Definition 1.6 (Irreducibility)**.** *An $\mathbb{F}-$algebraic set $\mathcal{Z} \subset \mathbb{F}^n$ is called irreducible if any decomposition $\mathcal{Z} = \mathcal{Z}_1 \cup \mathcal{Z}_2$ where $\mathcal{Z}_i$ are $\mathbb{F}-$algebraic sets, implies either $\mathcal{Z} = \mathcal{Z}_1$ or $\mathcal{Z} = \mathcal{Z}_2$.*

The algebraic set $\mathcal{Z} \subset \mathbb{C}^n$ is irreducible if and only if its vanishing ideal $I(\mathcal{Z})$ is prime (*cf.* [35, Prop. 3, Ch. 4, § 5]), otherwise it can be decomposed as the finite union of irreducible algebraic sets, *i.e.*

$$\mathcal{Z} = \mathcal{Z}_1 \cup \cdots \cup \mathcal{Z}_s,$$

with $\mathcal{Z}_i$ irreducible, and such a decomposition is unique up to supposing $\mathcal{Z}_i \not\subset \mathcal{Z}_j$ for $i \neq j$. If this is the case, the sets $\mathcal{Z}_i$ are called the irreducible components of $\mathcal{Z}$ and $I(\mathcal{Z}) = I(\mathcal{Z}_1) \cap \cdots \cap I(\mathcal{Z}_s)$.

Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set and let $I(\mathcal{Z}) = \langle f_1, \ldots, f_s \rangle$. The $s \times n$ matrix

$$Df = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_s}{\partial x_1} & \cdots & \frac{\partial f_s}{\partial x_n} \end{pmatrix}$$

is the Jacobian matrix associated to the polynomial system $f = (f_1, \ldots, f_s)$.

Definition 1.7 (Dimension and codimension)**.** *Let $\mathcal{Z} \subset \mathbb{C}^n$ be irreducible. The codimension of $\mathcal{Z}$ is the maximum rank c attained by Df on $\mathcal{Z}$, and its dimension is $d = n - c$. The dimension of an algebraic set is the maximum dimension d of its irreducible components, and its codimension is $c = n - d$.*

Remark that here we have defined the dimension of complex algebraic sets $\mathcal{Z}$, which typically does not correspond to the dimension of the real trace $\mathcal{Z} \cap \mathbb{R}^n$.

For $0 \leq i \leq d$, we denote by $\Omega_i(\mathcal{Z})$ the union of the irreducible components of $\mathcal{Z}$ of dimension i, and we call it its equidimensional component of dimension i. A set $\mathcal{Z}$ of dimension d is equidimensional if $\mathcal{Z} = \Omega_d(\mathcal{Z})$. Finally, algebraic sets of dimension one are typically called curves. Algebraic sets of codimension one are called hypersurfaces and their vanishing ideals are defined by a single polynomial.

A linear space $\mathcal{L} \subset \mathbb{C}^n$ is an algebraic set defined by polynomials of degree one. A linear space which is a hypersurface is called a hyperplane.

Definition 1.8 (Degree)**.** *The degree of an equidimensional algebraic set $\mathcal{Z} \subset \mathbb{C}^n$ of codimension c is the maximum cardinality of finite intersections $\mathcal{Z} \cap \mathcal{L}$ where $\mathcal{L} \subset \mathbb{C}^n$ runs over the collection of linear subspaces of $\mathbb{C}^n$ of dimension c. The degree of an algebraic set $\mathcal{Z}$ is the sum of the degrees of its equidimensional components.*

Let $\mathcal{Z} \subset \mathbb{C}^n$ be equidimensional of codimension c, and let $I(\mathcal{Z}) = \langle f_1, \ldots, f_s \rangle$. The singular locus of $\mathcal{Z}$, denoted by $\operatorname{sing}(\mathcal{Z})$, is the algebraic set defined by $f = (f_1, \ldots, f_s)$ and by all $c \times c$ minors of Df (*cf.* [62, Lect. 14]). If $\operatorname{sing}(\mathcal{Z}) = \emptyset$ we say that $\mathcal{Z}$ is smooth, otherwise singular. The points in $\operatorname{sing}(\mathcal{Z})$ are called singular, while points in $\operatorname{reg}(\mathcal{Z}) = \mathcal{Z} \setminus \operatorname{sing}(\mathcal{Z})$ are called regular. We use frequently the following theorem to prove regularity properties of algebraic sets.

Theorem 1.9 (Jacobian criterion, Th. 16.19 in [39])**.** *Let $\mathcal{Z} \subset \mathbb{C}^n$ and $I(\mathcal{Z}) = \langle f \rangle$.*

1. *If $\mathcal{Z}$ is equidimensional of dimension d, then the rank of Df at* $\operatorname{reg}(\mathcal{Z})$ *is $c = n - d$, and* $\operatorname{sing}(\mathcal{Z})$ *is defined by f and by the $c \times c$ minors of Df.*

2. *If the rank of Df is constant over $\mathcal{Z}$ and equal to c, then $\mathcal{Z}$ is smooth and equidimensional of dimension $d = n - c$.*

Definition 1.10 (Critical points, critical and regular values)**.** *Let $\mathcal{Z} \subset \mathbb{C}^n$ be smooth and equidimensional of codimension c, and let $I(\mathcal{Z}) = \langle f_1, \ldots, f_s \rangle$. Let $g : \mathbb{C}^n \to \mathbb{C}^m$ be an algebraic map. The critical points of the restriction of g to $\mathcal{Z}$ is the algebraic set denoted by* $\operatorname{crit}(g, \mathcal{Z})$ *and defined by $f = (f_1, \ldots, f_s)$ and by all $c + m$ minors of the Jacobian matrix*

$$D(f, g) = \begin{pmatrix} Df \\ Dg \end{pmatrix}.$$

Let $Im(g)$ denote the image of g. The points in $g(\operatorname{crit}(g, \mathcal{Z}))$ are called critical values of the restriction of g to $\mathcal{Z}$, while points in $Im(g) \setminus g(\operatorname{crit}(g, \mathcal{Z}))$ are called the regular values.

For $0 \leq i \leq n$, we denote by $\pi_i \colon \mathbb{C}^n \to \mathbb{C}^i$ the projection $\pi_i(x_1, \ldots, x_n) = (x_1, \ldots, x_i)$. Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set. The polar varieties associated to $\mathcal{Z}$ are defined as the sets of critical points of maps π_i restricted to $\mathcal{Z}$.

Definition 1.11 (Polar varieties)**.** *Let $f = (f_1, \ldots, f_s) \subset \mathbb{Q}[x]$, with $x = (x_1, \ldots, x_n)$, and $\mathcal{Z} = Z_{\mathbb{C}}(f) \subset \mathbb{C}^n$ be of dimension d. For $0 \leq i \leq d$, let $D(f, i)$ be the matrix obtained by eliminating the first i columns of Df. Let I_i be the ideal generated by f and by the $n - d$ minors of $D(f, i)$. For $0 \leq i \leq d$, the algebraic set $\mathcal{W}_i = Z_{\mathbb{C}}(I_i)$ is called the $i-$th polar variety of $\mathcal{Z}$.*

We remark that, in the previous definition, under the assumption that f generates a radical ideal and that $\mathcal{Z}$ is smooth and equidimensional, the polar variety $\mathcal{W}_i$ encodes the critical points of the restriction of the map π_i to $\mathcal{Z}$ (*cf.* [137]).

Semi-algebraic sets

Let $f \in \mathbb{Q}[x]$, with $x = (x_1, \ldots, x_n)$. A sign condition on f is one of the three possible formulas $f(x) > 0, f(x) < 0$ or $f(x) = 0$. Then, a semi-algebraic set is defined by a finite logical combination (using $\wedge$, $\vee$, or $\neg$) of sign conditions on a given finite set of polynomials. Equivalently:

Definition 1.12. *Let $f_1, \ldots, f_s \in \mathbb{Q}[x]$. The associated basic closed semi-algebraic set is the set*

$$S = \{x \in \mathbb{R}^n : f_i(x) \geq 0, \, i = 1, \ldots, s\}.$$

A set $S \subset \mathbb{R}^n$ is called semi-algebraic if it can be generated by a finite sequence of unions, intersections and complements from a collection of basic closed semi-algebraic sets.

The number s of polynomial inequalities that are needed to describe a basic semi-algebraic set S can be bounded by the inequality $s \le \frac{n(n+1)}{2}$, and this result is due to Bröcker and Scheiderer (*cf.* [24, Th. 10.4.8]). Remark here that also the set $\{x \in \mathbb{R}^n : f_i(x) > 0,\ i = 1, \ldots, s\}$ is semi-algebraic, and in this case s can be bounded linearly with respect to n, indeed $s \le n$ for some polynomials $f_1, \ldots, f_s \in \mathbb{Q}[x]$ (*cf.* [24]). Finally, the interior of a semi-algebraic set is semi-algebraic.

As for real algebraic sets, a basic semi-algebraic set does not preserve this property under linear projections. Moreover, the whole class of semi-algebraic sets is closed under this operation, as stated by Tarski-Seidenberg Principle.

Theorem 1.13 (Tarski-Seidenberg Principle, Th. 1.4.2 in [24])**.** *Let $S \subset \mathbb{R}^{n+d}$ be a semi-algebraic set, and let $\pi \colon \mathbb{R}^{n+d} \to \mathbb{R}^n$ be the projection on the first n variables. Then $\pi(S) \subset \mathbb{R}^n$ is a semi-algebraic set.*

Semi-algebraic sets are an important class of topological spaces, with special properties and arising in many contexts in mathematics. For this reason, a dedicated homological theory has been developed to measure their topological complexity (see [20, Ch. 7, §4] and references therein). So far, bounds on the number of connected components (and, more generally, on sums of Betti numbers) of algebraic or semi-algebraic sets are available in the literature. Results similar to the following theorem were established previously by J. Milnor and R. Thom.

Theorem 1.14 (Th. 1 in [18], Th. 1.1 in [19])**.** *Let $g = (g_1, \ldots, g_s) \subset \mathbb{Q}[x]$, with $x = (x_1, \ldots, x_n)$, such that $Z_{\mathbb{C}}(\langle g \rangle)$ has dimension k, and let $f = (f_1, \ldots, f_s) \subset \mathbb{Q}[x]$ be polynomials of degree at most d. Then the number of connected components of the semi-algebraic set $\{x \in \mathbb{R}^n : g = 0, f_i \ge 0, i = 1, \ldots, s\}$, is in $\binom{s}{k} (\mathcal{O}(d))^n$.*

The previous theorem implies that the number of connected components of any semi-algebraic set is singly exponential in the number of variables.

1.1.2 Algebraic transversality

The concept of transversality has its origins in differential geometry, to describe properties of intersections of manifolds. Intuitively, a family of vector subspaces $\mathcal{V}_1, \ldots, \mathcal{V}_t$ (or, more generally, of submanifolds) of a given vector space $\mathcal{E}$ (resp. a manifold) are transverse if and only if $\operatorname{codim}(\mathcal{V}_1 \cap \ldots \cap \mathcal{V}_t) = \operatorname{codim} \mathcal{V}_1 + \cdots + \operatorname{codim} \mathcal{V}_t$. If this is the case, the intersection of $\mathcal{V}_i$ has the expected dimension.

The transfer of transversality results in algebraic geometry started with the seminal paper of Mather [110] (*cf.* also [36, Ch. 3] and [6, 5]). A modern approach discussing the use of transversality in effective algebraic geometry is in [139, Ch. 4].

Algebraic properties

For the next definition, $\mathcal{Z} \subset \mathbb{C}^n$ is an algebraic set and $\mathscr{O} \subset \mathcal{Z}$ is a Zariski open subset of $\mathcal{Z}$.

Definition 1.15. *An algebraic property* P *(or, simply, a property) defined over the set* $\mathscr{O}$ *is a function*

$$\mathsf{P} \colon \mathscr{O} \to \{\texttt{true}, \texttt{false}\}.$$

We say that $x \in \mathcal{Z}$ *satisfies* P *if* $\mathsf{P}(x) = \texttt{true}$*. We say that* P *holds generically in* $\mathscr{O}$ *(or that it is an open condition on* $\mathscr{O}$*) if there exists a non-empty Zariski open set* $\mathscr{P} \subset \mathscr{O}$ *such that, for all* $x \in \mathscr{P}$*,* x *satisfies* P*.*

Examples of properties are listed here:

- Let $\mathsf{P} \colon \mathbb{M}_{n,n}(\mathbb{C}) \to \{\texttt{true}, \texttt{false}\}$ assign $\texttt{true}$ if M has full rank, $\texttt{false}$ otherwise. The preimage of $\texttt{true}$ is $\mathrm{GL}_n(\mathbb{C})$, hence P holds generically in $\mathbb{M}_{n,n}(\mathbb{C})$.

- Let $\mathsf{P} \colon \mathbb{R} \to \{\texttt{true}, \texttt{false}\}$ be the property satisfied by nonnegative real numbers. Then P does not hold generically since Zariski open subsets of $\mathbb{R}$ are complements of finite sets.

- Let $f = \sum_{i=0}^{d_1} f_i x^i, g = \sum_{i=0}^{d_2} g_i x^i \in \mathbb{Q}[x]$ be univariate polynomials. Identify f, g with their vectors of coefficients, and let $\mathbb{Q}[f, g]$ be the space of polynomials on $f = (f_0, \ldots, f_{d_1}), g = (g_0, \ldots, g_{d_2})$. The resultant $R(f, g)$ of the couple (f, g) is an element of $\mathbb{Q}[f, g]$, of degree $d_1 + d_2$ (*cf.* example of Section 5.2.2, on page 150). The map $\mathsf{P} \colon \mathbb{Q}[x] \times \mathbb{Q}[x]$ returning $\mathsf{P}(f, g) = \texttt{true}$ when (f, g) share no common factors is algebraic, and holds if and only if $R(f, g) \neq 0$, hence generically.

Transversality theory allows, via the two results discussed in the next sections, to produce sufficient conditions to establish that algebraic properties hold generically over their domains.

Sard's lemma

Let $\mathcal{Z} \subset \mathbb{C}^n$ be an equidimensional algebraic set of dimension d, and let $g \colon \mathbb{C}^n \to \mathbb{C}^s$ be an algebraic map. Remark that we do not assume that $\mathcal{Z}$ is smooth. By abuse of notation we denote by $\mathrm{crit}\,(g, \mathcal{Z})$ the set of critical points of the restriction of g to the regular points $\mathrm{reg}\,(\mathcal{Z})$ of the set $\mathcal{Z}$. Precisely, a point $x \in \mathrm{reg}\,(\mathcal{Z})$ belongs to $\mathrm{crit}\,(g, \mathcal{Z})$ if and only if the rank of $D(f, g)$ at x is less than $c + s$ where $c = n - d$ is the codimension of $\mathcal{Z}$, and $g \colon \mathbb{C}^n \to \mathbb{C}^s$.

The following lemma is the algebraic version of [36, Th. 3.5.1].

Lemma 1.16 (Sard's Lemma, Sec. 4.2 in [139], or Ch. 2, Sec. 6.2, Th .2 in [145])**.** *The set* $g(\mathrm{crit}\,(g, \mathcal{Z})) \subset \mathbb{C}^s$ *is contained in a proper hypersurface of* $\mathbb{C}^s$*.*

For a modern proof of Sard's Lemma we refer to [139, Prop. 4.2]. Sard's Lemma implies that the fiber over a generic point of the target space does not contain any critical points of the restriction of the map to the algebraic set, and hence is smooth. In the following example we explore a direct application.

Example 1.17. *Let $f \in \mathbb{Q}[x]$ be any non-constant polynomial. Its associated algebraic set $Z_{\mathbb{C}}(f) \subset \mathbb{C}^n$ is a hypersurface, hence equidimensional of codimension 1. Suppose that $Df \neq 0$ for all $x \in Z_{\mathbb{C}}(f)$, so that $Z_{\mathbb{C}}(f)$ is smooth.*

Let $g \colon \mathbb{C}^n \to \mathbb{C}$ be an algebraic map (represented by one polynomial $g \in \mathbb{Q}[x]$), and consider the restriction of g to $Z_{\mathbb{C}}(f)$. By Sard's Lemma (Lemma 1.16) the set of critical values of g on $Z_{\mathbb{C}}(f)$ is finite, denoted by $\{y_1, \ldots, y_t\}$. We deduce that, if $y \in \mathbb{C} \setminus \{y_1, \ldots, y_t\}$, then either $g^{-1}(y) = \{x \in \mathbb{C}^n : g(x) = y\}$ does not intersect $Z_{\mathbb{C}}(f)$, or the intersection is smooth and equidimensional of codimension 2 (by the Jacobian criterion, cf. Theorem 1.9).

We conclude by remarking that Example 1.17 shows how Sard's Lemma can be used to check transversality properties of generic intersections of algebraic varieties.

Thom's weak transversality

We move to a second result in algebraic transversality theory, dealing with the particular case of transversality to a point. We consider an algebraic map

$$\begin{array}{rccc} f: & \mathbb{C}^n \times \mathbb{C}^k & \longrightarrow & \mathbb{C}^s \\ & (x,y) & \longmapsto & f(x,y). \end{array}$$

For $y \in \mathbb{C}^k$, we denote by f_y the section map

$$\begin{array}{rccc} f_y: & \mathbb{C}^n & \longrightarrow & \mathbb{C}^s \\ & x & \longmapsto & f(x,y). \end{array}$$

Theorem 1.18 (Thom's Weak Transversality, Th 3.7.4 in [36])**.** *Let $\mathscr{U} \subset \mathbb{C}^n$ be a non-empty Zariski open set such that 0 is a regular value for the restriction of f to $\mathscr{U} \times \mathbb{C}^k$. Then there exists a non-empty Zariski open set $\mathscr{V} \subset \mathbb{C}^k$ such that, for $y \in \mathscr{V}$, 0 is a regular value of the map $f_y \colon \mathbb{C}^n \to \mathbb{C}^s$.*

We show via a simple example how this result can be used to establish algebraic properties.

Example 1.19. *Let $g = (g_1, \ldots, g_s) \in \mathbb{Q}[x]^s$ be any polynomial system. Consider the following map:*

$$\begin{array}{rccc} f: & \mathbb{C}^n \times \mathbb{C}^s & \longrightarrow & \mathbb{C}^s \\ & (x,y) & \longmapsto & (g_1 - y_1, \ldots, g_s - y_s). \end{array}$$

The $s \times (s+n)$ Jacobian matrix Df contains the submatrix $-\mathbb{I}_s$ of derivatives of polynomials $g_i - y_i$ with respect to variables in y. By the Jacobian criterion

(Theorem 1.9) we conclude that the algebraic set $Z_{\mathbb{C}}(f) \subset \mathbb{C}^n \times \mathbb{C}^s$ is empty or smooth, hence that 0 *is a regular value for* f*.*

By Thom's Weak Transversality Theorem (Theorem 1.18) we deduce that there exists a non-empty Zariski open set $\mathscr{V} \subset \mathbb{C}^s$ such that, if $y \in \mathscr{V}$, the zero set of f_y is empty or smooth. We deduce that for $y \in \mathscr{V}$, the fiber set $g^{-1}(y) = \{x \in \mathbb{C}^n : g(x) = y\}$ is either smooth or empty.

The previous example shows the intuitive fact that up to perturbing the level set of a polynomial function or map, then either the obtained set is empty, or one necessarily deduces expected dimension or regularity properties.

1.1.3 Noether position and properness

The notion of Noether position of an ideal I (or of the zero set $\mathcal{Z} = Z_{\mathbb{C}}(I)$) is related to the definition of dimension and on that of algebraic dependence in the associated quotient ring $\mathbb{C}[x] \big/ I(\mathcal{Z})$.

Normalization of algebraic sets

Let $A \subset B$ be a ring extension. We recall that an element $b \in B$ is integral over A if there exists a monic polynomial $f \in A[t]$ such that $f(b) = 0$. If all elements of B are integral over A, the extension $A \subset B$ is called integral. Moreover, elements $b_1, \ldots, b_s \in B$ are said to be algebraically independent over A if there are no $s-$variate polynomials with coefficients in A vanishing over $(b_1, \ldots, b_s)$.

Definition 1.20 (Noether Position, *cf.* [98, 39])**.** *Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d. We say that $\mathcal{Z}$ (or $I(\mathcal{Z})$) is in Noether position with respect to $x_1, \ldots, x_d$, if the extension*

$$\mathbb{C}[x_1, \ldots, x_d] \subset \mathbb{C}[x_1, \ldots, x_n] \big/ I(\mathcal{Z})$$

is integral.

For example, finite sets are all in Noether position (with $d = 0$).

Example 1.21. *An example of ideal which is not in Noether position is $I = \langle x_1 x_2 \rangle \subset \mathbb{C}[x_1, x_2]$. Indeed, the algebraic set $Z_{\mathbb{C}}(x_1 x_2) \subset \mathbb{C}^2$ has dimension 1, hence one is asking whether the inclusion of $\mathbb{C}[x_1]$ in the coordinate ring of I is integral. This is not the case since, e.g., the class of x_2 modulo I has as minimum polynomial over $\mathbb{C}[x_1]$ the polynomial $f(t) = x_2 t$. Remark that the fiber over 0 of the projection of $Z_{\mathbb{C}}(x_1 x_2)$ over the line $x_2 = 0$ consists of an entire line.*

Remark here that substituting e.g. $x_1 + x_2$ into x_1 yields the following action on the ideal I: $I \mapsto \langle x_1 x_2 + x_2^2 \rangle$ and puts the ideal I in Noether position with respect to x_1, since now x_2 is integral over $\mathbb{C}[x_1]$. Remark here that any fiber of the projection of $Z_{\mathbb{C}}(x_1 x_2 + x_2^2)$ over the line $x_2 = 0$ is finite of cardinality 2.

The behavior described in Example 1.21 is proved in a historical result in algebraic geometry, due to Hilbert.

Theorem 1.22 (Noether Normalization, [82], Th. 2.1 in [98], Th. 13.3 in [39])**.** *Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d. Then there exists $y_1, \ldots, y_d \in \mathbb{C}[x]$ such that the extension*

$$\mathbb{C}[y_1, \ldots, y_d] \subset \mathbb{C}[x_1, \ldots, x_n] \big/ I(\mathcal{Z})$$

is integral.

Geometrically, Noether's Theorem implies that for any algebraic set of dimension d there exists a finite-to-one map from the set to the affine space of dimension d (this is what is called a Noether normalization, *cf.* [145, Th. 10, Sec. 5.4]).

This theoretical result led to effective counterparts lying at the foundation of algorithms in computational algebraic geometry. Indeed, the Noether position can be recovered by applying generic linear changes of variables occurring in the ideal I [38, Cor. 2.4.4], hence by means of a probabilistic algorithm. Complexity bounds for this procedure have been computed (*cf.* [52, Th. 2.3]).

Proper maps

A classical notion in topology is that of proper maps.

Definition 1.23 (Proper Map)**.** *Let $\mathcal{X}, \mathcal{Y}$ be topological spaces, and let $f \colon \mathcal{X} \to \mathcal{Y}$ be a continuous function. The function f is called proper at a point $y \in \mathcal{Y}$ if there exists an open neighborhood $\mathcal{O} \subset \mathcal{Y}$ of y such that $f^{-1}(\overline{\mathcal{O}})$ is compact, where $\overline{\mathcal{O}}$ is the closure of $\mathcal{O}$ in the topology of $\mathcal{Y}$.*

We will mainly use Euclidean topological spaces, and frequently $\mathbb{R}^n$ or $\mathbb{C}^n$. The Noether position of an algebraic set $\mathcal{Z}$ is related to properness properties of the restriction of linear projections to $\mathcal{Z}$, as stated in the following proposition.

Proposition 1.24. *Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d. Then $\mathcal{Z}$ is in Noether position with respect to $x_1, \ldots, x_d$ if and only if the restriction to $\mathcal{Z}$ of the map $\pi_d \colon \mathbb{C}^n \to \mathbb{C}^d$ sending x to $(x_1, \ldots, x_d)$, is proper. If one of the previous conditions holds, the restriction of the map π_d to the real trace $\mathcal{Z} \cap \mathbb{R}^n$ is also proper.*

Proof. By [87, Prop. 3.2], the restriction of π_d to $\mathcal{Z}$ is proper if and only if the map $\pi_{d,*} \colon \mathbb{C}[x_1, \ldots, x_d] \to \mathbb{C}[x] \big/ I(\mathcal{Z})$ is finite. This, by [87, Prop. 3.2], holds if and only if the extension $\mathbb{C}[x_1, \ldots, x_d] \subset \mathbb{C}[x] \big/ I(\mathcal{Z})$ is integral, that is if and only if $\mathcal{Z}$ is in Noether position with respect to $x_1, \ldots, x_d$.

Suppose that the restriction of the map π_d to $\mathcal{Z}$ is proper, and let $y \in \pi_d(\mathcal{Z} \cap \mathbb{R}^n) \subset \mathbb{R}^d$. In particular $y \in \pi_d(\mathcal{Z})$, hence by the properness property there exists an open neighborhood $\mathscr{O} \subset \pi_d(\mathcal{Z})$ containing y such that $\pi_d^{-1}(\overline{\mathscr{O}})$ is compact. Then also $\mathscr{O} \cap \mathbb{R}^d$ is open in $\mathbb{R}^d$, it contains y, and $\pi_d^{-1}(\overline{\mathscr{O} \cap \mathbb{R}^d}) \cap \mathbb{R}^n$ is compact. *q.e.d.*

Coming back to Example 1.21, one can notice that the projection over the first variable of the algebraic set associated to $I = \langle x_1 x_2 \rangle$ is not a proper map, since it is not proper at the point 0. Indeed, we proved in the example that I is not in Noether position. Conversely, the set defined by $I = \langle x_1 x_2 + x_2^2 \rangle$ projects properly onto x_1, indeed the ideal I is in Noether position with respect to x_1.

In the last years, a lot of efforts have been made to develop the relations between Noether position and properties of algebraic maps. In particular, in [137] the authors exploit properness of linear projections to design algorithms for solving positive dimensional polynomial systems whose zero set is non-compact, and under regularity hypotheses.

Given $f \subset \mathbb{Q}[x]$ a polynomial system, let $\mathcal{V} = Z_{\mathbb{C}}(f) \subset \mathbb{C}^n$, $d = \dim \mathcal{V}$ and let $\mathcal{W}_i$ be the $i-$th polar variety associated to $\mathcal{V}$, defined on page 38. If $\mathcal{V}$ is smooth and equidimensional, and if its defining polynomial system generates a radical ideal, recall that $\mathcal{W}_i$ (defined by suitable minors of suitable truncations of the Jacobian matrix of $\mathcal{V}$) encodes the critical points of the restriction of the projection $\pi_i(x) = (x_1, \ldots, x_i)$ to $\mathcal{V}$. We also recall that for $\mathcal{Z} \subset \mathbb{C}^n$, $M^{-1}\mathcal{Z}$ denotes the set $\{x \in \mathbb{C}^n : M\,x \in \mathcal{Z}\}$.

Theorem 1.25 (Th.1 in [137])**.** *Suppose that $\mathcal{V}$ is smooth and equidimensional, and if its defining polynomial system generates a radical ideal. There exists a Zariski open set $\mathscr{O} \subset \mathrm{GL}_n(\mathbb{C})$ such that, if $M \in \mathscr{O} \cap \mathbb{M}_{n,n}(\mathbb{Q})$, and $i = 1, \ldots, d+1$, the restriction of the projection π_{i-1} to $M^{-1}\mathcal{W}_i$ is proper.*

In practice, Theorem 1.25 is proved by showing that a generic change of variables put the algebraic set $\mathcal{W}_i$ in Noether position with respect to $x_1, \ldots, x_{i-1}$. This result holds under regularity hypotheses for the input system, and one contribution of this thesis, contained in Chapter 2, is to generalize this fact by eliminating such regularity assumptions by defining a suitable decomposition of the algebraic set $\mathcal{Z}$.

1.2 Exact algorithms for polynomial system solving

In this section we discuss basic details in complexity theory, symbolic computation and effective real algebraic geometry. We do not claim to be comprehensive, and we refer to [30, 26] and to [20, 35] for completeness.

The problem of computing the set of complex solutions of a polynomial system $f = (f_1, \ldots, f_s)$, when this set is finite, lies at the basis of modern mathematics and finds numerous applications. Both numerical and exact algorithms have been developed in the last decades with the goal, especially for practical contexts, to isolate all real solutions. Since, typically, one deals with positive dimensional systems, a first step is to design algorithms to reduce the dimension, possibly exploiting the structure of the input polynomials.

In this thesis we are interested in solving determinantal polynomial systems in exact arithmetic. In particular, the output will be a finite set $\mathcal{Z} \subset \mathbb{C}^n$ encoded by a system of equations involving univariate rational functions. Every $x =$

$(x_1, \ldots, x_n) \in \mathcal{Z}$ verifies the following polynomial system:

$$\begin{cases} x_1 = \frac{q_1(t)}{q_0(t)} \\ \vdots \\ x_n = \frac{q_n(t)}{q_0(t)} \\ q_{n+1}(t) = 0, \end{cases}$$

for some univariate polynomials $q_0, q_1, \ldots, q_n, q_{n+1} \in \mathbb{Q}[t]$, such that q_0 and q_{n+1} are coprime, and for some root t of q_{n+1}. Moreover, the degree of q_{n+1} equals the cardinality of $\mathcal{Z}$ and there is a bijection between the roots of q_{n+1} and the points in $\mathcal{Z}$. The elements of $\mathcal{Z}$ can be represented by q_{n+1} and intervals isolating the corresponding root.

This univariate representation [136], dating back to Kronecker's work, can be computed in practice (using Gröbner bases, see Section 1.2.1) and reduces the input problem to the isolation of the roots of the univariate equation $q_{n+1}(t) = 0$. This reduced task can be performed via different algorithms (*cf., e.g.,* [123]).

The algorithms described in this thesis are exact and manipulate polynomial systems representing algebraic sets. This means that the input and the output are vectors of polynomials and that the intermediate operations are performed using admissible ring or field operations such as sums, subtractions, multiplications or divisions. We measure the complexity of our algorithms taking into account the following two quantities:

1. the total number of arithmetic operations $\{+, -, \times, \div\}$ performed over the ground field $\mathbb{Q}$;

2. the degree of the polynomial q_{n+1} in the output representation.

Obviously, one goal is to obtain algorithms for which the growth in terms of output size (represented by the degree of q_{n+1}, Point 2 above) does not imply exponential growth in terms of needed operations (Point 1 above).

To give asymptotic bounds on the previous two quantities, we use the classical notations $\mathcal{O}$ and $\tilde{\mathcal{O}}$ that we recall next. Let $f, g \colon \mathbb{N}^k \to \mathbb{N}$ be $k-$variate functions of integer variables with values in $\mathbb{N}$. We say that $g \in \mathcal{O}(f)$ if there exists a constant $c \in \mathbb{N}$ such that $g(n_1, \ldots, n_k) \leq cf(n_1, \ldots, n_k)$ for all $n_1, \ldots, n_k \in \mathbb{N}$. We say that $g \in \tilde{\mathcal{O}}(f)$ if there exists $a \in \mathbb{N}$ such that $g \in \mathcal{O}(f \, \log^a f)$.

1.2.1 Gröbner bases

The theory of Gröbner bases has been developed since Buchberger's PhD thesis in 1965. Its goal is to solve polynomial equations. Since his work, further improvements of Buchberger's algorithm have been achieved [35, Ch. 2, § 9], mainly in the last two decades [41, 42], and nowadays we have efficient tools and implementations for solving algebraic systems of equations using Gröbner bases [43].

Let $x = (x_1, \ldots, x_n)$. We associate to a vector with n nonnegative integer entries $i = (i_1, \ldots, i_n) \in \mathbb{N}^n$, a monomial $x^i = x_1^{i_1} \cdots x_n^{i_n} \in \mathbb{Q}[x]$ of degree $i_1 + \cdots + i_n$. A monomial ordering $>$ on $\mathbb{Q}[x]$ is a total ordering on $\mathbb{N}^n$ satisfying the following two properties: (1) if $i > j$ then $i + \ell > j + \ell$ for any $\ell \in \mathbb{N}^n$, and (2) every decreasing chain $i > j > \ell > \cdots$ stops. With respect to this ordering we say that $x^i > x^j$ if $i > j$ (*cf.* [35, Ch. 2, § 2]).

Given a polynomial $f \in \mathbb{Q}[x]$, with $x = (x_1, \ldots, x_n)$, we denote by $\mathrm{LT}_>(f)$ the leading term of f, *i.e.* the greatest term of f with respect to the order $>$. We denote by $\mathrm{LT}_>(I)$ the set of leading terms of elements of an ideal $I \subset \mathbb{Q}[x]$. A Gröbner basis of I with respect to $>$ is a set $G = \{g_1, \ldots, g_k\} \subset \mathbb{Q}[x]$ such that $\langle \mathrm{LT}_>(g_1), \ldots, \mathrm{LT}_>(g_k)\rangle = \langle \mathrm{LT}_>(I)\rangle$. One can suppose by definition that the elements of G are monic with respect to the ordering $>$, that is the coefficient of the leading term is 1. Moreover, if $g \in G$ is such that one of its terms can be generated by the leading terms of $G \setminus g$, then $G \setminus g$ is also a Gröbner basis of I [35, Ch. 2]. Any basis obtained by successively eliminating such elements from G, is called a reduced Gröbner basis.

Theorem 1.26 (Prop. 6, Ch 2, § 7 in [35])**.** *Every ideal $I \subset \mathbb{Q}[x]$, $I \neq \langle 0\rangle$, admits a unique reduced Gröbner basis.*

Reduced Gröbner bases can be computed via Buchberger's algorithm [29], which represents a generalization to polynomial systems of the Gaussian elimination algorithm for linear equations [153]. Implementations of this algorithm are available in many computer algebra systems, such as MAPLE [109] or MACAULAY2 [107]. Moreover, more efficient versions and speed-ups have been designed at the end of last century, in particular the algorithms F4 [41] and F5 [42], and those for the fast computation of change of monomial orderings [48] and [47]. These are implemented in the software FGB developed by J.-C. Faugère [43].

Gröbner bases can be used to solve in practice, among others, the following problems involving complex algebraic sets.

- Deciding the emptiness of an algebraic set. Indeed, it is sufficient to compute a reduced Gröbner basis of the polynomial ideal, remarking that the unique reduced Gröbner basis of $\langle 1\rangle$ is $\{1\}$.
- Computing the dimension of an algebraic set (*cf.* [35, Ch. 9,§. 3]).
- Verifying whether two polynomial systems define the same ideal. This is true if and only if they share the same reduced Gröbner basis.

We make frequently use of Gröbner bases computation in practice. Indeed, using Gröbner bases allows to meet the best practical performances for solving the aforementioned problems, mostly when dealing with structured polynomial systems (determinantal varieties, critical points varieties, boolean) as proved in recent results [45, 149, 120]. In particular, we use the new implementation of [50] for computing rational parametrizations.

1.2.2 Algorithms for computing rational parametrizations

In this section we report on exact algorithms for computing rational parametrizations of zero-dimensional algebraic sets.

The first one [53] computes a geometric resolution of the input, which can be used to obtain a rational parametrization. Let $f_1, \ldots, f_n, g \in \mathbb{Q}[x]$, with $x = (x_1, \ldots, x_n)$, be such that the zero set $\mathcal{Z} = Z_{\mathbb{C}}(\langle f_1, \ldots, f_n \rangle)$ defines a reduced regular sequence in the Zariski open set $g \neq 0$. In other words, for all $i = 1, \ldots, n$, for $\mathcal{Z}_i = Z_{\mathbb{C}}(\langle f_1, \ldots, f_i \rangle)$, the algebraic set $\overline{\mathcal{Z}_i \setminus Z_{\mathbb{C}}(g)}$ has dimension $n - i$ and the Jacobian matrix $D(f_1, \ldots, f_i)$ has full rank in $\mathcal{Z}_i \setminus Z_{\mathbb{C}}(g)$.

Denoting by δ_i the degree of $\mathcal{Z}_i$, and by $\delta = \max_i\{\delta_i\}$, the probabilistic algorithm in [53] computes a geometric resolution within $\tilde{\mathcal{O}}(n(nL + n^{\Omega})d^2\delta^2)$, with $\Omega \leq 4$, L the size of the straight-line program representing the input polynomials, and d a bound on their degree. Remark here that better bounds than the classical Bézout bound $\delta \leq d^n$ can be obtained in presence of structured input polynomials.

A second algorithm [88] also computes a geometric resolution, but uses subroutines based on symbolic homotopy techniques as in [74]. The complexity of this algorithm depends also on the combinatorial structure of input polynomials. Indeed, let $f \subset \mathbb{Q}[x]$ the input system with a finite number of solutions. Denote by N the sum of the cardinalities and by Q the maximum diameter of the supports of polynomials in f. Let e be the cardinality of $Z_{\mathbb{C}}(f)$. Let $\tilde{f}$ be a second system with the same support structure of f, and denote by e' the degree of the homotopy curve

$$Z_{\mathbb{C}}(tf + (1-t)\tilde{f}) \subset \mathbb{Q}[x,t],$$

The complexity of the algorithm in [88] is in $\mathcal{O}((n^2 N \log Q + n^{\omega+1})ee')$, where $\omega \leq \log 7$ is the exponent of matrix multiplication, *cf.* [151]. Also this second algorithm takes advantage of situations where the degrees e and e' are much lower that the singly exponential Bézout bound.

1.2.3 Solving over the reals

The theory of Gröbner bases is not sufficient to deal with decision problems over the real numbers. Indeed, the polynomial x^2+1 is a reduced Gröbner basis of the ideal it generates, while the real trace $Z_{\mathbb{R}}(\langle x^2+1 \rangle)$ is empty. Hence one cannot just rely on the computation of a Gröbner basis to establish the emptiness of real algebraic sets.

If the input algebraic set is zero-dimensional, *i.e.* finite, an intuitive definition of "solving the system over the reals" would be that of enumerating all the real solutions, obtaining a suitable description of the real set possibly via univariate representations (*cf.* Section 1.2.2). The following definition represents a natural generalization for positive-dimensional real algebraic sets.

Problem 1 (Real Root Finding Problem)**.** *Let $f \subset \mathbb{Q}[x]$ be a polynomial system encoding the algebraic set $\mathcal{Z} = Z_{\mathbb{C}}(f)$. The real root finding for $\mathcal{Z}$ is the problem of computing a finite set meeting each connected component of $\mathcal{Z} \cap \mathbb{R}^n$.*

An overview of the state of the art for exact algorithms solving Problem 1 has been given in the Introduction. Other possible exact approaches to real algebraic sets involve, *e.g.*, the construction of roadmaps, *i.e.* semi-algebraic subsets of dimension at most 1 (union of points and portions of curves) meeting each connected component of the input [139]. Among the main goals of this modern approach is indeed solving connectivity queries (namely, deciding whether two solution points lie on the same real component).

In [137] the authors designed an exact algorithm for solving the real root finding problem for algebraic sets satisfying genericity assumptions.

Theorem 1.27 (Th. 3 in [137])**.** *Let $f_1, \ldots, f_s \in \mathbb{Q}[x]$ of degree bounded by d, defining a smooth and equidimensional algebraic set $\mathcal{Z}$ of dimension e, and generating a radical ideal. Let L be the length of their straight-line program. There is a probabilistic algorithm computing a family of rational parametrizations, whose solutions meet every connected component of $\mathcal{Z} \cap \mathbb{R}^n$. The complexity is within $\mathcal{O}\tilde{}(Ln^{10}p(s+p)\mathsf{M}(d(n-e)\delta)^3)$, where $p = \binom{s}{n-e}\binom{n-1}{n-e}$, δ is an intrinsic degree quantity and $\mathsf{M}(t)$ is the number of operations needed to multiply two univariate polynomials of degree t.*

One gets from Theorem 1.27 that the real root finding problem can be solved within essentially $\mathcal{O}(d^{3n})$ arithmetic operations when the input algebraic set satisfies some regularity assumptions. Also, algorithms dealing with singular situations have been developed (see, *e.g.*, [140]).

It is finally worth to underline that solving Problem 1 would also solve, for the real set $\mathcal{Z} \cap \mathbb{R}^n$, the following second question, that we prefer to define over the more complete class of basic semi-algebraic sets.

Problem 2 (Emptiness Problem)**.** *Let $f_1, \ldots, f_s \in \mathbb{Q}[x]$, and let $S = \{x \in \mathbb{R}^n : f_i(x) \geq 0, i = 1, \ldots, s\}$ be the associated basic closed semi-algebraic set (cf. page 38). The emptiness problem for S is that of deciding whether S is empty or not, and in the negative case of computing a finite set meeting S.*

To solve Problem 2, one could think of choosing a dedicated polynomial criterion $f \in \mathbb{Q}[x]$, and to solve the associated optimization problem

$$\begin{aligned} \inf_{x \in \mathbb{R}^n} \; & f \\ s.t. \; & x \in S. \end{aligned}$$

For example, when f is the null polynomial (the zero element in $\mathbb{Q}[x]$), any element of S is a solution of the previous optimization program, provided that S is not empty. Exact algorithms for the polynomial optimization problem over real algebraic sets are available. We refer *e.g.* to Greuet's PhD thesis [57, 58] for the latest developments.

1.3 Optimization and convex algebraic geometry

1.3.1 Semidefinite programming

A symmetric matrix $M \in \mathbb{S}_m(\mathbb{R})$ defines a quadratic form over $\mathbb{R}^m$. The matrix M is called positive semidefinite (resp. definite) if the associated form is positive semidefinite (resp. definite), *i.e.* if $y'My \geq 0$ (resp. > 0) for all $y \neq 0$, where y' denotes the transpose of y.

A $m \times m$ symmetric matrix is positive semidefinite if and only if all its $2^m - 1$ principal minors are nonnegative, while a criterion for positive definiteness states that just all leading principal minors needs to be strictly positive (*cf.* [23, Sec. A.1.1]).

Spectrahedra

The set $\mathbb{S}_m^+(\mathbb{R})$ of positive semidefinite matrices is a convex cone in $\mathbb{S}_m(\mathbb{R})$, whose interior consists of positive definite matrices. An affine section of $\mathbb{S}_m^+(\mathbb{R})$, *i.e.* the intersection $\mathscr{S} = \mathcal{H} \cap \mathbb{S}_m^+(\mathbb{R})$ where $\mathcal{H}$ is an affine set, is also convex and it is called a spectrahedron. For every $\mathcal{H}$ there are $A_0, A_1, \ldots, A_n \in \mathbb{S}_m(\mathbb{R})$ such that

$$\mathscr{S} = \{x \in \mathbb{R}^n : A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n \succeq 0\}.$$

The formula $A(x) \succeq 0$ is called a linear matrix inequality (LMI). Spectrahedra generalize polyhedra and share them many properties, among which:

- they are convex basic semi-algebraic
- all their faces are exposed, that is they are defined as the intersection of a hyperplane with the spectrahedron itself

A substantial difference with polyhedra is that the number of linear inequalities defining a spectrahedron is infinite: indeed, one can write $\mathscr{S} = \{x \in \mathbb{R}^n : y'A(x)y \geq 0, \forall\, y \in \mathbb{R}^m\}$. Moreover, deciding whether a spectrahedron is in fact a polyhedron is NP-hard (*cf.* [17] and references therein). It is easy to check that a sufficient condition for representing $\mathscr{S}$ as a polyhedron is when matrices A_i commute, since in this case they are simultaneously diagonalizable.

Let $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$ be a symmetric pencil, and let $\mathscr{S}$ be the associated spectrahedron. The boundary B of $\mathscr{S}$ with respect to the Euclidean topology is a subset of the real determinantal hypersurface

$$\mathcal{D} = \{x \in \mathbb{R}^n : \det A(x) = 0\},$$

and in particular $\mathcal{D}$ is the closure of B in the Zariski topology (the smallest algebraic set containing B).

Optimizing over spectrahedra

While linear programming deals with linear optimization over polyhedra, its generalization to spectrahedra is called semidefinite programming. A typical semidefinite program is expressed in its primal and dual formulation, as:

$$\begin{aligned} p^* = \inf_{X \in \mathbb{S}_m} \quad & (C, X) \\ s.t. \quad & X \succeq 0 \\ & (A_j, X) = b_j, j = 1, \ldots, n \end{aligned}$$

$$\begin{aligned} d^* = \sup_{y \in \mathbb{R}^n} \quad & b'y \\ s.t. \quad & C - y_1 A_1 - \cdots - y_n A_n \succeq 0 \end{aligned}$$

where $C, A_j \in \mathbb{S}_m$ and $b \in \mathbb{R}^n$ are given matrices and vectors, and (C, X) denotes the standard scalar product, *i.e.* the trace of the matrix $C'X = CX$. While weak duality always holds ($p^* \geq d^*$) strong duality is related to the geometry of the primal and dual feasible sets. We say that a program is strictly feasible if the feasible set contains a positive definite matrix.

Theorem 1.28 (Lemma 1.3 in [102])**.** *If the primal program is strictly feasible, and the dual is feasible, then $p^* = d^*$ and the infimum is attained. If the primal program is feasible and the dual is strictly feasible, then $p^* = d^*$ and the supremum is attained.*

For an example with $p^* < d^*$ we refer, *e.g.*, to [23, Ex. 2.14]. From a computational viewpoint, semidefinite programming can be solved efficiently via numerical algorithms (*cf.* [22, Ch. 4]). Indeed one could use, *e.g.*, the ellipsoid algorithm to compute an approximate solution, and this procedure, when the accuracy $\epsilon > 0$ is fixed, can be performed in polynomial-time [59]. A numerical certificate is then produced with a complexity which depends polynomially on m, n and $\log(1/\epsilon)$.

As for linear programs, primal-dual interior-point algorithms have also been developed since the end of the 1980's for semidefinite programming. Their complexity is polynomial in the size of parameters of the input problem at any fixed desired accuracy, and it relies on the hypothesis that the feasible set $\mathscr{S}$ is not empty and that a feasible point $x^* \in \mathscr{S}$ is given. Hence, it is quite important to combine such methods with algorithms able to efficiently compute sample points on $\mathscr{S}$. Typically, interior-point methods produce maximal rank solution points, while finding low rank solutions is, conversely, a hard task. The reader can refer to [1] for an interesting discussion about unpredictability of the rank at optimal solutions, and next paragraph for further results. Numerical solvers dedicated to semidefinite programming, implementing primal-dual interior-point methods, have been developed, for example: SEDUMI [152] or MOSEK [113] or based on different approaches, such as LMILAB [105] or PENSDP [125]. The reader can refer to [159] for combinatorial conditions ensuring the existence of low rank solutions. Moreover, techniques based on projection methods for general conic optimization problems have been developed [79].

In [91], Khachiyan and Porkolab showed that the feasibility of a semidefinite program, that is, deciding the emptiness of a spectrahedron $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$, can be tested in $\mathcal{O}(nm^4) + m^{\mathcal{O}(\min\{n,m^2\})}$ operations over integer numbers of bit-size in $\ell m^{\mathcal{O}(\min\{n,m^2\})}$, where ℓ bounds the bit-size of the input integer entries of $A(x)$. Remark that when the size of the matrix m is fixed, the result of Khachiyan and Porkolab gives a complexity which is linear in the number of variables. Moreover, when $\mathscr{S}$ is not empty, bounds on the bit-size of the feasible solutions are given (precisely in $\ln^{\mathcal{O}(\min\{n,m^2\})}$). On the other hand, this algorithm strongly relies on Renegar's quantifier elimination [133], and no explicit estimates on the exponents are given; hence it does not lead to efficient implementations in practice.

The rank at an optimal solution

Since spectrahedra are convex sets and their faces are exposed, the solution of a semidefinite program coincides with an entire face of $\mathscr{S}$, and any solution is encoded by a singular matrix.

Under genericity assumptions on the input parameters $C, A_j \in \mathbb{S}_m$ and $b \in \mathbb{R}^n$ of the aforementioned primal-dual semidefinite programs, the optimal matrices X and $A(y) = C - y_1A_1 - \cdots - y_nA_n$ are positive semidefinite and satisfy the complementary condition $A(y)X = 0$ (*cf.* [117, Th. 3]).

Proposition 1.29 (Prop. 5 in [117]). *Suppose that input data $C, A_j \in \mathbb{S}_m$ and $b \in \mathbb{R}^n$ are generic. Let r and $m - r$ be the ranks of the optimal matrices $A(y)$ and X. Then these values range over the Pataki's inequality*

$$\binom{m-r+1}{2} \leq n \qquad \binom{r+1}{2} \leq \binom{m+1}{2} - n.$$

Hence, the geometry of spectrahedra and that of the loci of rank defects associated to the linear pencil are strictly related. Indeed, the optimization over $\mathscr{S}$ can be also performed by minimizing the same linear function over the set of matrices of rank at most r in the pencil, where r is the expected rank. In this thesis, we follow this idea by reducing the LMI constraint $A(x) \succeq 0$ to a sequence of optimization problems over the loci of rank defects of $A(x)$.

If $\mathscr{S} = \{x \in \mathbb{R}^n : A_0 + x_1A_1 + \cdots + x_nA_n \succeq 0\}$ is the spectrahedron, with rational input data matrices $A_0, A_1, \ldots, A_n$, the solution $A(x^*)$ is a symmetric matrix whose entries are algebraic numbers, but typically not rational. Upper bounds for their algebraic degree, giving an estimate of the complexity of solving exactly the program itself, have been established in [117].

Concerning the rank at a given solution of a semidefinite program, the next result implies that the minimizers with maximal rank correspond to points in the relative interior of the optimal face.

Theorem 1.30 (Lemma 1.4 in [102]). *Let $A(x)$ be a symmetric pencil, and let $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$ be the associated spectrahedron. Then:*

- *If x^* is such that $\operatorname{rank} A(x^*) = \max_{x \in \mathscr{S}} \operatorname{rank} A(x)$, then $\ker A(x^*) \subset \ker A(x)$ for all $x \in \mathscr{S}$.*
- *Let $c'x$ be a linear function, let $\mathcal{F} \subset \mathscr{S}$ be the face of minimizers to the associated semidefinite program, and let $x^* \in \mathcal{F}$ be such that $\operatorname{rank} A(x^*) = \max_{x \in \mathcal{F}} \operatorname{rank} A(x)$. Then $\ker A(x^*) \subset \ker A(x)$ for all $x \in \mathcal{F}$.*

Example 1.31. *Consider the symmetric pencil*

$$A(x) = \begin{pmatrix} 1 - x_1 & x_2 & 0 \\ x_2 & 1 + x_1 & 0 \\ 0 & 0 & -x_2 \end{pmatrix}.$$

Its associated spectrahedron $\mathscr{S} = \{x \in \mathbb{R}^2 : A(x) \succeq 0\} = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 - 1 \leq 0, x_2 \leq 0\}$ is a half-disk. Its algebraic boundary is shaped by the cubic hypersurface

$$\det A(x) = -x_2(1 - x_1^2 - x_2^2).$$

We remark here that the boundary of $\mathscr{S}$ contains the two real singular points of the determinantal variety, i.e. $(-1, 0)'$ and $(1, 0)'$.

Let e_1, e_2, e_3 be the standard Euclidean basis of $\mathbb{R}^3$. The matrix A has rank 1 at these two points and $\ker A(-1, 0)$ contains the vector space $\langle e_2, e_3\rangle$, while $\ker A(1, 0)$ contains $\langle e_1, e_3\rangle$. Consider the closed line segment $\mathcal{F} = [-1, 1] \times \{0\} \subset \mathscr{S}$, which is the optimal face when maximizing the linear function $f(x_1, x_2) = x_2$. One can easily check that, consistently with Theorem 1.30, the kernel of $A(t, 0)$ for $t \in (-1, 1)$ is constant and equal to $\langle e_3\rangle = \ker A(-1, 0) \cap \ker A(1, 0)$. Any point in the relative interior of $\mathcal{F}$ generate rank 2 semidefinite matrices.

Following the central path generated in interior-point methods for maximizing x_2 on $\mathscr{S}$ gives as output an approximation of an element of the interior of $\mathcal{F}$. Point $(-1, 0)$ (and symmetrically $(1, 0)$) is special since it maximizes an infinite family of semidefinite programs, parametrized by a convex cone.

Spectrahedral shadows

Another difference between spectrahedra and polyhedra is that spectrahedra are not closed under projection. Also, the dual body of a spectrahedron is not, in general, a spectrahedron, but it can be represented as the projection of a spectrahedron, namely a spectrahedral shadow (*cf.* [131] or [23, Ch. 5]).

A spectrahedral shadow is represented as follows:

$$S = \left\{x \in \mathbb{R}^n : \exists\, y \in \mathbb{R}^d, A(x, y) \succeq 0\right\},$$

for some linear pencil A in two groups of variables x, y. The variables $y \in \mathbb{R}^d$ are called lifting or slack variables. By Tarski-Seidenberg theorem, S is also semi-algebraic, but in general not basic. Recent results characterize the boundary structure of spectrahedral shadows [148]. A central conjecture in convex algebraic geometry asks whether every convex semi-algebraic set is the projection of a spectrahedron. Scheiderer proved that this conjecture holds for planar convex semi-algebraic sets [143], *i.e.* for $n = 2$, even though there are no explicit constructions of the pencil A and no bound on the number d of lifting variables.

Solving this conjecture constructively or, in general, being able to express a set S as a projection of a spectrahedron $\mathscr{S}$ reduces, in particular, the problem of optimizing linear functions over S to a semidefinite program. The main drawback of this reduction is that one does not have, in general, reasonable bounds on the number of lifting variables y to be added (or on the size of the matrix A) to get the semidefinite representation.

1.3.2 Applications

Semidefinite programming models a large class of optimization or decision queries. We give in this section a gallery of examples and motivation to develop an effective theory for semidefinite optimization. In general, we deal with computationally hard instances (NP-hard, NP-complete ...) and the goal is to build relaxations that can be solved efficiently.

Nonnegative polynomials and sums of squares

Deciding nonnegativity of a multivariate polynomial $f \in \mathbb{R}[x]$ of degree greater than 4, *i.e.* proving that $f \geq 0$ everywhere in $\mathbb{R}^n$, is NP-hard. Consequently, it is of tremendous importance to write down a proof or, more precisely, a certificate establishing nonnegativity. Artin's positive answer to Hilbert's 17th problem implies that one can write a certificate for f as a sum of squares of rational functions.

Theorem 1.32 (Artin, 1927)**.** *A polynomial $f \in \mathbb{R}[x]$ is nonnegative in $\mathbb{R}^n$ if and only if there exist $g, h_i \in \mathbb{R}[x]$ such that $fg^2 = \sum h_i^2$.*

Since the polynomial g in the previous theorem is the g.c.d. of the denominators of the rational functions, its degree in this representation can be high and hence one would ask to obtain a certificate with $g = 1$. Recently, Lombardi, Perrucci and Roy announced a uniform bound for the degree of polynomials involved in Hilbert's representation, namely a tower of five exponentials in the degree and number of variables of input polynomials [106].

Hilbert proved in 1888 (*cf.* [83]) that every nonnegative polynomial $f \in \mathbb{R}[x]$, of degree $2d$, admits a decomposition as a sum of squares of polynomials, if and only if either $n = 1$, or $d = 1$, or $(n, 2d) = (2, 4)$. The effective counterpart of this result tells that such a decomposition can be computed solving a semidefinite program. Indeed, suppose that f has degree $2d$ and let v be the vector of length

$\binom{n+d}{d}$ containing all monomials of degree at most d. Then a sum-of-squares decomposition for f can be computed by finding a matrix $X \in \mathbb{S}_{\binom{n+d}{d}}$ such that

$$\begin{aligned} X &\succeq 0 \\ f &= v' X v \end{aligned}$$

cf. for example [134, 121, 128, 100, 99, 116].

This boils down to deciding the emptiness of a spectrahedron, and the algorithm developed in Chapter 4 of this manuscript gives an answer by means of an exact approach. The main goal of designing exact algorithms for linear matrix inequalities and more general semidefinite programs, is to obtain, at least for inputs of modest size, exact multi-precision algebraic representations of the solutions, and a certificate for the emptiness of the associated feasible sets that cannot be provided by numerical algorithms.

Polynomial optimization

The general polynomial optimization problem asks to minimize a polynomial function $f \in \mathbb{Q}[x]$ over a basic semi-algebraic set:

$$\begin{aligned} f^* = \inf_{x \in \mathbb{R}^n} \; & f \\ s.t. \; & f_i(x) \geq 0, \, i = 1, \ldots, s. \end{aligned}$$

The set $S = \{x \in \mathbb{R}^n : f_i(x) \geq 0, \, i = 1, \ldots, s\}$ is, in general, non-convex and unbounded. In any case, one can rewrite it as a positivity query for multivariate polynomials:

$$\begin{aligned} f^* = \sup_{\lambda \in \mathbb{R}} \; & \lambda \\ s.t. \; & f - \lambda \geq 0 \;\text{ on } S. \end{aligned}$$

Results as Schmüdgen's or Putinar's Positivstellensatz allow to obtain certificates for polynomials that are nonnegative over semi-algebraic sets. Indeed, given $f_1, \ldots, f_s \in \mathbb{R}[x]$, the set

$$M(f_1, \ldots, f_s) = \left\{ g_0 + \sum_{j=1}^{s} g_j f_j \, : \, g_i \text{ is a sum of squares }, i = 1, \ldots, s \right\}$$

is called the quadratic module of $f_1, \ldots, f_s$, while the set

$$P(f_1, \ldots, f_s) = \left\{ \sum_{\delta \in \{0,1\}^s} g_\delta \prod_i f_i^{\delta_i} \, : \, g_\delta \text{ is a sum of squares }, \delta \in \{0,1\}^s \right\}$$

is called the preordering of $f_1, \ldots, f_s$.

Theorem 1.33 (Schmüdgen Positivstellensatz, [146]). *If the semi-algebraic set $S = \{x \in \mathbb{R}^n : f_1(x) \geq 0, \ldots, f_s(x) \geq 0\}$ is compact, then for $f \in \mathbb{R}[x]$, $f > 0$ on S if and only if $f \in P(f_1, \ldots, f_s)$.*

Under compactness assumptions, Schmüdgen's Positivstellensatz builds a certificate of positivity for f on S. Putinar's version also assume a generic condition (which is, in general, slightly stronger than the compactness of S) on the polynomials f_i.

Theorem 1.34 (Putinar Positivstellensatz, [129]). *Suppose that the quadratic module $M(f_1, \ldots, f_s)$ is archimedean, that is there exists $N \in \mathbb{N}$ such that $N - \sum_i x_i^2 \in M(f_1, \ldots, f_s)$. Then for $f \in \mathbb{R}[x]$, $f > 0$ on S if and only if $f \in M(f_1, \ldots, f_s)$.*

Both Schmüdgen and Putinar's theorem can be used effectively by fixing an upper bound on the degree of the sum-of-squares multipliers, and solving the associated semidefinite program. While Putinar's Positivstellensatz admits at most $s+1$ sum-of-squares multipliers in the representation of f, in Schmüdgen's version one could have a singly exponential number of multipliers with respect to the number s of polynomials defining the semi-algebraic set. Recently, Nie and Schweighofer computed upper bounds on the degree of sum-of-squares multipliers for Putinar Positivstellensatz [119], and Schweighofer for Schmüdgen Positivstellensatz [147].

When S is a real algebraic set, dedicated methods have been developed. The algorithms in [57], based on dimensional reduction and on the construction of polar varieties, allow to solve this optimization problem in exact arithmetic with a complexity in $\widetilde{\mathcal{O}}(D^6(\sqrt[3]{2}(s+1)(D-1))^{3n})$, where D bounds the degrees of all polynomials. For general semi-algebraic sets, one can use the approach via border bases in [4]. Other methods use resultants, homotopy continuation or linear algebra over the gradient ideal $\langle \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \rangle$ of f (*cf.* [122] for a general overview).

Lasserre established in [100] that such general optimization problems can be reduced, when S is compact, to a hierarchy of convex optimization problems modeled by linear matrix inequalities, that is to semidefinite programs. Under mild genericity assumptions on input data, the sequence of minimizers converges to the optimal solution after a finite number of iterations [118]. We refer to [99] for an exhaustive description of the duality theory of nonnegative polynomials and moments, on which Lasserre's method strongly relies. Based on this and further developments, the software GLOPTIPOLY [73] implements the method of moments and is able to tackle problems of remarkably high size.

We conclude by remarking that, since with semidefinite programming one can compute positivity certificates for multivariate polynomials over semi-algebraic sets, an exact algorithm solving SDP would lead to exact certificates and solutions to the polynomial optimization problem.

Control theory

For the use of linear matrix inequalities in control theory we refer, in general, to the historical survey [25], and to [78] for a modern viewpoint.

We recall, as example, that among the first LMI appearing in mathematics is the Lyapunov condition for the convergence of the solutions of the ordinary differential equation $\dot{x} = M\,x$, which reads

$$\begin{aligned} P &\succ 0 \\ M'P + PM &\prec 0, \end{aligned}$$

with P a matrix of variables. If such a P exists, a Lyapunov quadratic function certifying the asymptotic stability of the origin can be built directly, by Lyapunov theorem.

Other problems of interest in systems control can be modeled as LMI, such as the generalized eigenvalue problem. Finally, LMI and semidefinite programming arise naturally when modeling the analysis or synthesis of nonlinear dynamical systems (see *e.g.* [155]), or in nonlinear optimal control with polynomial data [75, 32].

Feasibility certificates for semi-algebraic sets

In this thesis we are interested in obtaining proofs in exact arithmetic that some real algebraic or semi-algebraic sets is empty or not.

Results such as Real or Complex Nullstellensatz for algebraic sets (*cf.* Theorem 1.4 and 1.5), or Farkas Lemma for linear programming, or the Positivstellesatz for semi-algebraic sets (*cf.* Theorem 1.33 and 1.34), represent dedicated algebraic certificates for the feasibility of semi-algebraic sets (*cf.* [3, 121]). We finally give below a general equivalent form of the Positivstellensatz underlining this aspect.

Theorem 1.35 (Positivstellensatz certificate, Th. 3.15 in [102])**.** *Let $\{f_i\}_{1,\ldots,r}$, $\{g_j\}_{1,\ldots,s}$ and $\{h_\ell\}_{1,\ldots,t} \subset \mathbb{R}[x]$, and let $S = \{x \in \mathbb{R}^n : f_i(x) \neq 0, g_j(x) \geq 0, h_\ell(x) = 0,\ i = 1,\ldots,r,\ j = 1,\ldots,s,\ \ell = 1,\ldots,t\}$. Then $S = \emptyset$ if and only if*

$$\prod_{i=1}^{r} f_i^{2d_i} + \sum_{\delta \in \{0,1\}^s} \sigma_\delta \prod_{j=1}^{s} g_j^{\delta_J} + \sum_{\ell=1}^{t} \tau_\ell h_\ell = 0$$

for some $d_i \in \mathbb{N}$, σ_δ sum of squares of polynomials in $\mathbb{R}[x]$, and $\tau_\ell \in \mathbb{R}[x]$.

Recently, Klep and Schweighofer have proved the following certificate for the emptiness of spectrahedra. For a symmetric pencil $A(x)$ and its spectrahedron $\mathscr{S}$, we define the associated quadratic module

$$M_A = \left\{ \sigma + \sum_j v_j' A v_j \,:\, \sigma \text{ is a sum of squares and } v_j \in \mathbb{R}^m \right\}.$$

Theorem 1.36 (Th. 2.2.5 in [96])**.** *The spectrahedron* $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$ *is empty if and only if* $-1 \in M_A$*.*

Part II

Contributions

Chapter 2

Projections

The correspondence between algebraic and geometric properties of real or complex sets lies at the foundation of algebraic geometry. Moreover, such properties are often exploited to prove the correctness of exact algorithms. Especially for probabilistic algorithms, whose correctness typically depends on the choice of some parameters (such as linear changes of variables), the goal is to prove that algebraic properties (such as smoothness of algebraic sets or closure of polynomial maps) hold generically with respect to this choice. Within this framework, we consider the relation between Noether position and properness of the restriction of linear projections to algebraic sets.

This chapter contains a generalization to the non-smooth and non-equidimensional case of the algebraic properties proved in [137]. The results that we present appeared in the work [68], accepted for publication in Journal of Symbolic Computation, 2015.

2.1 Introduction

The results of this chapter follow and build on more than a decade of efforts in establishing connections between Noether position of complex algebraic sets and properness properties of linear projections restricted to their real traces. Indeed, this has already been subject of investigation in several works [137, 138, 53], since proving that an algebraic set is in Noether position represents a useful tool for real solving in computational algebraic geometry.

We establish in our main result, Theorem 2.1, that any algebraic set $\mathcal{Z}$ can be decomposed as a suitable union of algebraic subsets $\mathcal{O}_i(\mathcal{Z})$ such that a generic change of variables allows to simultaneously put in Noether position the set itself, together with any element $\mathcal{O}_i(\mathcal{Z})$. Moreover, since the image of any set $\mathcal{O}_i(\mathcal{Z})$ via the canonical projection $(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_{i+1})$ contains the critical values of the restriction of the map itself to $\mathcal{Z}$, this implies that the image of any connected component $\mathcal{C} \subset \mathcal{Z} \cap \mathbb{R}^n$ by this map is closed.

Throughout the chapter, we denote by $\mathcal{Z} \subset \mathbb{C}^n$ an algebraic set, and we let d be its dimension. If $A \subset B$ is a ring extension, we recall that $b \in B$ is integral over A if it is the root of a monic polynomial $y^k + a_{k-1}y^{k-1} + \cdots + a_1 y + a_0$ where $a_i \in A$, for $i = 0, \ldots, k-1$. The extension $A \subset B$ is integral if all elements of B are integral over A.

We recall that $\mathcal{Z}$, or equivalently its defining ideal $I(\mathcal{Z}) \subset \mathbb{C}[x]$, with $x = (x_1, \ldots, x_n)$, is in Noether position with respect to $x_1, \ldots, x_d$, if the ring extension

$$\mathbb{C}[x_1, \ldots, x_d] \subset \mathbb{C}[x_1, \ldots, x_n] \big/ I(\mathcal{Z})$$

is integral, that is if every element of $\mathbb{C}[x_1, \ldots, x_n] \big/ I(\mathcal{Z})$ is integral over $\mathbb{C}[x_1, \ldots, x_d]$. We also recall (*cf.* Definition 1.23, page 43) that a continuous map $f \colon \mathcal{X} \to \mathcal{Y}$ between topological spaces $\mathcal{X}, \mathcal{Y}$, is proper at $y \in \mathcal{Y}$ if there is an open neighborhood $\mathcal{O} \subset \mathcal{Y}$ of y such that $f^{-1}(\overline{\mathcal{O}})$ is compact in $\mathcal{X}$, where $\overline{\mathcal{O}}$ is the closure of $\mathcal{O}$ in the topology of $\mathcal{Y}$. A proper map $f \colon \mathcal{X} \to \mathcal{Y}$ is also a closed map [36].

2.1.1 Prior work

The relation between Noether position of algebraic sets and properness and closure properties of algebraic maps, and, in particular, the idea of exploiting such a relation for algorithms in real algebraic geometry, has been first introduced and discussed by Safey El Din and Schost in [137] and further in [138].

The algorithm in [137] deals with polynomial systems solving over the real numbers. Under the following hypotheses:

- $f_1, \ldots, f_s \in \mathbb{Q}[x]$ generate a radical ideal $I = \langle f_1, \ldots, f_s \rangle$, and
- $\mathcal{V} = Z_{\mathbb{C}}(I)$ is smooth and equidimensional,

the authors prove that computing sections of polar varieties of generic projections of $\mathcal{V}$ is sufficient to compute at least one point per connected component of $\mathcal{V} \cap \mathbb{R}^n$. Remarkably, this algorithm does not need compactness assumptions.

To prove correctness of the algorithm in [137], a properness property for all polar varieties of $\mathcal{V}$ up to a generic linear change of variable is proved. Denoting by $\mathcal{W}_i$ the $i-$th polar variety of $\mathcal{V}$ (*cf.* Section 1.1.3) we recall the statement of Theorem 1.25 on page 44:

Suppose that $\mathcal{V}$ is smooth and equidimensional, and if its defining polynomial system generates a radical ideal. There exists a Zariski open set $\mathscr{O} \subset \mathrm{GL}_n(\mathbb{C})$ such that, if $M \in \mathscr{O} \cap \mathbb{M}_{n,n}(\mathbb{Q})$, and $i = 1, \ldots, d+1$, the restriction of the projection π_{i-1} to $M^{-1}\mathcal{W}_i$ is proper.

The key idea to prove the previous Theorem relies on the notion of Noether position, and essentially on [137, Prop. 1]. In this chapter we give a strict generalization of this result for algebraic sets without any regularity assumption.

2.1.2 Notation

Singular-critical decomposition of algebraic sets

In this section we define a decomposition of recursive type of general algebraic sets with respect to their equidimensional components, singular locus and critical loci of linear projections. Such a decomposition is admissible without any special assumption and it has to be understood as a variant of the classical flag of polar varieties that can be defined over the input algebraic set.

Let $\mathscr{Z}$ be an algebraic set of dimension d. For $i \in \{0, \dots, d\}$, in Section 1.1.1 we have denoted by $\Omega_i(\mathscr{Z})$ the equidimensional component of $\mathscr{Z}$ of dimension i, that is the union of its irreducible components of dimension i. We now define two operators acting on algebraic sets.

Singular Decomposition Operator. Let

$$\mathscr{S} : \mathscr{Z} \to \mathscr{S}(\mathscr{Z}),$$

be the operator sending a given algebraic set $\mathscr{Z} \subset \mathbb{C}^n$ to $\mathscr{S}(\mathscr{Z})$, defined as the union of the algebraic sets:

- $\Omega_0(\mathscr{Z}) \cup \cdots \cup \Omega_{d-1}(\mathscr{Z})$
- $\operatorname{sing}(\Omega_d(\mathscr{Z}))$.

Critical Decomposition Operator. Furthermore, for $1 \leq i \leq n$, let π_i be the canonical projection $(x_1, \dots, x_n) \mapsto (x_1, \dots, x_i)$ (and for $i = 0$ let π_0 be the constant map π_0 whose codomain is the singleton $\{\bullet\}$). We denote by

$$\mathscr{C} : (\pi_i, \mathscr{Z}) \to \mathscr{C}(\pi_i, \mathscr{Z}),$$

the map sending $(\pi_i, \mathscr{Z})$ to the Zariski closure of the union of the algebraic sets:

- $\Omega_0(\mathscr{Z}) \cup \cdots \cup \Omega_{i-1}(\mathscr{Z})$
- the union, for $j \geq i$, of $\operatorname{crit}(\pi_i, \operatorname{reg}(\Omega_j(\mathscr{Z})))$.

Recursive Singular-Critical Decomposition. We define now a recursive collection of algebraic subsets of a fixed $\mathscr{Z} \subset \mathbb{C}^n$, of dimension d. For any $M \in \mathrm{GL}_n(\mathbb{C})$, the collection $\{\mathcal{O}_i(M^{-1}\mathscr{Z}) : i \in \{0, \dots, d\}\}$ of algebraic sets is defined as follows:

- $\mathcal{O}_d(M^{-1}\mathscr{Z}) = M^{-1}\mathscr{Z}$
- $\mathcal{O}_i(M^{-1}\mathscr{Z}) = \mathscr{S}(\mathcal{O}_{i+1}(M^{-1}\mathscr{Z})) \cup \mathscr{C}(\pi_{i+1}, \mathcal{O}_{i+1}(M^{-1}\mathscr{Z})) \cup \mathscr{C}(\pi_{i+1}, M^{-1}\mathscr{Z})$, for $i \in \{0, \dots, d-1\}$.

For $M = \mathbb{I}_n$ we denote $\mathcal{O}_i(\mathscr{Z})$ by $\mathcal{O}_i$. For $i \in \{0, \dots, d\}$, we denote $I(\mathcal{O}_i)$ by I_i and $I(\mathcal{O}_i(M^{-1}\mathscr{Z}))$ by $I_i \circ M$ when the choice of $\mathscr{Z}$ is clear.

Symbolic change of variables

Let B be a $n \times n$ matrix of unknowns. We denote by $\mathbb{Q}(B) \supset \mathbb{Q}$ the field extension of $\mathbb{Q}$ generated by the entries of B. For $f \in \mathbb{Q}[x]$, $x = (x_1, \ldots, x_n)$, we denote by $f \circ B \in \mathbb{Q}(B)[x]$ the polynomial such that $(f \circ B)(x) = f(B\,x)$ for all $x \in \mathbb{C}^n$.

Let $I = \langle f_1, \ldots, f_s \rangle \subset \mathbb{Q}[x]$ be any ideal, and let $\mathcal{Z} = Z_{\mathbb{C}}(I)$. We denote the ideal $\langle f_1 \circ B, \ldots, f_s \circ B \rangle \subset \mathbb{Q}(B)[x]$ by $I \circ B$ and $Z_{\mathbb{C}}(I \circ B)$ by $B^{-1}\mathcal{Z}$, which is a subset of $\mathbb{C}^{n+n^2}$. Finally, for $i \in \{0, \ldots, d = \dim \mathcal{Z}\}$, we denote by $I_i \circ B = I(\mathcal{O}_i(B^{-1}\mathcal{Z}))$, when there is no ambiguity on $\mathcal{Z}$.

2.1.3 Main result and chapter outline

Main result

The main result of this chapter establishes properness properties of the restrictions of linear projections to the algebraic sets $\mathcal{O}_i$ introduced in the previous section. It is encoded by the following main theorem. Its proof can be obtained by merging the proof of Propositions 2.2, 2.4 and 2.5. It states that, up to generic changes of variables, the dimension of sets $\mathcal{O}_i$ is at most i, that they are in Noether position with respect to $x_1, \ldots, x_i$, and to conclude that the images of any connected component of $\mathcal{Z} \cap \mathbb{R}^n$ via the maps π_i are closed sets.

Theorem 2.1 (Main Theorem 1)**.** *Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d. There exists a non-empty Zariski open set $\mathscr{M} \subset \mathrm{GL}_n(\mathbb{C})$ such that, if $M \in \mathscr{M} \cap \mathbb{M}_{n,n}(\mathbb{Q})$, the following holds:*

1. *for $i = 0, \ldots, d$, the set $\mathcal{O}_i(M^{-1}\mathcal{Z})$ has dimension $\leq i$ and is in Noether position*
2. *for any connected component $\mathcal{C} \subset \mathcal{Z} \cap \mathbb{R}^n$, for $i = 0, \ldots, d$ and for $w \in \mathbb{R}^i$ in the boundary of $\pi_i(M^{-1}\mathcal{C})$, the set $\pi_i^{-1}(w) \cap M^{-1}\mathcal{C}$ is non-empty and contained in $\mathcal{O}_{i-1}(M^{-1}\mathcal{Z}) \cap M^{-1}\mathcal{C}$.*

Outline

Chapter 2 is organized as follows. In Section 2.2 we define, first, two algebraic properties. The first involves the set $\mathcal{O}_i$ introduced above, and states that it is in Noether position with respect to $x_1, \ldots, x_i$. The second is defined for a given algebraic set $\mathcal{Z} \subset \mathbb{C}^n$, and states that the critical values (*cf.* Definition 1.10, page 38) of the restriction of the projection π_i to any connected component of $\mathcal{Z} \cap \mathbb{R}^n$ is contained on the image of the algebraic set $\mathcal{O}_{i-1}$ by the map itself.

We prove the logical dependency of these properties and that both hold generically. This represents the first result towards Theorem 2.1, and its proof is in Section 2.2.1. The second result concerns Assumption 2 in Theorem 2.1 and is proved in Section 2.2.2. The proof of an intermediate result is finally given in Section 2.3.

2.2 Projections and Noether position

2.2.1 Properties

We define two algebraic properties and prove that both hold generically. These involve the Singular-Critical Decomposition of the algebraic set $\mathcal{Z}$ introduced previously in Section 2.1.2. A rigorous definition of algebraic property can be found in Definition 1.15, on page 40.

The first property depends on a fixed algebraic set, and its domain is the locally closed set $\mathrm{GL}_n(\mathbb{C})$ of non-singular $n \times n$ matrices. Let $\mathcal{Z}$ be any algebraic set of dimension d:

Property $\mathsf{P}(\mathcal{Z})$

Let $M \in \mathrm{GL}_n(\mathbb{C})$. We say that M satisfies $\mathsf{P}(\mathcal{Z})$ when, for all $i = 0, \ldots, d$,

1. *$\mathcal{O}_i(M^{-1}\mathcal{Z})$ has dimension $\leq i$, and*
2. *$\mathcal{O}_i(M^{-1}\mathcal{Z})$ is in Noether position with respect to $x_1, \ldots, x_i$.*

The second property is defined over the Cartesian product of $\mathrm{GL}_n(\mathbb{C})$ with the collection of algebraic subsets of $\mathbb{C}^n$ of dimension d. It is defined as the logical conjunction of d properties $\mathsf{Q} = \mathsf{Q}_1 \wedge \cdots \wedge \mathsf{Q}_d$:

Property Q

Let $M \in \mathrm{GL}_n(\mathbb{C})$. Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d, and let $i = 1, \ldots, d$. We say that $M^{-1}\mathcal{Z}$ satisfies Q_i when for any connected component $\mathcal{C} \subset \mathcal{Z} \cap \mathbb{R}^n$ the boundary of $\pi_i(M^{-1}\mathcal{C})$ is contained in $\pi_i(\mathcal{O}_{i-1}(M^{-1}\mathcal{Z}) \cap M^{-1}\mathcal{C})$. We say that $M^{-1}\mathcal{Z}$ satisfies Q if $M^{-1}\mathcal{Z}$ satisfies Q_i for all $i = 1, \ldots, d$.

We prove a first result which relates properties $\mathsf{P}(\mathcal{Z})$ and Q. Formally, it states that the following logical implication is true:

$$\forall \mathcal{Z} \subset \mathbb{C}^n \text{ algebraic set}, \quad \forall M \in \mathrm{GL}_n(\mathbb{C}) \quad (\mathsf{P}(\mathcal{Z}) \Longrightarrow \mathsf{Q})\,.$$

Proposition 2.2. *Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d, and let $M \in \mathrm{GL}_n(\mathbb{C})$. If M satisfies $\mathsf{P}(\mathcal{Z})$ then $M^{-1}\mathcal{Z}$ satisfies Q.*

Proof. To keep notations simple, suppose without loss of generality that M is the identity matrix $\mathbb{I}_n$. We also define the following integer-valued function over $\mathcal{Z}$: for $y \in \mathcal{Z}$, let

$$J(y) = \min\{j : y \in \mathcal{O}_j\}.$$

We proceed by decreasing induction on the index i. First, we claim that $\mathcal{Z}$ satisfies Q_d. Indeed, let $\mathcal{C} \subset \mathcal{Z} \cap \mathbb{R}^n$ be a connected component, and let $x \in \mathbb{R}^d$ belong to the boundary of $\pi_d(\mathcal{C})$. By [87, Lem. 3.10], $\mathsf{P}(\mathcal{Z})$ implies that the restriction of π_d to $\mathcal{O}_d$ is proper, hence closed. We deduce that the restriction of π_d to $\mathcal{O}_d \cap \mathcal{C}$ is closed, and that $x \in \pi_d(\mathcal{O}_d \cap \mathcal{C})$. Let $y \in \mathcal{O}_d \cap \mathcal{C}$ such that $\pi_d(y) = x$. If $J(y) \leq d-1$, the claim follows straightforwardly. Otherwise, suppose $J(y) = d$. We deduce that $y \in \operatorname{reg} \Omega_d(\mathcal{Z})$ (the set of regular points of $\Omega_d(\mathcal{Z})$, introduced on page 37). By the Implicit Function Theorem [150, Th. 2.12] we conclude that $y \in \operatorname{crit}(\pi_d, \operatorname{reg} \Omega_d(\mathcal{Z})) \subset \mathscr{C}(\pi_d, \mathcal{Z}) \subset \mathcal{O}_{d-1}$, which is a contradiction.

Suppose now that $\mathcal{Z}$ satisfies Q_{i+1}. We proceed in two steps.

First step. We claim that the boundary of $\pi_i(\mathcal{C})$ is included in $\pi_i(\mathcal{O}_i \cap \mathcal{C})$. Indeed, let $x \in \mathbb{R}^i$ belong to the boundary of $\pi_i(\mathcal{C})$. Let $p \colon \mathbb{R}^{i+1} \to \mathbb{R}^i$ be the map sending $(x_1, \dots x_{i+1})$ to $(x_1, \dots x_i)$. For $\delta > 0$, let $B_\delta \subset \mathbb{R}^i$ be the ball of center x and radius δ, and let $B'_\delta = p^{-1}(B_\delta) \subset \mathbb{R}^{i+1}$. We claim that B'_δ meets both $\pi_{i+1}(\mathcal{C})$ and its complementary in $\mathbb{R}^{i+1}$. Indeed this is a consequence of the following immediate equalities

$$\pi_i^{-1}(B_\delta) \cap \mathcal{C} = \pi_{i+1}^{-1} \circ p^{-1}(B_\delta) \cap \mathcal{C} = \pi_{i+1}^{-1}(B'_\delta) \cap \mathcal{C}$$

and $\pi_i^{-1}(B_\delta) \cap \mathcal{C} \neq \emptyset$ and $B_\delta \cap (\mathbb{R}^i \setminus \pi_i(\mathcal{C})) \neq \emptyset$. Since B'_δ is connected, B'_δ meets also the boundary of $\pi_{i+1}(\mathcal{C})$. Since $\mathcal{Z}$ satisfies Q_{i+1}, for every $\delta > 0$ there exists $y_\delta \in \mathcal{O}_i \cap \mathcal{C}$ such that $\pi_{i+1}(y_\delta) \in B'_\delta$, and so $\pi_i(y_\delta) \in B_\delta$. Thus, x lies in the closure of $\pi_i(\mathcal{O}_i \cap \mathcal{C})$. This image is closed and our claim follows.

Second step. We prove that $\mathcal{Z}$ satisfies Q_i. Let $x \in \mathbb{R}^i$ belong to the boundary of $\pi_i(\mathcal{C})$. From the previous point, we deduce that there exists $y \in \mathcal{O}_i \cap \mathcal{C}$ such that $\pi_i(y) = x$. Suppose that for all such y, $J(y) = i$. Fix $y \in \mathcal{O}_i \setminus \mathcal{O}_{i-1}$ such that $\pi_i(y) = x$. In particular, $y \in \mathcal{O}_i \setminus \mathscr{S}(\mathcal{O}_i)$, and thus, we deduce that $y \in \operatorname{reg} \Omega_i(\mathcal{O}_i)$. Next, since $x \in \pi_i(\Omega_i(\mathcal{O}_i) \cap \mathcal{C})$ and lies on the boundary of $\pi_i(\mathcal{C})$, we deduce that x lies on the boundary of $\pi_i(\Omega_i(\mathcal{O}_i) \cap \mathcal{C})$. By the Implicit Function Theorem [150, Th. 2.12], we deduce that $y \in \operatorname{crit}(\pi_i, \operatorname{reg} \mathcal{O}_i) \subset \mathscr{C}(\pi_i, \mathcal{O}_i) \subset \mathcal{O}_{i-1}$, which is a contradiction since we assumed that $J(y) = i$.

We conclude that $\mathcal{Z}$ satisfies Q_i. By induction, $\mathcal{Z}$ satisfies $\mathsf{Q}_1, \dots, \mathsf{Q}_d$, whence $\mathsf{Q}_1 \wedge \cdots \wedge \mathsf{Q}_d = \mathsf{Q}$. *q.e.d.*

We prove next that, for any d and any algebraic set $\mathcal{Z}$ of dimension d, $\mathsf{P}(\mathcal{Z})$ holds generically in $\mathrm{GL}_n(\mathbb{C})$. We assume for the moment the following intermediate result, whose proof is given in Section 2.3.

Given $\mathcal{Z} \subset \mathbb{C}^n$, we recall that in Section 2.1.2 we let $I_i = I(\mathcal{O}_i(\mathcal{Z}))$ and $I_i \circ M = I(\mathcal{O}_i(M^{-1}\mathcal{Z}))$.

Lemma 2.3. *Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d. There exists a non-empty Zariski open set $\mathscr{M} \subset \mathrm{GL}_n(\mathbb{C})$ such that, if $M \in \mathscr{M} \cap \mathbb{M}_{n,n}(\mathbb{Q})$, the following holds. Let $i \in \{0, \dots, d\}$, and let $\mathcal{P}$ be one of the prime components of $I_i \circ M$, and let $r = \dim Z_\mathbb{C}(\mathcal{P})$. Then $r \leq i$ and the ring extension $\mathbb{C}[x_1, \dots, x_r] \longrightarrow \mathbb{C}[x_1, \dots, x_n]/\mathcal{P}$ is integral.*

We prove in Proposition 2.4 that for any algebraic set $\mathcal{Z}$ of dimension d, the set of matrices M such that $\mathsf{P}(\mathcal{Z})$ holds, is Zariski dense in $\mathrm{GL}_n(\mathbb{C})$, hence in $\mathbb{M}_{n,n}(\mathbb{C})$. This implies that, for any algebraic set $\mathcal{Z}$, $\mathsf{P}(\mathcal{Z})$ is generic in $\mathbb{M}_{n,n}(\mathbb{C})$, and allows to conclude, applying Proposition 2.2, that Q holds generically.

Proposition 2.4. *Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d. There exists a non-empty Zariski open set $\mathscr{M} \subset \mathrm{GL}_n(\mathbb{C})$ such that, if $M \in \mathscr{M} \cap \mathbb{M}_{n,n}(\mathbb{Q})$, M satisfies $\mathsf{P}(\mathcal{Z})$.*

Proof. First, we remark that Point 2 of property $\mathsf{P}(\mathcal{Z})$ implies Point 1 of $\mathsf{P}(\mathcal{Z})$ (cf. [145, Ch. 1.5.4]). Hence it is sufficient to prove Point 2.

Let $\mathscr{M} \subset \mathrm{GL}_n(\mathbb{C})$ be the set defined in Lemma 2.3. It is non-empty and Zariski open, and for $M \in \mathscr{M} \cap \mathbb{M}_{n,n}(\mathbb{Q})$ and $i \in \{0, \ldots, d\}$, we deduce by Lemma 2.3 that any irreducible component of $\mathcal{O}_i(M^{-1}\mathcal{Z})$ is in Noether position, and has dimension at most i. Hence $\mathcal{O}_i(M^{-1}\mathcal{Z})$ is in Noether position and has dimension at most i. We deduce that Point 2 of $\mathsf{P}(\mathcal{Z})$ holds and hence that M satisfies $\mathsf{P}(\mathcal{Z})$. *q.e.d.*

2.2.2 Fibers over critical values

This section contains a generalization of [68, Lem. 19]. Indeed, we give in Proposition 2.5 a dimension result about fibers over critical values of linear projections restricted to the set $\mathcal{Z}$.

The main idea is to relate the geometry of intersections of fibers of π_i and the connected components of $\mathcal{Z} \cap \mathbb{R}^n$, with the decomposition of the algebraic set $\mathcal{Z}$, given in Section 2.1.2.

Proposition 2.5. *Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d, and let $M \in \mathrm{GL}_n(\mathbb{C})$ satisfy $\mathsf{P}(\mathcal{Z})$. Let $\mathcal{C} \subset \mathcal{Z} \cap \mathbb{R}^n$ be a connected component, $i \in \{0, \ldots, d\}$ and let $w \in \mathbb{R}^i$ belong to the boundary of $\pi_i(M^{-1}\mathcal{C})$. Then $\pi_i^{-1}(w) \cap M^{-1}\mathcal{C}$ is non-empty and contained in $\mathcal{O}_{i-1}(M^{-1}\mathcal{Z}) \cap M^{-1}\mathcal{C}$.*

Proof. We suppose without loss of generality that $M = \mathbb{I}_n$. Proposition 2.2 implies that if $w \in \mathbb{R}^i$ belongs to the boundary of $\pi_i(\mathcal{C})$, there exists $x \in \mathcal{O}_{i-1} \cap \mathcal{C}$ such that $\pi_i(x) = w$. Hence $(\mathcal{O}_{i-1} \cap \mathcal{C}) \cap (\pi_i^{-1}(w) \cap \mathcal{C}) \neq \emptyset$.

Let $J \colon \mathcal{Z} \to \{0, 1, \ldots, d\}$ be the function defined in the proof of Proposition 2.2. Suppose by contradiction that there exists $x \in \pi_i^{-1}(w) \cap \mathcal{C}$ such that $J(x) = j > i-1$, that is $x \in \mathcal{O}_j \setminus \mathcal{O}_{j-1}$. We deduce that $x \in \operatorname{reg} \Omega_j(\mathcal{O}_j) \cap \mathcal{C}$. Since $w = \pi_i(x)$ lies on the boundary of $\pi_i(\mathcal{C})$, we deduce that $\pi_j(x)$ lies on the boundary of $\pi_j(\Omega_j(\mathcal{O}_j) \cap \mathcal{C})$. Moreover, since $x \in \operatorname{reg} \Omega_j(\mathcal{O}_j) \cap \mathcal{C}$, by the Implicit Function Theorem [150, Th. 2.12] x is a critical point of the restriction of π_j to $\mathcal{O}_j$. Hence, $x \in \operatorname{crit}(\pi_j, \mathcal{O}_j) \subset \mathscr{C}(\pi_j, \mathcal{O}_j) \subset \mathcal{O}_{j-1}$. We conclude that $J(x) \leq j-1$, which contradicts the hypothesis. *q.e.d.*

2.3 Proofs of intermediate results

We refer to the notation introduced in Section 2.1.2, and we prove a result which implies Lemma 2.3. It consists of the generalization of [137, Prop. 1] to the non-equidimensional case. The proof of Lemma 2.3 can be deduced by applying *mutatis mutandis* the proof of [137, Prop. 2, Sec .2.4].

The prime components of an ideal I are the radical ideals of the primary ideals of a minimal primary decomposition of I (*cf.* Section 1.1.1 or [35, Th. 9, Ch. 4, § 7]).

Lemma 2.6. *Let $\mathcal{Z} \subset \mathbb{C}^n$ be an algebraic set of dimension d, and let $i \in \{0,\ldots,d\}$. Let $\mathcal{P} \subset \mathbb{Q}(B)[x]$ be one of the prime components of $I_i \circ B$, and let $r = \dim Z_{\mathbb{C}}(\mathcal{P})$. Then $r \leq i$ and the ring extension $\mathbb{Q}(B)[x_1,\ldots,x_r] \longrightarrow \mathbb{Q}(B)[x_1,\ldots,x_n]/_{\mathcal{P}}$ is integral.*

Proof. Our reasoning is by decreasing induction on i, the index i of the projection maps π_i. Let $i = d$, hence $I_d \circ B = I(B^{-1}\mathcal{Z})$. Let $\mathcal{P}$ be a prime component of $I_d \circ B$, and write $r = \dim Z_{\mathbb{C}}(\mathcal{P})$. The algebraic set $Z_{\mathbb{C}}(\mathcal{P})$ is irreducible of dimension $r \leq d$ and then $Z_{\mathbb{C}}(\mathcal{P}) \subset \Omega_r(B^{-1}\mathcal{Z})$. We conclude by applying Noether Normalization Lemma (*cf.* [98, Th. 2.1] or [104]).

Suppose that the statement is true for $i+1$. We write $\mathcal{O}_i$ instead of $\mathcal{O}_i(B^{-1}\mathcal{Z})$. One gets

$$I_i \circ B = I(\mathscr{S}(\mathcal{O}_{i+1})) \cap I(\mathscr{C}(\pi_{i+1},\mathcal{O}_{i+1})) \cap I(\mathscr{C}(\pi_{i+1},B^{-1}\mathcal{Z})).$$

Let $\mathcal{P}$ be a prime component of $I_i \circ B$, and let $r = \dim Z_{\mathbb{C}}(\mathcal{P})$. Then it is a prime component of one of the three ideals in the previous decomposition. We investigate below all possible cases:

First case. $I(\mathscr{S}(\mathcal{O}_{i+1})) \subset \mathcal{P}$. Hence

$$I(\Omega_0(\mathcal{O}_{i+1})) \cap \cdots \cap I(\Omega_i(\mathcal{O}_{i+1})) \cap I(\mathrm{sing}\,(\Omega_{i+1}(\mathcal{O}_{i+1}))) \subset \mathcal{P}.$$

Combined with the fact that $\mathcal{P}$ is prime, this implies that

- either $I(\Omega_j(\mathcal{O}_{i+1})) \subset \mathcal{P}$, for some $j \in \{0,\ldots,i\}$; then one gets $r \leq i$ and by the induction assumption that

  $$\mathbb{Q}(B)[x_1 \ldots x_r] \longrightarrow \mathbb{Q}(B)[x_1 \ldots x_n]/_{\mathcal{P}}$$

  is integral;

- or $I(\mathrm{sing}\,(\Omega_{i+1}(\mathcal{O}_{i+1}))) \subset \mathcal{P}$.

Assume that $I(\mathrm{sing}\,(\Omega_{i+1}(\mathcal{O}_{i+1}))) \subset \mathcal{P}$. Since $\dim(\Omega_{i+1}(\mathcal{O}_{i+1})) = i+1$ by definition, we deduce that

$$r = \dim Z_{\mathbb{C}}(\mathcal{P}) \leq \dim \mathrm{sing}\,(\Omega_{i+1}(\mathcal{O}_{i+1})) \leq i.$$

Let $f \circ B = \{f_1 \circ B, \ldots, f_s \circ B\}$ be a set of generators of the ideal $I(\Omega_{i+1}(\mathcal{O}_{i+1}))$. Then, since $f \circ B$ define a radical ideal and $\Omega_{i+1}(\mathcal{O}_{i+1})$ is equidimensional,

$$I(\mathrm{sing}\,(\Omega_{i+1}(\mathcal{O}_{i+1}))) = \sqrt{\langle f \circ B, g_1, \ldots, g_N \rangle},$$

where $g_1, \ldots, g_N$ are the minors of order $(n-i-1) \times (n-i-1)$ of the Jacobian matrix $D(f \circ B)$. We prove below by induction on t that for any prime $\mathcal{Q}$ associated to $\langle f \circ B, g_1, \ldots, g_t \rangle$,

$$\mathbb{Q}(B)[x_1 \ldots x_r] \longrightarrow \mathbb{Q}(B)[x_1 \ldots x_n] \,/_{\mathcal{Q}}$$

is integral. Taking $t = N$ concludes the proof.

The induction assumption implies the claimed property for $t = 0$. Assume that for any prime $\mathcal{Q}'$ associated to $\langle f \circ B, g_1, \ldots, g_t \rangle$, the extension $\mathbb{Q}(B)[x_1 \ldots x_r] \longrightarrow \mathbb{Q}(B)[x_1 \ldots x_n] \,/_{\mathcal{Q}'}$ is integral. We prove below that for any prime $\mathcal{Q}$ associated to $\langle f \circ B, g_1, \ldots, g_{t+1} \rangle$, the extension

$$\mathbb{Q}(B)[x_1 \ldots x_r] \longrightarrow \mathbb{Q}(B)[x_1 \ldots x_n] \,/_{\mathcal{Q}}$$

is integral.

Indeed, any prime $\mathcal{Q}$ associated to $\langle f \circ B, g_1, \ldots, g_{t+1} \rangle$ is a prime associated to $\mathcal{Q}' + \langle g_{t+1} \rangle$. Suppose that $g_{t+1} \notin \mathcal{Q}'$ (otherwise, the conclusion follows immediately) and let $r' = \dim Z_{\mathbb{C}}(\mathcal{Q}')$.

By Krull's Principal Ideal Theorem [39, Sec. 8.2.2], $\mathcal{Q}' + \langle g_{t+1} \rangle$ is equidimensional of dimension $r' - 1$. Following the same argumentation of the proof of [137, Prop. 1], the ideal $\mathcal{Q}' + \langle g_{t+1} \rangle$ contains a monic polynomial in $x_{r'}$, so that the extension

$$\mathbb{Q}(B)[x_1 \ldots x_{r'-1}] \longrightarrow \mathbb{Q}(B)[x_1 \ldots x_n] \Big/ \mathcal{Q}' + \langle g_{t+1} \rangle$$

is integral, and we conclude.

Second case. $I(\mathscr{C}(\pi_{i+1}, \mathcal{O}_{i+1})) \subset \mathcal{P}$. Recall that $\mathscr{C}(\pi_{i+1}, \mathcal{O}_{i+1})$ is the union of the algebraic sets $\mathrm{crit}\,(\pi_{i+1}, \mathrm{reg}\,(\Omega_{i+1}(\mathcal{O}_{i+1})))$ and $\Omega_j(B^{-1}\mathcal{Z})$ for $j \in \{0, \ldots, i\}$.

When $I(\Omega_j(B^{-1}\mathcal{Z})) \subset \mathcal{P}$, the induction assumptions allow to conclude. Thus we focus on the case when $\mathcal{P} \supset I(\mathrm{crit}\,(\pi_{i+1}, \mathrm{reg}\,(\Omega_{i+1}(\mathcal{O}_{i+1}))))$. The latter ideal is built as follows. Let $\langle f \circ B \rangle = I(\Omega_{i+1}(\mathcal{O}_{i+1}))$, and let $g_1, \ldots, g_N$ be the $(n-i-1) \times (n-i-1)$ minors of the matrix obtained by deleting the first i columns of $D(f \circ B)$. Let $J = I(\mathrm{sing}\,(\Omega_{i+1}(\mathcal{O}_{i+1})))$. The following equality is immediate:

$$I(\mathrm{crit}\,(\pi_{i+1}, \mathrm{reg}\,(\Omega_{i+1}(\mathcal{O}_{i+1})))) = \sqrt{\langle f \circ B, g_1, \ldots, g_N \rangle} : J^{\infty}.$$

We deduce that $\mathcal{P}$ is a prime component of $\sqrt{\langle f \circ B, g_1, \ldots, g_N \rangle}$ such that $Z_{\mathbb{C}}(\mathcal{P})$ is not contained in $\mathrm{sing}\,(\Omega_{i+1}(\mathcal{O}_{i+1}))$. The integral ring extension property is already proved (by induction) for every component of $\langle f \circ B \rangle$, and hence we proceed as in the first point.

Third case. $I(\mathscr{C}(\pi_{i+1}, B^{-1}\mathcal{Z})) \subset \mathcal{P}$. Again, recall that $\mathscr{C}(\pi_{i+1}, B^{-1}\mathcal{Z})$ is the union of $\Omega_j(B^{-1}\mathcal{Z})$ for $j \in \{0, \ldots, i\}$ and the union for $r' \geq i$ of the sets $\mathrm{crit}\,(\pi_i, \mathrm{reg}\,(\Omega_{r'}(B^{-1}\mathcal{Z})))$ of critical points of the restriction of π_i to the regular locus of $\Omega_{r'}(B^{-1}\mathcal{Z})$.

For $r' \geq i+1$, we can assume $I(\mathrm{crit}\,(\pi_{i+1}, \mathrm{reg}\,(\Omega_{r'}))) \subset \mathcal{P}$. The proof follows exactly the same argumentation as the one in the second point. *q.e.d.*

Chapter 3

Real root finding for low rank linear matrices

Linear matrices and their loci of rank defects arise as natural algebraic structures both in theoretical and in applicative contexts, of particular interest in real algebraic geometry, polynomial optimization, control theory. A common feature of problems involving these objects is that low rank is synonymous of structure and sparsity. The algorithm described in this chapter gives a solution to the so-called real root finding problem for algebraic sets, in the context of structured polynomial systems generated by rank deficiencies in linear matrices.

The results that are presented are part, or generalizations, of the previous works [68], accepted for publication in Journal of Symbolic Computation, and [70], submitted to Journal of Complexity.

3.1 Introduction

Let $m, n \in \mathbb{N}_*$. Consider $n+1$ square rational matrices of size m, stored in a vector

$$A = (A_0, A_1, \ldots, A_n) \in \mathbb{M}^{n+1}_{m,m}(\mathbb{Q}),$$

where $\mathbb{M}^{n+1}_{m,m}(\mathbb{Q})$ is the $(n+1)-$fold Cartesian product $\mathbb{M}_{m,m}(\mathbb{Q}) \times \cdots \times \mathbb{M}_{m,m}(\mathbb{Q})$. We associate to the vector A (and we use the same symbol, with abuse of notation) the affine map

$$\begin{array}{rccl} A: & \mathbb{C}^n & \to & \mathbb{M}_{m,m}(\mathbb{C}) \\ & x & \mapsto & A_0 + x_1 A_1 + \cdots + x_n A_n, \end{array}$$

where x is the complex vector $(x_1, \ldots, x_n) \in \mathbb{C}^n$. We also refer to the matrix $A(x)$ as a linear matrix or linear pencil.

Hence, we identify elements of $\mathbb{M}^{n+1}_{m,m}(\mathbb{Q})$ with affine maps as above. Moreover, for a fixed $x \in \mathbb{C}^n$, it is straightforward to remark that the instantiated matrix $A(x)$ defines a linear map $A(x)\colon \mathbb{C}^m \to \mathbb{C}^m$. Hence, the map A associates vectors in $\mathbb{C}^n$ to linear endomorphisms of $\mathbb{C}^m$.

3.1.1 Problem statement

Let $m, n \in \mathbb{N}_*$, and $0 \leq r \leq m-1$. Given matrices $A_0, A_1, \ldots, A_n \in \mathbb{M}_{m,m}(\mathbb{Q})$, we define the algebraic set

$$\mathcal{D}_r = \{x \in \mathbb{C}^n : \operatorname{rank} A(x) \leq r\} = Z_{\mathbb{C}}(\text{minors}\,(r+1, A)),$$

and the corresponding real algebraic set

$$\mathcal{D}_r \cap \mathbb{R}^n = \{x \in \mathbb{R}^n : \operatorname{rank} A(x) \leq r\} = Z_{\mathbb{R}}(\text{minors}\,(r+1, A)).$$

The main goal of this chapter is to solve the following problem.

Problem 3 (Real root finding for determinantal varieties)**.** *Design an efficient exact algorithm whose input is the vector $A = (A_0, A_1, \ldots, A_n) \in \mathbb{M}_{m,m}(\mathbb{Q})$ and the integer r, and whose output is either*

(A) the empty list, if and only if $\mathcal{D}_r \cap \mathbb{R}^n = \emptyset$, or

(B) a rational parametrization encoding a finite subset of $\mathbb{R}^n$ meeting each connected component of $\mathcal{D}_r \cap \mathbb{R}^n$.

More explicitly, solving Problem 3 would provide answers to the following questions involving a given pencil A and the corresponding real determinantal varieties $\mathcal{D}_r \cap \mathbb{R}^n$:

- Is $\mathcal{D}_r \cap \mathbb{R}^n$ empty?
- If it is not empty, how to estimate the number of its connected components?
- If it is not empty, how to compute sample points $x^* \in \mathcal{D}_r \cap \mathbb{R}^n$?

3.1.2 Main result and chapter outline

Strategy and main result

The method we propose to solve Problem 3 is threefold. We first give a brief synopsis of our strategy. We recall that the definition of $\mathcal{D}_r$ is in Section 3.1.1.

Step 1: Regularization. For $m \geq 2$, $\mathcal{D}_r$ is, typically, a singular algebraic set. Its singularities arise generically from degenerate subsets where the rank defect is greater than one. We remark that this is not always the case (*cf.* the example of Kummer spectrahedra in [130]). Consequently, we first substitute $\mathcal{D}_r$ with the lifted set:

$$\mathcal{V}_r(A, U, S) = \big\{(x, y) \in \mathbb{C}^n \times \mathbb{C}^{m(m-r)} : A(x)Y(y) = 0, UY(y) - S = 0\big\}.$$

for some full matrices $U \in \mathbb{M}_{m-r,r}(\mathbb{Q})$, $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$. The lifted set $\mathcal{V}_r$ is smooth and equidimensional under genericity assumptions on the defining linear matrix A. Furthermore, the semi-algebraic set obtained by projecting the lifted set over $\mathbb{C}^n$ is dense in $\mathcal{D}_r$. The lifted set has no more determinantal structure, but it is defined by bilinear equations, and hence its degree is well controlled. One obtains finally a representation of $\mathcal{D}_r$ as the projection of a set having regularity properties.

Step 2: Reduction. The dimension of the lifted set equals the dimension of $\mathcal{D}_r$. Hence, one deals with positive dimensional complex sets (and *a priori* also with positive dimensional real sets). To obtain subsets with controlled dimension (actually ≤ 0) we apply a dedicated critical point method based on Lagrangian systems. Namely, we compute critical points of the restriction of linear projections to the lifted set, and to keep its bilinear structure, we choose these linear projections appropriately. The image of any connected component of $\mathcal{D}_r \cap \mathbb{R}^n$ via these maps turns out to be (Euclidean) closed sets.

Step 3: Recursion. Finally, Step 2 is iterated over a generic fiber of the projection maps. If the regularity assumptions needed to perform the second step always hold, one can also extract a finite sample subset of the pre-image of the previous maps.

The main result of this chapter is the following theorem. Its proof can be obtained by merging the proof of Theorem 3.9, page 86, yielding the correctness of our algorithm, with that of Theorem 3.17, page 97, which computes a bound for its complexity.

The algorithm presented in the following theorem is probabilistic, and relies on regularity assumptions on the incidence variety (*i.e.* smoothness and equidimensionality) and on its defining polynomial system (that must define a radical ideal). Also, we assume as hypothesis that the low rank loci associated to the input pencil A have the expected dimension.

We prove that the set of input data for which the genericity assumptions are not satisfied is included in a proper algebraic subset of its parameter space; similarly, the set of parameters chosen during the procedure, for which the output is not correct, lies in a proper algebraic subset of its parameter space.

Theorem 3.1 (Main Theorem 2). *Let A be a $n-$variate linear matrix of size m, and let $r \leq m-1$. Let f be the polynomial system defining the incidence variety $\mathcal{V}_r$. Suppose that the ideal $\langle f \rangle$ is radical, that $\mathcal{V}_r$ is smooth and equidimensional and that, for $0 \leq p \leq r$, $\mathcal{D}_p$ is empty or has the expected codimension $(m-p)^2$.*

There exists a probabilistic algorithm such that, with input (A, r), returns in output a rational parametrization $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \subset \mathbb{Q}[t]$ of degree in

$$\mathcal{O}\left((n-(m-r)^2)\binom{n+m(m-r)}{n}^3\right)$$

within
$$\mathcal{O}\left((n+m^2-r^2)^7\binom{n+m(m-r)}{n}^6\right)$$
arithmetic operations over $\mathbb{Q}$. The set defined by q contains at least one point in each connected component of the real algebraic set $\mathcal{D}_r \cap \mathbb{R}^n$.

Outline

Chapter 3 is organized as follows. In Section 3.3 we describe the algorithm, starting from the definition of the algebraic sets that are built during its execution. We also write down its formal description and a proof of its correctness. The main results towards the correctness theorem are then proved in Section 3.5. Section 3.4 shows how one can compute explicit bounds for the degree of the output representation of the solution and we also report on a formula for the complexity. Results of numerical experiments showing the practical performance of our algorithm are provided in Chapter 5, Section 5.2.

3.2 Some examples

Linear matrices are interesting structures in different contexts, as shown in the following three examples. We will come back to these examples in Section 5.2.2.

Example 3.2 (Joint probability matrices)**.** *The first example comes from statistics. Consider two given discrete random variables, with possible states in the set $\{1, \ldots, m\}$. Hence, these random variables can take integer values between 1 and m with the joint probability distribution represented by the square matrix*
$$A(x) = \begin{pmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & & \vdots \\ x_{m,1} & \cdots & x_{m,m} \end{pmatrix},$$
where for $1 \leq i, j \leq m$, $x_{i,j}$ is the probability that the first variable is equal to i and the second is equal to j, that is the joint probability. We deduce that the entries of the matrix $A(x)$ are nonnegative and that they sum up to one, hence that any possible configuration of joint probabilities lies in the simplex
$$\left\{A(x) \in \mathbb{M}_{m,m}(\mathbb{R}) : x_{i,j} \geq 1, \sum_{i,j} x_{i,j} = 1\right\}.$$
After discarding the nonnegativity condition on the entries of A, the associated low rank locus with expected rank r is the algebraic set
$$\mathcal{D}_r \cap \mathbb{R}^{m^2} = \left\{x \in \mathbb{R}^{m^2} : \operatorname{rank} A(x) \leq r, \sum_{i,j} x_{i,j} = 1\right\}.$$

Low rank of such probability matrices represents statistical independence. More precisely, the set $\mathcal{D}_1 \cap \mathbb{Q}^{m^2}$ contains all possible joint probability matrices corresponding to independent distributions, while for $r > 1$ the set $\mathcal{D}_r$ corresponds to mixtures of r independent probability distributions, cf. [86, Sec. 2]. In our context, the matrix X is linear of size m in $n = m^2 - 1$ variables, since one can perform the substitution

$$x_{1,1} \;\leftarrow\; 1 - \sum_{(i,j)\neq(1,1)} x_{i,j}.$$

Example 3.3 (Sylvester matrices)**.** *Suppose that*

$$\begin{aligned} f(y) &= f_0 + f_1 y + \cdots + f_{d_1} y^{d_1} \\ g(y) &= g_0 + g_1 y + \cdots + g_{d_2} y^{d_2} \end{aligned}$$

are two univariate polynomials of degree d_1 and d_2. It is classically known (cf. [51, Ch. 12]) that these two polynomials share a common root if and only if the determinant of the associated $(d_1 + d_2) \times (d_1 + d_2)$ Sylvester matrix

$$A = \begin{pmatrix} f_0 & f_1 & f_2 & \dots & f_{d_1} & 0 & \dots & 0 \\ 0 & f_0 & f_1 & f_2 & \dots & f_{d_1} & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & f_0 & f_1 & f_2 & \dots & f_{d_1} \\ g_0 & g_1 & \dots & g_{d_2} & 0 & \dots & \dots & 0 \\ 0 & g_0 & g_1 & \dots & g_{d_2} & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & g_0 & g_1 & \dots & \dots & g_{d_2} \end{pmatrix}$$

vanishes. The polynomial $R(f, g) = \det A \in \mathbb{Q}[f_i, g_j]$, of degree $d_1 + d_2$, is called the resultant of the couple (f, g). The resultant $R_{d_1,d_2}(f, g)$ vanishes if and only if the greatest common divisor of f and g is not a constant. The matrix A is linear in the coefficients of polynomials f, g and in Section 5.2.2 we will provide numerical experiments when such coefficients are expressed as affine polynomials on a given set of unknowns.

Example 3.4 (Hurwitz stability)**.** *Let $f(z) = f_0 z^m + f_1 z^{m-1} + \ldots + f_{m-1} z + f_m$ be a real univariate polynomial of complex variable z, of degree at most m. We consider the associated $m \times m$ Hurwitz matrix*

$$A = \begin{pmatrix} f_1 & f_3 & f_5 & \cdots & 0 \\ f_0 & f_2 & f_4 & & \vdots \\ 0 & f_1 & f_3 & & \vdots \\ \vdots & 0 & \vdots & \ddots & 0 \\ 0 & \vdots & & & f_m \end{pmatrix}.$$

This matrix is obtained as follows. For $i = 1, \ldots, m$, the diagonal element at row i and column i is the coefficient f_i. Then, for $i = 1, \ldots, m$, the $i-th$ column contains the coefficients $\{f_0, f_1, \ldots, f_m\}$ in decreasing order (with respect to the index j of f_j) until there is enough room, otherwise it contains zeros. For example, the Hurwitz matrix of the cubic polynomial $3 + 5z - 2z^2 + 4z^3$ is

$$\begin{pmatrix} -2 & 3 & 0 \\ 4 & 5 & 0 \\ 0 & -2 & 3 \end{pmatrix}.$$

Suppose that the polynomial f is monic, that is $f_0 = 1$. In this case f is said to be stable if all its complex zeroes have strictly negative real part. This definition applies in particular when f is the characteristic polynomial of the matrix M of the linear ordinary differential equation

$$\frac{dx}{dt}(t) = M\,x(t),$$

and in this case the stability of f implies asymptotic convergence to the origin, of the solutions, for any initial condition.

Denote by $A[i]$ the leading principal submatrix, of size i, of A, that is the square matrix obtained by selecting the first i rows and i columns of A. By the Hurwitz Criterion [92, Sec. 4.3], stability holds whenever the determinants of all leading principal submatrices of A are strictly positive. While checking whether f is stable or not amounts to computing the sign of polynomial expressions of the vector $(f_1, \ldots, f_m)$, the topology of the stability locus

$$S_m = \Big\{(f_1, \ldots, f_m) \in \mathbb{R}^m \ : \ \det A[i] > 0, \ i = 1, \ldots, m\Big\}$$

is not totally understood, apart from limited low-dimensional cases. This is a basic open semi-algebraic set, non-convex for $m \geq 3$. An interesting problem in control theory consists of finding the point in the closure of S_m which is nearest to a given unstable polynomial, that is to minimize the Euclidean distance function to the set S_m. Degree bounds for this optimization problem are known, cf. [37].

3.3 The algorithm

3.3.1 A geometrical viewpoint

Algebraic sets defined by rank defects in linear pencils can be represented as affine sections of the locus of rank defects in the whole space of $m \times m$ matrices (*cf.* [40] for a historical survey via commutative algebra). Let

$$X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & & \vdots \\ x_{m,1} & \cdots & x_{m,m} \end{pmatrix}$$

be a $m\times m$ matrix whose entries are variables $x_{i,j}$. Let $\widetilde{x}$ be the vector of variables $x_{i,j}$. Then, for $0 \leq r \leq m-1$, the ideal $I = \langle \text{minors}\,(r+1, X)\rangle \subset \mathbb{Q}[\widetilde{x}]$ is generated by homogeneous polynomials and its associated zero set is the affine cone of $m\times m$ matrices of rank at most r. We denote it by $\sigma_r = Z_{\mathbb{C}}(I) \subset \mathbb{M}_{m,m}(\mathbb{C})$ (we recall that $\mathbb{M}_{m,m}(\mathbb{C})$ is the set of $m \times m$ matrices with complex entries).

Moreover, if $A(x) = A_0 + x_1A_1 + \cdots + x_nA_n$, and $a_{\ell,i,j}$ is the entry of A_ℓ at $i-$th row and $j-$th column, then the algebraic set $\mathcal{D}_r$ is actually obtained by instantiating $x_{i,j}$ in I to $a_{0,i,j} + a_{1,i,j}x_1 + \ldots + a_{n,i,j}x_n$ for $i,j = 1,\ldots,m$, by eliminating variables $x_{i,j}$, and by taking the zero locus.

Proposition 3.5. *There exists a non-empty Zariski open set $\mathscr{A} \subset \mathbb{M}^{n+1}_{m,m}(\mathbb{C})$ such that, if $A \in \mathscr{A} \cap \mathbb{M}^{n+1}_{m,m}(\mathbb{Q})$, for all $0 \leq r \leq m-1$, the following holds:*

- *the algebraic set $\mathcal{D}_r$ is either empty or it has dimension $n-(m-r)^2$;*
- *for $n \geq (m-r)^2$, the degree of $\mathcal{D}_r$ is given by the Giambelli-Porteous-Thom formula:*
$$\prod_{i=0}^{m-r-1} \frac{i!(m+i)!}{(m-1-i)!(m-r+i)!}.$$

Proof. Let $0 \leq r \leq m-1$. By [28, Prop. 1.1], the set $\sigma_r \subset \mathbb{M}_{m,m}(\mathbb{C})$ is irreducible of codimension $(m-r)^2$ and dimension $r(2m-r)$. Let $x = (x_1,\ldots,x_n)$ and let

$$I = \langle \text{minors}\,(r+1, X)\rangle + \langle x_{i,j} - a_{0,i,j} - a_{1,i,j}x_1 - \cdots - a_{n,i,j}x_n\rangle_{1\leq i,j\leq m}$$

as an ideal of $\mathbb{Q}[\widetilde{x}, x]$. The set $Z_{\mathbb{C}}(\text{minors}\,(r+1,X)) \subset \mathbb{C}^{m^2+n}$ is irreducible of codimension $(m-r)^2$ and dimension $m^2+n-(m-r)^2$ (in fact, variables $x_1,\ldots,x_n$ are free, and this set is the cylinder over σ_r in $\mathbb{C}^{m^2+n}$). If linear forms $x_{i,j} - a_{0,i,j} - a_{1,i,j}x_1 - \cdots - a_{n,i,j}x_n$ are generic, then $Z_{\mathbb{C}}(I) \subset \mathbb{C}^{m^2+n}$ is empty or equidimensional of dimension $n-(m-r)^2$ (Bertini's theorem, see [145]).

Let $\pi\colon \mathbb{C}^{m^2+n} \to \mathbb{C}^n$ be the projection $\pi(\widetilde{x}, x) = x$. Let V be an irreducible component of $Z_{\mathbb{C}}(I)$. Then V has dimension $n-(m-r)^2$. For $x \in \pi(V)$, the fiber $\pi^{-1}(x)$ is finite of degree 1. By the Theorem on the Dimension of Fibers [145, Sec. 6.3, Th. 7], $\dim \pi(V) = \dim V = n-(m-r)^2$. Moreover, $\deg \pi(Z_{\mathbb{C}}(I)) = \deg Z_{\mathbb{C}}(I)$. The latter degree equals the degree of $Z_{\mathbb{C}}(\langle \text{minors}\,(r+1,X)\rangle) \subset \mathbb{C}^{m^2}$, which is actually given by Giambelli-Porteous-Thom formula, see for example [10, Ch. II, § 4].

We conclude that for all $0 \leq r \leq m-1$ there exists a non-empty Zariski open set $\mathscr{A}^{(r)} \subset \mathbb{M}^{n+1}_{m,m}(\mathbb{C})$ such that if $A \in \mathscr{A}^{(r)} \cap \mathbb{M}^{n+1}_{m,m}(\mathbb{Q})$, then either $\mathcal{D}_r$ is empty or it has the expected dimension $n-(m-r)^2$ and the expected degree given by the above formula. We conclude by defining $\mathscr{A} = \bigcap_r \mathscr{A}^{(r)}$ which is non-empty and Zariski open. *q.e.d.*

3.3.2 Notation

We recall first the notation for linear changes of variables, given on page 35. For $M \in \mathrm{GL}_n(\mathbb{C})$, and a polynomial $f \in \mathbb{C}[x]$, with $x = (x_1, \ldots, x_n)$, we denote by $f \circ M$ the polynomial such that $(f \circ M)(x) = f(M\,x)$ for all $x \in \mathbb{C}^n$. If $f = (f_1, \ldots, f_s) \subset \mathbb{C}[x]$, and $\mathscr{Z} = Z_{\mathbb{C}}(f)$, we denote by $M^{-1}\mathscr{Z}$ the set defined by $\langle f_1 \circ M, \ldots, f_s \circ M \rangle$, that is $M^{-1}\mathscr{Z} = \{x \in \mathbb{C}^n : M\,x \in \mathscr{Z}\}$.

Let $A = (A_0, A_1, \ldots, A_n) \in \mathbb{M}_{m,m}^{n+1}(\mathbb{Q})$ be the given vector of square matrices with rational coefficients. For any $M \in \mathrm{GL}_n(\mathbb{Q})$, with the same spirit as above we use the notation $A \circ M$ to denote the linear matrix $A(M\,x)$.

Incidence varieties

As introduced in Section 3.1.2, the algorithm first builds an algebraic set over the determinantal variety $\mathcal{D}_r$. Indeed, the first motivation is to remove singularities. To do this, we introduce $m(m-r)$ variables $y = (y_{1,1}, \ldots, y_{m,m-r})$, stored in an $m \times (m-r)$ matrix

$$Y(y) = \begin{pmatrix} y_{1,1} & \cdots & y_{1,m-r} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ y_{m,1} & \cdots & y_{m,m-r} \end{pmatrix},$$

and for any full rank matrix $U \in \mathbb{M}_{m-r,m}(\mathbb{Q})$ and $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$, we define the algebraic set

$$\mathcal{V}_r(A, U, S) = \big\{(x, y) \in \mathbb{C}^n \times \mathbb{C}^{m(m-r)} : A(x)Y(y) = 0, UY(y) - S = 0\big\}.$$

By basic linear algebra arguments, $Y(y)$ has full rank if and only if there exist U and S as above such that $UY(y) = S$. Hence the linear entries of $UY(y) - S$ encode the information that the columns of $Y(y)$ are linearly independent, and $\mathcal{V}_r(A, U, S)$ is an algebraic subset of the locally closed set $\mathscr{Z} \cap \mathscr{O}$, where $\mathscr{Z} = Z_{\mathbb{C}}(A(x)Y(y))$ and $\mathscr{O} = \{(x, y) \in \mathbb{C}^n \times \mathbb{C}^{m(m-r)} : \operatorname{rank} Y(y) = m - r\}$.

We refer to $\mathcal{V}_r(A, U, S)$ as the incidence variety over $\mathcal{D}_r$ (with parameters U, S), and we use the short-hand notation $\mathcal{V}_r$ when parameters A, U and S are clear. Remark that the projection of $\mathcal{V}_r$ on the space of x variables is contained in $\mathcal{D}_r$. We also define the polynomial system

$$\begin{array}{rrcl} f(A, U, S): & \mathbb{C}^{n+m(m-r)} & \to & \mathbb{C}^{(2m-r)(m-r)} \\ & (x, y) & \mapsto & (A(x)Y(y),\ UY(y) - S) \end{array}$$

whose zero locus is exactly $\mathcal{V}_r$.

Finally, given $t \in \mathbb{C}$, consider the map

$$\begin{array}{rrcl} f_t: & \mathbb{C}^{n+m(m-r)} & \to & \mathbb{C}^{(2m-r)(m-r)+1} \\ & (x, y) & \mapsto & (f(A, U, S),\ x_1 - t). \end{array}$$

Its zero locus $\mathcal{V}_{r,t} = Z_{\mathbb{C}}(f_t)$ is the intersection of $\mathcal{V}_r$ with the hyperplane $x_1 - t = 0$. We remark that $f_t = f(\widetilde{A}, U, S)$ with $\widetilde{A} = (A_0 + tA_1) + x_2A_2 + \cdots + x_nA_n$.

Lagrange system

Let A, U, S as in the previous section, and let $f = f(A, U, S)$. We denote by $D_1 f$ the matrix obtained from Df by removing the first column (that is, the partial derivatives with respect to x_1). Let $z \in \mathbb{C}^{(2m-r)(m-r)}$ be a vector of variables.

Let $v \in \mathbb{C}^{(2m-r)(m-r)}$. We define the map

$$\begin{array}{rccc} \ell(A \circ M, U, S, v) : & \mathbb{C}^{n+(3m-r)(m-r)} & \to & \mathbb{C}^{n+(3m-r)(m-r)} \\ & (x, y, z) & \mapsto & (f(A \circ M, U, S), z' D_1 f,\ v'z - 1) \end{array}$$

We call $\ell(A \circ M, U, S, v)$ the Lagrange system associated to parameters A, M, U, S and v. The algebraic set $Z_{\mathbb{C}}(\ell(A \circ M, U, S, v))$ is denoted by $\mathcal{Z}(A \circ M, U, S, v)$.

3.3.3 Description

Input-output data representation

The input of the algorithm described in this chapter is a linear matrix $A(x) = A_0 + x_1A_1 + \cdots + x_nA_n$ encoded by the vector of its defining $m \times m$ matrices $A = (A_0, A_1, \ldots, A_n)$ with coefficients in $\mathbb{Q}$, and an integer $r \in \{0, \ldots, m-1\}$.

During the execution of the algorithm, one has to represent polynomials with rational coefficients: we use in this case the standard dense representation of polynomials, that is we represent $f \in \mathbb{Q}[x]$, of degree $\leq d$, with the vector of its $\binom{n+d}{n}$ coefficients.

The initial problem is reduced to isolating the real solutions of an algebraic set $\mathcal{Z} \subset \mathbb{C}^n$ of dimension at most 0. We encode the output $\mathcal{Z}$ by means of a rational parametrization, that is by a polynomial system $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \in \mathbb{Q}[t]^{n+2}$ such that:

- q_0 and q_{n+1} are coprime (that is, they do not share complex roots);
- $\mathcal{Z} = \left\{ \left(\frac{q_1(t)}{q_0(t)}, \ldots, \frac{q_n(t)}{q_0(t)} \right) \in \mathbb{C}^n : q_{n+1}(t) = 0 \right\}$;
- $\mathcal{Z}$ is in one-to-one correspondence with the roots of q_{n+1}.

A discussion about rational parametrization and their use for representing zero-dimensional ideals has been given in Section 1.2.2.

Genericity properties

The algorithm works under some regularity assumptions on the input data (A, r). We describe next these assumptions and how they are checked in practice.

Property G_1. *A linear matrix $A \in \mathbb{M}_{m,m}^{n+1}(\mathbb{Q})$ satisfies G_1 if* $\operatorname{sing} \mathcal{D}_r = \mathcal{D}_{r-1}$.

A proof of the fact that G_1 is generic in $\mathbb{M}_{m,m}^{n+1}(\mathbb{Q})$ is given further, in Proposition 3.6.

Property G_2. *A polynomial system $f = (f_1, \ldots, f_s) \in \mathbb{Q}[x]^s$ satisfies G_2 if:*

- *$\langle f \rangle$ is radical, and*
- *$Z_{\mathbb{C}}(\langle f \rangle)$ is either empty or smooth and equidimensional.*

We provide below a sufficient condition for a polynomial system $f = (f_1, \ldots, f_s) \in \mathbb{Q}[x]^s$ to satisfy G_2. Indeed, suppose that the rank of the Jacobian matrix Df is constant and equal to c on the algebraic set $Z_{\mathbb{C}}(\langle f \rangle)$. We deduce by applying the Jacobian Criterion (*cf.* [39, Th. 16.19] or Theorem 1.9) that the ideal generated by f is radical and that $Z_{\mathbb{C}}(\langle f \rangle)$ is either empty or smooth and equidimensional of codimension c.

Property G_3. *$(A, r) \in \mathbb{M}_{m,m}^{n+1}(\mathbb{Q}) \times \mathbb{Z}$ satisfies G_3 if for all $0 \le p \le r$, $\mathcal{D}_p$ is either empty or has codimension $(m-p)^2$.*

We finally define property G as follows.

> **Property G**
>
> Let $A \in \mathbb{M}_{m,m}^{n+1}(\mathbb{Q})$. Then A satisfies G if
>
> - A satisfies G_1, and
> - $f(A, U, S)$ satisfies G_2, for any U, S full rank, and
> - $(A, m-1)$ satisfies G_3.

Property G can be checked in practice using Gröbner bases, as follows:

1. Let f be the polynomial system defined by the minors of size $r+1$ of A, and let c be the codimension of $\mathcal{D}_r$ (this can be computed using Gröbner bases, *cf.* [35, Ch. 9,§. 3]). The singular locus $\operatorname{sing} \mathcal{D}_r$ is the set of points where f and the $c \times c$ minors of Df vanish. Then we use Gröbner bases computations to check that $\operatorname{sing} \mathcal{D}_r$ and $\mathcal{D}_{r-1}$ define the same set.

2. To check G_2, we use the Jacobian criterion as described above. We compute the codimension c of $Z_{\mathbb{C}}(f)$ using Gröbner bases (*cf.* [35, Ch. 9,§. 3]). Then we check whether the set defined by f and by the $c \times c$ minors of Df is empty or not, by computing a reduced Gröbner basis of $\langle f, \operatorname{minors}(c, Df) \rangle$. If this is $\{1\}$, then f satisfies G_2.

3. The dimension of a complex algebraic set, hence property G_3, can be computed using Gröbner bases, as above.

We prove in Proposition 3.6 that G holds generically and that it is generically inherited in the next recursive calls of $\mathsf{LowRankRec}$.

Main subroutines

The algorithm uses different subroutines. The first group of these routines manipulates vectors of polynomials:

- IsReg

  Input: Parameters A, U, S;

  Output: `true` if A satisfies G, `false` otherwise.

- RatPar

  Input: $f \subset \mathbb{Q}[x]$;

  Output: an error message if $\dim Z_{\mathbb{C}}(\langle f \rangle) > 0$, otherwise a rational parametrization $q \subset \mathbb{Q}[t]$ of its solutions.

The second group of subroutines manipulates rational parametrizations of finite sets:

- Project

  Input: a rational parametrization of a set $\mathcal{Z} \subset \mathbb{Q}[x_1, \ldots, x_N]$ and a subset $\{x_1, \ldots, x_i\} \subset \{x_1, \ldots, x_N\}$;

  Output: a rational parametrization of the projection of $\mathcal{Z}$ on the space generated by $\{x_1, \ldots, x_i\}$.

- Lift

  Input: a rational parametrization of a set $\mathcal{Z} \subset \mathbb{Q}[x_1, \ldots, x_N]$ and $t \in \mathbb{C}$;

  Output: a rational parametrization of $\{(t, x) : x \in \mathcal{Z}\}$.

- Image

  Input: a rational parametrization of a set $\mathcal{Z} \subset \mathbb{Q}[x_1, \ldots, x_N]$ and a matrix $M \in \mathrm{GL}_N(\mathbb{Q})$;

  Output: a rational parametrization of $M^{-1}\mathcal{Z}$.

- Union

  Input: two rational parametrizations of $\mathcal{Z}_1, \mathcal{Z}_2 \subset \mathbb{Q}[x_1, \ldots, x_N]$

  Output: a rational parametrization of $\mathcal{Z}_1 \cup \mathcal{Z}_2$.

Formal description

With input (A, r), the algorithm, which is called LowRank, draws randomly matrices U and S and checks whether A satisfies G by applying the subroutine IsReg. If this is the case, it calls a recursive subroutine called LowRankRec with the same input (A, r). Otherwise it returns an error message.

LowRank(A, r)

Input: A $n-$variate linear matrix $A(x)$ of size m, encoded by the $m^2(n+1)$ rational entries of $A_0, A_1, \ldots, A_n$, and an integer $0 \leq r \leq m-1$;

Output: Either an error message stating that genericity assumptions are not satisfied, or the output of LowRankRec (*cf.* below).

Procedure:

1. Choose randomly $U \in \mathbb{Q}^{(m-r)\times m}, S \in \mathbb{Q}^{(m-r)\times(m-r)}$;
2. If IsReg(A, U, S) = `false` then return("error: data are not generic");
3. else return LowRankRec(A, r).

The subroutine LowRankRec that is called after the control on regularity assumptions, is recursive.

LowRankRec(A, r)

Input: A $n-$variate linear matrix $A(x)$ of size m, encoded by the $m^2(n+1)$ rational entries of $A_0, A_1, \ldots, A_n$, and an integer $0 \leq r \leq m-1$;

Output: Either an error message stating that genericity assumptions are not satisfied, or a rational parametrization $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \in \mathbb{Q}[t]^{n+2}$ such that for every connected component $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$, there exists $t_0 \in Z_{\mathbb{R}}(q_{n+1})$ with $(q_1(t_0)/q_0(t_0), \ldots, q_n(t_0)/q_0(t_0)) \in \mathcal{C}$.

Procedure:

1. If $n < (m-r)^2$ then return an empty list;
2. If $n = (m-r)^2$ then return $\mathsf{Project}(\mathsf{RatPar}(f(A, U, S)), x)$;
3. Choose randomly $M \in \mathrm{GL}(n, \mathbb{Q})$, $v \in \mathbb{Q}^{(2m-r)(m-r)}$;
4. $\mathsf{P} \leftarrow \mathsf{Image}(\mathsf{Project}(\mathsf{RatPar}(l(A \circ M, U, S, v)), x), M^{-1})$;
5. Choose randomly $t \in \mathbb{Q}$; $A \leftarrow (A_0 + tA_1, A_2, \ldots, A_n)$;
6. $\mathsf{Q} \leftarrow \mathsf{Lift}(\mathsf{LowRankRec}(A, r), t)$;
7. return $\mathsf{Union}(\mathsf{Q}, \mathsf{P})$.

It takes as input the couple (A, r). Then:

- if $n < (m-r)^2$ it returns an empty list: recall that Proposition 3.5 implies that, for generic input data, $\mathcal{D}_r = \emptyset$;

- if $n = (m-r)^2$, it directly computes a rational parametrization of the projection of the finite set $\mathcal{V}_r$ on $(x_1, \ldots, x_n)$, which coincides with $\mathcal{D}_r$;

- if $n > (m-r)^2$, it computes a rational parametrization q of the polynomial ideal $\langle \ell(A \circ M, U, S, v) \rangle$, where M, v are chosen randomly, by applying RatPar to the generators of the Lagrange ideal and then Project to eliminate Lagrange multipliers z and kernel variables y. Then it chooses a random value $t \in \mathbb{Q}$ and recalls recursively LowRankRec with input $(\widetilde{A}, r)$ where $\widetilde{A}$ is obtained by setting x_1 to t in $A \circ M$.

Before intersecting the incidence variety with the hyperplane defined by $x_1 - t = 0$, eliminating one variable and calling the recursive subroutine (*cf.* Step 5 and 6 of LowRankRec), the algorithm recovers the solutions in the previous coordinate system by applying Image with input q and M^{-1} (*cf.* Step 4).

After the recursive call, the subroutines Lift and Union recover a rational parametrization of a finite subset of $\mathbb{R}^n$ containing one point per connected component of $\mathcal{D}_r \cap \mathbb{R}^n$ (*cf.* Step 6 and 7).

3.3.4 Correctness

To prove that LowRank is correct, we have to show that with input (A, r) such that the genericity hypotheses are satisfied, it returns a correct output (after a finite number of arithmetic operations), and that, otherwise, it returns an error message stating that the necessary assumptions are not satisfied.

Intermediate results

The correctness theorem is proved in Theorem 3.9 and follows from intermediate results that are described next. The first result is a regularity theorem for the incidence varieties $\mathcal{V}_r$, and its proof is in Section 3.5.1, page 97.

Proposition 3.6. *Let $m, n, r \in \mathbb{N}$, with $0 \leq r \leq m-1$.*

1. *There exists a non-empty Zariski-open set $\mathscr{A} \subset \mathbb{M}_{m,m}^{n+1}(\mathbb{C})$ such that if $A \in \mathscr{A} \cap \mathbb{M}_{m,m}^{n+1}(\mathbb{Q})$, A satisfies G;*
2. *If A satisfies G, there exists a non-empty Zariski open set $\mathscr{T} \subset \mathbb{C}$ such that if $t \in \mathscr{T} \cap \mathbb{Q}$, and $\tilde{A} = (A_0 + tA_1) + x_2A_2 + \cdots + x_nA_n$, then $\tilde{A}$ satisfies G.*

The second result deals with Lagrange systems $\ell(A \circ M, U, S, v)$. We prove that its set of solutions is finite and that ℓ satisfies G_2. The proof of this result is contained in Section 3.5.3, page 103.

Proposition 3.7. *Let $A \in \mathbb{M}_{m,m}^{n+1}(\mathbb{C})$ be such that A satisfies G. Let $c = (2m - r)(m-r)$. Then there exist non-empty Zariski open sets $\mathscr{V} \subset \mathbb{C}^c$ and $\mathscr{M}_1 \subset \mathrm{GL}_n(\mathbb{C})$ such that, if $v \in \mathscr{V} \cap \mathbb{Q}^c$ and $M \in \mathscr{M}_1 \cap \mathbb{M}_{n,n}(\mathbb{Q})$, for $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$ and $U \in \mathbb{M}_{m-r,m}(\mathbb{Q})$ full rank, the following holds*

1. *$\mathcal{Z}(A \circ M, U, S, v)$ is finite and $\ell(A \circ M, U, S, v)$ satisfies G_2;*
2. *the projection of $\mathcal{Z}(A \circ M, U, S, v)$ on the space of variables (x, y) contains the set of critical points of the restriction of π_1 to $\mathcal{V}_r$.*

The last proposition states that in generic coordinates the image of each connected component $\mathcal{C}$ of $\mathcal{D}_r \cap \mathbb{R}^n$ by each map $\pi_i(x) = (x_1, \ldots, x_i)$ is a closed subset of the real line $\mathbb{R}$ (for the Euclidean topology). In addition it states that the pre-images of values $t \in \mathbb{R}$ lying on the boundary of $\pi_1(\mathcal{C})$ can be computed as projections of critical points of π_1 restricted to $\mathcal{V}_r$. Its proof is in Section 3.5.4, page 107.

Proposition 3.8. *There exist non-empty Zariski open sets $\mathscr{U} \subset \mathbb{M}_{m-r,m}(\mathbb{C})$ and $\mathscr{M}_2 \subset \mathrm{GL}_n(\mathbb{C})$ such that if $U \in \mathscr{U} \cap \mathbb{M}_{m-r,m}(\mathbb{Q})$ and $M \in \mathscr{M}_2 \cap \mathbb{M}_{n,n}(\mathbb{Q})$, and if A satisfies G, and $d = \dim \mathcal{D}_r$, then for any connected component $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$, the following holds:*

1. *for $i = 1, \ldots, d$, the image $\pi_i(M^{-1}\mathcal{C})$ is closed;*
2. *if $t \in \mathbb{R}$ lies on the boundary of $\pi_1(M^{-1}\mathcal{C})$ then $\pi_1^{-1}(t) \cap M^{-1}\mathcal{C}$ is finite and there exists $(x, y) \in \mathcal{V}_r(A \circ M, U, S)$ such that $\pi_1(x, y) = t$.*

Logical dependencies of Zariski open sets

Propositions 3.6, 3.7 and 3.8 suggest that in order to prove that $\mathsf{LowRank}$ is correct, both input data and parameters chosen internally must belong to predetermined dense open subsets of their parameter spaces. We formalize next this fact, also explicitating the hidden dependencies between the choice of different parameters.

We recall that $\mathscr{A} \subset \mathbb{M}^{n+1}_{m,m}(\mathbb{C})$ denotes the non-empty Zariski open set defined in Proposition 3.6. Remark that, since the set $\mathscr{A}$ is obtained by the regularity result of Proposition 3.6 via a non-constructive proof, we cannot explicitly give the equations defining the complement of this Zariski open set. Hence, we cannot *a priori* verify whether A belongs to $\mathscr{A}$ or not. Conversely, whether or not A verifies G can be easily checked by testing the emptiness of complex algebraic sets, for example via Gröbner bases, using the Jacobian criterion (*cf.* Theorem 1.9 on page 37).

Furthermore, remark that the number of recursive calls of $\mathsf{LowRankRec}$ is $n-(m-r)^2$ (0 if $n<(m-r)^2$). At the first call of $\mathsf{LowRankRec}$, the input matrix is linear in n variables, at the first recursive call it is linear in $n-1$ variables, at the last call it is linear in $(m-r)^2$ variables. We denote by:

- $\mathscr{T}^{(j)}, \mathscr{U}^{(j)}, \mathscr{M}_1^{(j)}, \mathscr{M}_2^{(j)}$ and $\mathscr{V}^{(j)}$ the non-empty Zariski open sets defined by Propositions 3.6, 3.7 and 3.8, associated to the $(n-j+1)-$th recursion call of $\mathsf{LowRankRec}$;
- $M^{(j)}, v^{(j)}$ and $t^{(j)}$ the parameters chosen respectively at Step 3 and 5 of the $(n-j+1)-$th recursion call of $\mathsf{LowRankRec}$;
- $(A^{(j)}, r)$, the input of the $(n-j+1)-$th recursion call of $\mathsf{LowRankRec}$.

The existence of $\mathscr{T}^{(j)}$ at each call is guaranteed in Proposition 3.6, as follows:

$$A^{(j)} \text{ satisfying } \mathsf{G} \implies \exists\, \mathscr{T}^{(j)} \subset \mathbb{C},$$

while that of $\mathscr{U}^{(j)}, \mathscr{M}_1^{(j)}, \mathscr{M}_2^{(j)}$ and $\mathscr{V}^{(j)}$ follow similarly from Propositions 3.7:

$$\begin{aligned} A^{(j)} \text{ satisfying } \mathsf{G} &\implies \exists\, \mathscr{V}^{(j)} \subset \mathbb{C}^c, \text{ and} \\ A^{(j)} \text{ satisfying } \mathsf{G} &\implies \exists\, \mathscr{M}_1^{(j)} \subset \mathrm{GL}_j(\mathbb{C}), \end{aligned}$$

and Proposition 3.8:

$$\begin{aligned} A^{(j)} \text{ satisfying } \mathsf{G} &\implies \exists\, \mathscr{U}^{(j)} \subset \mathbb{M}_{m-r,r}(\mathbb{C}), \text{ and} \\ A^{(j)} \text{ satisfying } \mathsf{G} &\implies \exists\, \mathscr{M}_2^{(j)} \subset \mathrm{GL}_j(\mathbb{C}). \end{aligned}$$

Hypotheses on input data and parameters

We refer to notation introduced in the previous section about Zariski open sets $\mathscr{A}$, $\mathscr{T}^{(j)}$, $\mathscr{U}^{(j)}$, $\mathscr{M}_1^{(j)}$, $\mathscr{M}_2^{(j)}$ and $\mathscr{V}^{(j)}$. Let $(A, r) \in \mathbb{M}^{n+1}_{m,m}(\mathbb{Q}) \times \mathbb{Z}$ be the input of LowRank, and U, S be parameters chosen at Step 1. The first hypothesis is about input data A and parameters U, S.

Hypothesis H_1

H_1 holds if

- U is full-rank and $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$;
- $A \in \mathscr{A} \cap \mathbb{M}^{n+1}_{m,m}(\mathbb{Q})$.

Now, parameter U is chosen once (at Step 1 of LowRank) but one has to ensure that it belongs to sets $\mathscr{U}^{(j)}$ for all j; conversely, parameters $M^{(j)}, v^{(j)}$ and $t^{(j)}$ are chosen differently at each call. Finally, the choices of random parameters performed during the whole algorithm should satisfy the following hypothesis.

Hypothesis H_2

H_2 holds if, for all $j = (m-r)^2, \ldots, n$

- $U \in \mathscr{U}^{(j)} \cap \mathbb{M}_{m-r,m}(\mathbb{Q})$;
- $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$;
- $M^{(j)} \in \mathscr{M}_1^{(j)} \cap \mathscr{M}_2^{(j)} \cap \mathbb{M}_{j,j}(\mathbb{Q})$;
- $v^{(j)} \in \mathscr{V}^{(j)} \cap \mathbb{Q}^{(2m-r)(m-r)}$;
- $t^{(j)} \in \mathscr{T}^{(j)} \cap \mathbb{Q}$.

Correctness theorem

We can prove the correctness theorem, which assumes Hypothesis H_1 and H_2. Let $A \in \mathbb{M}^{n+1}_{m,m}(\mathbb{Q})$ be a $n-$variate linear matrix with rational coefficients, and let $0 \leq r \leq m-1$. Let $U, S, t^{(j)}, M^{(j)}$ and $v^{(j)}$ be the parameters chosen during LowRank.

Theorem 3.9. *Let (A, r) be the input of* LowRank*. If H_1 and H_2 hold, then the output of* LowRank *is a rational parametrization whose associated set of solutions intersects each connected component of $\mathcal{D}_r \cap \mathbb{R}^n$.*

Proof. The proof is by induction on n, the number of variables appearing in A. We first eliminate some basic case. If $n < (m-r)^2$, since H_1 holds, then the set $\mathcal{D}_r$ is empty. Hence $\mathcal{D}_r \cap \mathbb{R}^n$ is empty and the algorithm returns a correct output (the empty list). Thereafter, we proceed by induction on $n \geq (m-r)^2$.

For $n = (m-r)^2$, since H_1 holds, Proposition 3.6 implies that $\mathcal{D}_r$ and $\mathcal{V}_r$ are finite. We deduce that RatPar returns a rational parametrization of $\mathcal{V}_r$ and the routine Project returns a parametrization of $\mathcal{D}_r$.

Let $n > (m-r)^2$. Let U, S be the random matrices chosen at Step 1 of $\mathsf{LowRank}$. The induction hypothesis implies that for any $(n-1)-$variate linear matrix $\tilde{A}$ such that $(\tilde{A}, U, S)$ satisfies G, $\mathsf{LowRank}$ returns the expected output when H_1 and H_2 hold. Let A be the input matrix of size m, n variables, and let $0 \leq r \leq m-1$. Let $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$ be a connected component. Let $M \in \mathrm{GL}_n(\mathbb{C})$ be the matrix chosen at Step 3 of $\mathsf{LowRankRec}$ with input (A, r), and consider the projection map $\pi_1 \colon (x_1, \ldots, x_n) \to x_1$ restricted to $\mathcal{V}_r$. Since H_1 and H_2 hold, by Proposition 3.8, $\pi_1(M^{-1}\mathcal{C})$ is closed, hence either $\pi_1(M^{-1}\mathcal{C}) = \mathbb{R}$ or $\pi_1(M^{-1}\mathcal{C})$ is a closed set with non-empty boundary. We claim that, in both cases, $\mathsf{LowRank}$ with input (A, r) returns a point which lies in the connected component $M^{-1}\mathcal{C}$. This is proved next.

First case. Suppose $\pi_1(M^{-1}\mathcal{C}) = \mathbb{R}$. We deduce that for all $t \in \mathbb{Q}$, the set $\pi_1^{-1}(t)$ intersects $M^{-1}\mathcal{C}$, so $\pi_1^{-1}(t) \cap M^{-1}\mathcal{C} \neq \emptyset$. In particular this is true when t is the parameter chosen at Step 5 of $\mathsf{LowRankRec}$ with input (A, r). Let $A^{(n-1)}$ be the $(n-1)-$variate linear matrix obtained from $A \circ M$ by substituting $x_1 = t$. Observe that $\pi_1^{-1}(t) \cap M^{-1}\mathcal{C}$ is the union of some connected components of the determinantal variety $\mathcal{D}_r^{(n-1)} \cap \mathbb{R}^{n-1} = \{x \in \mathbb{R}^{n-1} : \text{rank } A^{(n-1)} \leq r\}$ (understood as a subset of the real hyperplane $\pi_1^{-1}(t) \cap \mathbb{R}^n$). Since H_1 holds, $(A^{(n-1)}, U, S)$ satisfies G; we deduce by the induction hypothesis that the $\mathsf{LowRankRec}$ returns a rational parametrizations whose solutions contain one point in each connected component of $\mathcal{D}_r^{(n-1)} \cap \mathbb{R}^{n-1}$, and so at least one point in $M^{-1}\mathcal{C}$. The subroutines Lift and Union at Step 6 and 7 conclude the proof.

Second case. Suppose $\pi_1(M^{-1}\mathcal{C}) \neq \mathbb{R}$. By Proposition 3.8, $\pi_1(M^{-1}\mathcal{C})$ is closed. Since $M^{-1}\mathcal{C}$ is connected, $\pi_1(M^{-1}\mathcal{C})$ is a closed interval, and since $\pi_1(M^{-1}\mathcal{C}) \neq \mathbb{R}$ there exists t in the boundary of $\pi_1(M^{-1}\mathcal{C})$ such that $\pi_1(M^{-1}\mathcal{C}) \subset [t, +\infty)$ or $\pi_1(M^{-1}\mathcal{C}) \subset (-\infty, t]$. Suppose without loss of generality that $\pi_1(M^{-1}\mathcal{C})$ is contained in $[t, +\infty)$, so that t is a local minimum of π_1, attained on $M^{-1}\mathcal{C}$.

Let $x = (t, x_2, \ldots, x_n) \in M^{-1}\mathcal{C}$. By Proposition 3.8, there exists $y \in \mathbb{C}^{m(m-r)}$ such that $(x, y) \in \mathcal{V}_r$. We claim that there exists $z \in \mathbb{C}^{(2m-r)(m-r)}$ such that (x, y, z) lies in $\mathcal{Z}(A \circ M, U, S, v)$. Then, we conclude that the point $x \in M^{-1}\mathcal{C}$ appears among the solutions of the rational parametrization P obtained at Step 4 of $\mathsf{LowRankRec}$: correction of the algorithm follows. The claim is proved next.

Let $\mathcal{C}' \subset \mathcal{V}_r \cap \mathbb{R}^{n+m(m-r)}$ be the connected component such that $M^{-1}\mathcal{C}' \subset M^{-1}\mathcal{V}_r \cap \mathbb{R}^{n+m(m-r)}$ contains (x, y). First, we prove that $t = \pi_1(x, y)$ lies on the boundary of $\pi_1(M^{-1}\mathcal{C}')$. Indeed, suppose that there exists $(\widetilde{x}, \widetilde{y}) \in M^{-1}\mathcal{C}'$ such that $\pi_1(\widetilde{x}, \widetilde{y}) < t$. Since $M^{-1}\mathcal{C}$ is connected, and since it is a connected component of a real algebraic set, there exists a continuous semi-algebraic map $\tau \colon [0, 1] \to M^{-1}\mathcal{C}'$ with $\tau(0) = (x, y)$ and $\tau(1) = (\widetilde{x}, \widetilde{y})$. Let $\pi_x \colon \mathbb{R}^{n+m(m-r)} \to \mathbb{R}^n$ be the map $\pi_x(x, y) = x$. Then $\pi_x \circ \tau$ is the composition of continuous semi-algebraic maps, and hence is continuous and semi-algebraic. Moreover $(\pi_x \circ \tau)(0) = x$ and $(\pi_x \circ \tau)(1) = \widetilde{x}$. Since $(\pi_x \circ \tau)(\theta) \in \mathcal{D}_r$ for all $\theta \in [0, 1]$, then $\widetilde{x} \in M^{-1}\mathcal{C}$. Since $\pi_1(\widetilde{x}) < t$ we obtain a contradiction. So $\pi_1(x, y)$ lies on the boundary of $\pi_1(M^{-1}\mathcal{C}')$.

Since (A, U, S) satisfies G, hence $(A \circ M, U, S)$ satisfies G and by the Implicit

Function Theorem [150, Th. 2.12] one deduces that (x, y) is a critical point of π_1 restricted to $\mathcal{V}_r$ and that, by Proposition 3.7, there exists $z \in \mathbb{C}^{(2m-r)(m-r)}$ such that $(x, y, z) \in \mathcal{Z}(A \circ M, U, S, v)$, as claimed. *q.e.d.*

3.4 Degree bounds and complexity

We give arithmetic complexity bounds on the number of operations in $\mathbb{Q}$ needed to perform the algorithm LowRank, and degree bounds for the rational parametrization given in terms of the output (Section 3.4.1). These essentially derive from the multilinear structure of systems encoding the critical points (Lagrange systems) and rely on the computation of a multilinear Bézout bound.

The complexity of LowRank is expressed as a function involving the size of parameters (number of variables, size of matrices and the expected rank: m, n, r) and the bounds on the degree of rational parametrizations. It is computed in Section 3.4.2.

3.4.1 The degree of the output

Multilinear bounds

To perform this analysis we suppose that A satisfies G. Let us consider the subroutine RatPar at the first recursion step of LowRank. Hence its input consists of either the generators of the ideal of the incidence variety $f(A \circ M, U, S)$, if $n = (m-r)^2$, or of the Lagrange systems $\ell(A \circ M, U, S, v)$, if $n > (m-r)^2$. In both cases, the input system is bilinear in two or three groups of variables.

If $x^{(1)}, \ldots, x^{(p)}$ are groups of variables, and $f \in \mathbb{Q}[x^{(1)}, \ldots, x^{(p)}]$, we say that the multidegree of f is $(d_1, \ldots, d_p)$ if its degree with respect to the group $x^{(j)}$ is d_j for $j = 1, \ldots, p$. For example, the multidegree of $x_1^2 y_2 x_2 + y_1^5$ with groups $x = (x_1, x_2)$ and $y = (y_1, y_2)$, is $(3, 5)$.

Proposition 3.10. *Let $A \in \mathbb{M}_{m,m}^{n+1}(\mathbb{Q})$, $0 \leq r \leq m-1$ and let U, S and M, v be respectively the parameters chosen at step 1 of* LowRank *and at step 3 of* LowRankRec*, with input (A, r). Suppose that H_1 and H_2 hold. Let q be the rational parametrization returned by* RatPar*. Then:*

1. *if $n = (m-r)^2$, the degree of q is bounded from above by*

$$\delta(m, n, r) = \binom{m(m-r)}{(m-r)^2};$$

2. *if $n > (m-r)^2$, the degree of q is bounded from above by*

$$\delta(m, n, r) = \sum_{k \in \mathcal{F}_{m,n,r}} \binom{m(m-r)}{n-k} \binom{n-1}{k+(m-r)^2-1} \binom{r(m-r)}{k},$$

with $\mathcal{F}_{m,n,r} = \{k : \max\{0, n-m(m-r)\} \leq k \leq \min\{n-(m-r)^2, r(m-r)\}\}$.

Proof of Assertion 1. Since H_1 holds, the dimension of $\mathcal{V}_r$ is zero. Consequently, the degree of q is the degree of $\mathcal{V}_r$. Since the entries of $f(A \circ M, U, S)$ are bilinear in x, y, one can compute the associated Multilinear Bézout bound, see *e.g.* [139, Ch. 11].

One can reduce the polynomial system f to an equivalent form as follows. From equations $UY(y) - S = 0$ one can eliminate $(m-r)^2$ variables $y_{i,j}$, for example those corresponding to the last $m-r$ rows of $Y(y)$. With abuse of notation, we denote by $f \subset \mathbb{Q}[x, y_{1,1}, \ldots, y_{r,m-r}]$ the polynomial system obtained after this reduction. It is constituted by $m(m-r)$ polynomials of multidegree bounded by $(1,1)$ with respect to $x = (x_1, \ldots, x_n)$ and $y = (y_{1,1}, \ldots, y_{r,m-r})$.

By [139, Prop. 11.1.1], $\deg Z_{\mathbb{C}}(f)$ is bounded by the sum of the coefficients of

$$(s_x + s_y)^{m(m-r)} \quad \mod \left\langle s_x^{n+1}, s_y^{r(m-r)+1} \right\rangle \subset \mathbb{Z}[s_x, s_y].$$

Since $n + r(m-r) = m(m-r)$, and since $(s_x + s_y)^{m(m-r)}$ is homogeneous of degree $m(m-r)$, the aforementioned bound equals the coefficient of $s_x^n s_y^{r(m-r)}$ in the expansion of $(s_x + s_y)^{m(m-r)}$, that is exactly $\binom{m(m-r)}{(m-r)^2}$. *q.e.d.*

Proof of Assertion 2. The proof is similar to that of Assertion 1. Let f be the simplified system defined above. We consider variables $z = (1, z_2, \ldots, z_{m(m-r)})$ (we can put $z_1 = 1$ since the Lagrange system is defined over the Zariski open set $z \neq 0$). The simplified Lagrange system is constituted by polynomials (f, g, h) where $(g, h) = z' D_1 f$, that is by:

- $m(m-r)$ polynomials of multidegree bounded by $(1, 1, 0)$;
- $n-1$ polynomials of multidegree bounded by $(0, 1, 1)$;
- $r(m-r)$ polynomials of multidegree bounded by $(1, 0, 1)$.

Since H_1 and H_2 hold, the zero set $Z_{\mathbb{C}}(f, g, h)$ is finite or empty, and its degree equals that of q. By [139, Prop. 11.1.1] this value is bounded from above by the sum of the coefficients of any representative of the class of

$$(s_x + s_y)^{m(m-r)} (s_y + s_z)^{n-1} (s_x + s_z)^{r(m-r)}$$

modulo $\langle s_x^{n+1}, s_y^{r(m-r)+1}, s_z^{m(m-r)} \rangle \subset \mathbb{Z}[s_x, s_y, s_z]$. As above, by homogeneity of the previous polynomial and by counting the degrees, this sum is given by the coefficient of the monomial $s_x^n s_y^{r(m-r)} s_z^{m(m-r)-1}$ in the expansion

$$\sum_{i,j,k} \binom{m(m-r)}{i} \binom{n-1}{j} \binom{r(m-r)}{k} s_x^{i+k} s_y^{m(m-r)-i+j} s_z^{n-1-j+r(m-r)-k}.$$

The coefficient is obtained by setting the equalities $i+k=n$, $m(m-r)-i+j=r(m-r)$ and $n-1-j+r(m-r)-k=m(m-r)-1$. These equalities imply $i+k=n=j+k+(m-r)^2=j+k+i-j=i+k$ and consequently one deduces the claimed expression. *q.e.d.*

The previous proposition states that the number of complex solutions computed at step 2 of LowRankRec, is at most $\delta(m,n,r)$. We deduce the following estimate on the total number of complex solutions.

Corollary 3.11. *Suppose that the hypothesis of Proposition 3.10 are satisfied. Then* LowRank *returns a rational parametrization whose degree is less than or equal to*

$$\Delta(m,n,r) = \binom{m(m-r)}{(m-r)^2} + \sum_{j=(m-r)^2+1}^{\min\{n,m^2-r^2\}} \delta(m,j,r).$$

Proof. Since H_1 holds, for $n<(m-r)^2$, $\mathcal{D}_r$ is empty and the algorithm returns the empty list. For m,j,r let $\mathcal{F}_{m,j,r}$ be the set of indices defined in Proposition 3.10. Observe that, since $j>\max\{0,j-m(m-r)\}$ always holds, then $\delta(m,j,r)=0$ if and only if $\mathcal{F}_{m,j,r}=\emptyset$, if and only if $j-m(m-r)>r(m-r)$, that is when $j>m^2-r^2$. Hence, the upper bound $\Delta(m,n,r)$ is deduced straightforwardly from bounds given in Proposition 3.10. *q.e.d.*

By the previous proof, we also deduce that to evaluate the arithmetic complexity of RatPar we can restrict the analysis to recursive calls of LowRankRec with input linear matrices $A^{(j)}$ with $(m-r)^2 \leq j \leq m^2-r^2$ (for the notation $A^{(j)}$ *cf.* page 86).

We finally remark that, as shown in Table 3.1, page 108, the bounds given by Proposition 3.10 and Corollary 3.11 are typically not optimal. Indeed, we give in Table 3.1, page 108, some numerical values of these bounds, comparing with the effective degrees. We remark that these bounds equal 0 (and hence are optimal) when the algorithm does not compute any solution.

Further estimates

We conclude the analysis on the degree-size of the output representation by giving a final estimate of the bound provided in Corollary 3.11. The following Proposition states that the output degree is essentially bounded by the product of the dimension of $\mathcal{D}_r$ and of a natural combinatorial bound.

Proposition 3.12. *For all m,n,r, with $r \leq m-1$, $\delta(m,n,r) \leq \binom{n+m(m-r)}{n}^3$. Moreover, if we let d be the dimension of $\mathcal{D}_r$, $\delta_{max}=\delta_{max}(m,r)$ be the maximum of $\delta(m,j,r)$ when j runs between $(m-r)^2$ and m^2-r^2, and $\Delta(m,n,r)$ be the bound computed in Corollary 3.11, if H_1 and H_2 hold, one has*

$$\Delta(m,n,r) \in \mathcal{O}\left(d\,\delta_{max}\right) \subset \mathcal{O}\left(d\binom{n+m(m-r)}{n}^3\right).$$

Proof. The estimate for $\delta(m,n,r)$ comes straightforwardly from the formula

$$\binom{a+b}{a}^3 = \sum_{i_1,i_2,i_3=0}^{\min(a,b)} \binom{a}{i_1}\binom{b}{i_1}\binom{a}{i_2}\binom{b}{i_2}\binom{a}{i_3}\binom{b}{i_3}$$

applied to $a = n$ and $b = m(m-r)$, and from the expression of $\delta(m,n,r)$ computed in Proposition 3.10. Moreover, by Corollary 3.11, one deduces

$$\begin{aligned}\Delta(m,n,r) &= \binom{m(m-r)}{(m-r)^2} + \sum_{j=(m-r)^2+1}^{\min\{n,m^2-r^2\}} \delta(m,j,r) \leq \\ &\leq \binom{m(m-r)}{(m-r)^2} + (n-(m-r)^2)\delta_{max}.\end{aligned}$$

Since H_1 and H_2 hold, the expected dimension of $\mathcal{D}_r$ is $d = n-(m-r)^2$, while the term $\binom{m(m-r)}{(m-r)^2}$ is negligible in the previous sum. *q.e.d.*

On the number of connected components

We finally give an upper bound on the number of connected components of the set $\mathcal{D}_r \cap \mathbb{R}^n$, for average input data. Indeed, the degree of the rational parametrization returned by $\mathsf{LowRank}$ implies such an upper bound, and has to be compared with classical bounds. Remark that for any $0 \leq r \leq m-1$, the set $\mathcal{D}_r$ is defined by at most $\binom{m}{r+1}^2$ polynomials of degree at most $r+1$: we deduce that the number of connected components of $\mathcal{D}_r \cap \mathbb{R}^n$ is in $\mathcal{O}(r)^n$ (*cf.* the bound given on page 39).

Proposition 3.13. *Suppose that H_1 and H_2 hold and that the dimension d of $\mathcal{D}_r$ is strictly positive. The maximum number of connected components of the algebraic set $\mathcal{D}_r \cap \mathbb{R}^n$ is bounded from above by $\Delta(m,n,r)$ and hence is in $\mathcal{O}(d\,\delta_{max})$. Moreover, for subfamilies of sets $\mathcal{D}_r$ where the size m is fixed, the number of connected components of $\mathcal{D}_r \cap \mathbb{R}^n$ is in $\mathcal{O}(n^{1+3m(m-r)})$.*

Proof. If Hypothesis H_1 and H_2 hold, by Corollary 3.11, $\Delta(m,n,r)$ is a bound on the degree (and hence on the cardinality of the solution set) of the output parametrization q; since q parametrizes a finite set meeting each connected component of $\mathcal{D}_r \cap \mathbb{R}^n$, we conclude. The asymptotic when m is fixed derives from Proposition 3.12 and from the following trivial facts:

- $d \leq n$;
- if m is fixed, then $\binom{n+m(m-r)}{n} \in \mathcal{O}(n^{m(m-r)})$.

q.e.d.

We recall that Hypotheses H_1 and H_2 are open conditions on the input data A, r and on the recursive choices of parameters performed by $\mathsf{LowRank}$. We deduce hence by Proposition 3.13 that for average data matrices $A \in \mathbb{M}_{m,m}^{n+1}(\mathbb{Q})$, the number of connected components of the associated real low rank locus $\mathcal{D}_r \cap \mathbb{R}^n$ is essentially the product of the dimension of $\mathcal{D}_r$ with a combinatorial bound.

We finally stress the fact that this result is part of a historical approach due first to Thom and Milnor, using the critical points method to obtain bounds for the expected number of components of real algebraic sets. Indeed, this is strictly linked to the problem of finding sample sets of real solutions of a polynomial system.

3.4.2 Complexity analysis

We recall that our complexity model takes into account the number of arithmetic operations over the ground field $\mathbb{Q}$ that are performed during LowRank. It is given by the sum of the contributions of all subroutines, that is by:

- the contribution of IsReg at step 2 of the preamble: we suppose that regularity assumptions are satisfied, so that one does not have to estimate this complexity;
- the contribution of the composition of Project and RatPar with input $f(A^{(m-r)^2}, U, S)$ at step 2 of the first call of LowRankRec;
- for j from $\min\{n, m^2 - r^2\}$ downto $(m-r)^2+1$, the contribution of RatPar, Project, Image, Lift and Union inside the $(n-j+1)-$th recursive call of LowRankRec.

Complexity of RatPar

We start by estimating the complexity of the computation of the rational parametrization. This complexity depends on the choice of the algorithm performing RatPar.

As established in Section 1.2.2, there exist different algorithms for solving this task. We prefer to refer to the homotopy algorithm in [88], because it allows to deduce complexity estimates as a direct function of the number of complex solutions that are computed. It uses symbolic homotopy techniques as in [74]. We remark that both algorithms in [88] and [53] allow to build a geometric resolution of the Lagrange system and hence to compute a rational parametrization of its solutions. Both complexities rely on geometric invariants of algebraic sets built during the procedure. We base our complexity analysis on the symbolic homotopy algorithm of [88] since it allows to directly exploit the degree bounds computed in Section 3.4.1.

In the paper [68] we used the algorithm in [53], which relies on the computation of the maximum degree of intermediate varieties of Lagrange systems: we finally observe that such degrees can be also bounded by multilinear Bézout bounds, and that their asymptotic behavior is similar to that of $\delta(m, n, r)$ (*cf.* Section 3.4.1).

We suppose that $(m-r)^2 < n \leq m^2 - r^2$, since the complexity of RatPar with input $f(A \circ M, U, S)$ is controlled by that with input $\ell(A \circ M, U, S, v)$. We bound below the complexity of the computation of the rational parametrization performed by RatPar where the input linear matrix of $\mathsf{LowRankRec}$ is $n-$variate.

We suppose that the input of RatPar is the simplified system ℓ defined in the proof of Assertion 2 of Proposition 3.10, page 89, consisting in

- $m(m-r)$ polynomials of multidegree bounded by $(1, 1, 0)$;
- $n-1$ polynomials of multidegree bounded by $(0, 1, 1)$;
- $r(m-r)$ polynomials of multidegree bounded by $(1, 0, 1)$.

We denote by

$$\begin{aligned}
\Delta_{xy} &= \{1, x_i, y_j, x_i y_j : i = 1, \ldots, n, j = 1, \ldots, r(m-r)\} \\
\Delta_{yz} &= \{1, y_j, z_k, y_j z_k : j = 1, \ldots, r(m-r), k = 2 \ldots, m(m-r)\} \\
\Delta_{xz} &= \{1, x_i, z_k, x_i z_k : i = 1, \ldots, n, k = 2, \ldots, m(m-r)\}
\end{aligned}$$

the supports of the aforementioned three groups of polynomials. Suppose, further, that $\widetilde{\ell} \subset \mathbb{Q}[x, y, z]$ is a polynomial system such that:

- the length of $\tilde{\ell}$ equals that of ℓ;
- for $i = 1, \ldots, n-1+m^2-r^2$, the support of $\tilde{\ell}_i$ equals that of ℓ_i;
- the solutions of $\tilde{\ell}$ are known.

Since ℓ consists of bilinear polynomials in three groups of variables, $\widetilde{\ell}$ can be easily built by considering products of linear forms. Indeed, from a polynomial ℓ_i, with support in Δ_a, one can build a polynomial $\widetilde{\ell_i}$ as follows:

- if $a = xy$, generate linear forms $g_{i,x} \in \mathbb{Q}[x]$, $g_{i,y} \in \mathbb{Q}[y]$, and let $\widetilde{\ell_i} = g_{i,x} g_{i,y}$;
- if $a = yz$, generate linear forms $g_{i,y} \in \mathbb{Q}[y]$, $g_{i,z} \in \mathbb{Q}[z]$, and let $\widetilde{\ell_i} = g_{i,y} g_{i,z}$;
- if $a = xz$, generate linear forms $g_{i,x} \in \mathbb{Q}[x]$, $g_{i,z} \in \mathbb{Q}[z]$, and let $\widetilde{\ell_i} = g_{i,x} g_{i,z}$.

Hence, we suppose that the solutions of $\widetilde{\ell}$ are known since they can be computed by solving systems of linear equations. Subsequently, we build the homotopy

$$t\ell + (1-t)\tilde{\ell} \subset \mathbb{Q}[x, y, z, t]. \tag{3.1}$$

where t is a new variable. The zero set of (3.1) is a $1-$dimensional algebraic set, that is a curve. From [88, Prop. 6.1], if the solutions of $\tilde{\ell}$ are known, one can compute a rational parametrization of the solution set of system (3.1) within $\mathcal{O}((\tilde{n}^2 N \log Q + \tilde{n}^{\omega+1})ee')$ arithmetic operations over $\mathbb{Q}$, where:

- $\tilde{n}$ is the number of variables in ℓ;

- $N = m(m-r)\#\Delta_{xy} + (n-1)\#\Delta_{yz} + r(m-r)\#\Delta_{xz}$;
- $Q = \max\{\|q\| : q \in \Delta_{xy} \cup \Delta_{yz} \cup \Delta_{xz}\}$;
- e is the number of isolated solutions of ℓ;
- e' is the degree of the curve $Z_{\mathbb{C}}(t\ell + (1-t)\tilde{\ell})$;
- ω is the exponent of matrix multiplication.

The theoretical definition of ω is pretty technical and can be sketched as follows. The exponent of matrix multiplication is the infimum of the set of real numbers ω sharing the following property: two $n \times n$ matrices can be computed within $\mathcal{O}(n^{\omega+\epsilon})$ many multiplications, for any $\epsilon > 0$. We briefly recall that $\omega \leq \log_2 7 \approx 2.807$ by Strassen's algorithm (*cf.* [151]) and that the best upper bound is now set to 2.3728639, thanks to Le Gall [97].

We first prove the following technical lemma. We refer to notation introduced in Proposition 3.10, page 88. Further, we use the result of Lemma 3.14 to compute the complexity of RatPar in Proposition 3.15, page 96.

Lemma 3.14. *Let $\mathcal{F}_{m,n,r}$ be the set defined in Proposition 3.10, and suppose $\mathcal{F}_{m,n,r} \neq \emptyset$. Let $\delta(m,n,r)$ be the bound defined in Proposition 3.10, and let e' be the degree of $Z_{\mathbb{C}}(t\ell + (1-t)\tilde{\ell})$. Hence*

$$e' \in \mathcal{O}\left((n + m^2 - r^2) \min\{n, m(m-r)\}\, \delta(m,n,r)\right).$$

Proof. The system $t\ell + (1-t)\tilde{\ell}$ is bilinear with respect to groups of variables x, y, z, t. We exploit this bilinearity by computing Multilinear Bézout bounds, as in Proposition 3.10, page 88.

By [139, Prop. 11.1.1], e' is bounded by the sum of the coefficients of

$$q = (s_x + s_y + s_t)^{m(m-r)}(s_y + s_z + s_t)^{n-1}(s_x + s_z + s_t)^{r(m-r)}$$

modulo $I = \langle s_x^{n+1}, s_y^{r(m-r)+1}, s_z^{m(m-r)}, s_t^2 \rangle \subset \mathbb{Z}[s_x, s_y, s_z, s_t]$. It is easy to check that $q = q_1 + s_t(q_2 + q_3 + q_4) + g$ with s_t^2 that divides g and

$$\begin{aligned}
q_1 &= (s_x + s_y)^{m(m-r)}(s_y + s_z)^{n-1}(s_x + s_z)^{r(m-r)} \\
q_2 &= m(m-r)s_t(s_x + s_y)^{m(m-r)-1}(s_y + s_z)^{n-1}(s_x + s_z)^{r(m-r)} \\
q_3 &= (n-1)s_t(s_x + s_y)^{m(m-r)}(s_y + s_z)^{n-2}(s_x + s_z)^{r(m-r)} \\
q_4 &= r(m-r)s_t(s_x + s_y)^{m(m-r)}(s_y + s_z)^{n-1}(s_x + s_z)^{r(m-r)-1},
\end{aligned}$$

whence $q \equiv q_1 + q_2 + q_3 + q_4 \mod I$. The stated bound is given by the sum of the contributions and follows straightforwardly by the estimates given below.

Contribution of q_1. It is the sum of its coefficients modulo the ideal

$$I' = \langle s_x^{n+1}, s_y^{r(m-r)+1}, s_z^{m(m-r)} \rangle.$$

This has been already computed in Proposition 3.10, and coincides with $\delta(m,n,r)$.

Contribution of q_2. Write $q_2 = m(m-r)s_t\tilde{q}_2$ with $\tilde{q}_2 \in \mathbb{Z}[s_x, s_y, s_z]$. Consequently we need to compute the sum of the coefficients of $\tilde{q}_2$, modulo I' (defined above), multiplied by $m(m-r)$. Observe that $\deg \tilde{q}_2 = n-2+m^2-r^2$ and that maximal powers admissible modulo I' are $s_x^n, s_y^{r(m-r)}, s_z^{m(m-r)-1}$. Hence, three configurations give a contribution.

(A) The coefficient of the monomial $s_x^{n-1}s_y^{r(m-r)}s_z^{m(m-r)-1}$ in $\tilde{q}_2$, that is

$$\Sigma_A = \sum_{k=0}^{r(m-r)} \binom{m(m-r)-1}{n-1-k}\binom{n-1}{k-1+(m-r)^2}\binom{r(m-r)}{k}.$$

(B) The coefficient of the monomial $s_x^{n}s_y^{r(m-r)-1}s_z^{m(m-r)-1}$ in $\tilde{q}_2$, that is

$$\Sigma_B = \sum_{k=0}^{r(m-r)} \binom{m(m-r)-1}{n-k}\binom{n-1}{k-1+(m-r)^2}\binom{r(m-r)}{k}.$$

(C) The coefficient of the monomial $s_x^{n}s_y^{r(m-r)}s_z^{m(m-r)-2}$ in $\tilde{q}_2$, that is

$$\Sigma_C = \sum_{k=0}^{r(m-r)} \binom{m(m-r)-1}{n-k}\binom{n-1}{k-2+(m-r)^2}\binom{r(m-r)}{k}.$$

The contribution of q_2 equals $m(m-r)(\Sigma_A+\Sigma_B+\Sigma_C)$, and one can easily check that $\Sigma_A \leq \delta(m,n,r)$ and $\Sigma_B \leq \delta(m,n,r)$. Remember that we assume that $\delta(m,n,r) > 0$, whence $\mathcal{F}_{m,n,r} \neq \emptyset$. We claim that $\Sigma_C \leq (1+\min\{n, m(m-r)\})\,\delta(m,n,r)$. Consequently, we conclude that the contribution of q_2 is

$$m(m-r)(\Sigma_A+\Sigma_B+\Sigma_C) \in O\left(m(m-r)\,\min\{n, m(m-r)\}\,\delta(m,n,r)\right).$$

Let us prove this claim. First, write

$$\begin{aligned}
\chi_1 &= \max\{0, n-m(m-r)\} & \chi_2 &= \min\{r(m-r), n-(m-r)^2\}\\
\alpha_1 &= \max\{0, n+1-m(m-r)\} & \alpha_2 &= \min\{r(m-r), n+1-(m-r)^2\}
\end{aligned}$$

for the indices such that $\delta(m,n,r)$ sums over $\chi_1 \leq k \leq \chi_2$ and Σ_C over $\alpha_1 \leq k \leq \alpha_2$. Remark that $\chi_1 \leq \alpha_1$ and $\chi_2 \leq \alpha_2$. Finally, denote by $\varphi(k)$ the $k-$th term in the sum defining Σ_C, and by $\gamma(k)$ the $k-$th term in the sum defining $\delta(m,n,r)$.

For all indices k admissible for both $\delta(m,n,r)$ and Σ_A, that is for $\alpha_1 \leq k \leq \chi_2$, one gets, by basic properties of binomial coefficients, that

$$\varphi(k) \leq \Psi(k)\,\gamma(k) \qquad \text{with } \Psi(k) = \frac{n-k}{m(m-r)-n+1+k}.$$

When k runs over all admissible indices, the rational function $\Psi(k)$ is non-decreasing monotone, and its maximum is attained in $\Psi(\chi_2)$ and is bounded by $\min\{n, m(m-r)\}$. Three possible cases can occur:

1. $\alpha_1 = 0$. Hence $\chi_1 = 0$, $\alpha_2 = r(m-r)$ and $\chi_2 = r(m-r)$. We deduce straightforwardly from the above discussion that $\Sigma_C \leq \min\{n, m(m-r)\}\delta(m,n,r)$;

2. $\alpha_1 = n - m(m-r) + 1$ and $\chi_1 = n - m(m-r)$. We deduce that $\chi_2 = \alpha_2 = r(m-r)$ and that $\Sigma_C = \sum_{k=\alpha_1}^{\chi_2} \varphi(k) \leq \varphi(\alpha_1) + \min\{n, m(m-r)\}\,\delta(m,n,r) \leq (1 + \min\{n, m(m-r)\})\,\delta(m,n,r)$;

3. $\chi_1 = 0$ and $\alpha_1 = n - m(m-r) + 1$. Hence, we deduce the chain of inequalities $0 \leq n - m(m-r) + 1 \leq 1$. Hence, either this case coincides with case 2 (if $n = m(m-r)$) or we deduce that $n = m(m-r) - 1$, and we fall into case 1.

Contributions of q_3 and q_4. Following exactly the same path as in the case of q_2, one respectively deduces that:

- the contribution of q_3 is in $\mathcal{O}\left(n \min\{n, m(m-r)\}\,\delta(m,n,r)\right)$;
- that of q_4 is in $\mathcal{O}\left(r(m-r) \min\{n, m(m-r)\}\,\delta(m,n,r)\right)$.

q.e.d.

Proposition 3.15. *Let $n > (m-r)^2$. Let $A \in \mathbb{M}_{m,m}^{n+1}(\mathbb{Q})$, $0 \leq r \leq m-1$ and let M, U, S, v be the parameters chosen before the first recursive step of* LowRank*. Let $\delta(m,n,r)$ be the bound defined in Proposition 3.10. Then, if* H_1 *and* H_2 *hold,* RatPar *returns a rational parametrization within*

$$\widetilde{\mathcal{O}}\left((n + m^2 - r^2)^7\,\delta(m,n,r)^2\right)$$

arithmetic operations.

Proof. Following the notation introduced on page 93, $\tilde{n} = n - 1 + m^2 - r^2$, the bound for e is $\delta(m,n,r)$ and is given in Proposition 3.10 and a bound for e' is given in Lemma 3.14, and is in $\widetilde{\mathcal{O}}(\tilde{n}^2\delta)$. Moreover, $N \in O(nmr(m-r)^2)$, and hence $N \in O(\tilde{n}^3)$. The proof follows from [88, Prop. 6.1], since the maximum diameter of $\Delta_{xy}, \Delta_{yz}, \Delta_{xz}$ is bounded above by $\tilde{n}$, that is $Q \leq \tilde{n}$. *q.e.d.*

Complexity of subroutines

For these complexity bounds, we refer to those given in [139, Lem. 10.1], [139, Lem. 10.3], [139, Lem. 10.5] and [139, Lem. 10.6], from which they are obtained straightforwardly.

Proposition 3.16. *Let $\delta(m,n,r)$ be the bound defined in Proposition 3.10. At the first recursion step of* LowRankRec*, the following holds. The complexity of*

- Project *is in* $\mathcal{O}^\sim\left((n+m^2-r^2)^2\,\delta(m,n,r)^2\right)$;
- Lift *is in* $\mathcal{O}^\sim\left((n+m^2-r^2)\,\delta(m,n,r)^2\right)$;
- Image *is in* $\mathcal{O}^\sim\left((n+m^2-r^2)^2\,\delta(m,n,r)+(n+m^2-r^2)^3\right)$;
- Union *is in* $\mathcal{O}^\sim\left((n+m^2-r^2)\,\delta(m,n,r)^2\right)$.

Complexity theorem

The estimates provided in this section yield the following complexity theorem.

Theorem 3.17. *Suppose that* H_1 *and* H_2 *hold. Then* LowRank *with input* (A, r) *returns a rational parametrization of a finite set meeting each connected component of* $\mathcal{D}_r \cap \mathbb{R}^n$ *within*

$$\mathcal{O}^\sim\left((n+m^2-r^2)^7\binom{n+m(m-r)}{n}^6\right)$$

arithmetic operations over $\mathbb{Q}$.

Proof. This complexity estimate is obtained by applying:

- the complexity bound for RatPar, computed in Proposition 3.15;
- the complexity bounds for the minor subroutines, computed in Proposition 3.16;
- the bound $\delta(m,n,r) \leq \binom{n+m(m-r)}{n}^3$ computed in Proposition 3.12.

q.e.d.

3.5 Proofs of regularity

This section hosts the proof of Proposition 3.6, 3.7 and 3.8 shown, in this order, in Sections 3.5.1, 3.5.3 and 3.5.4. We provide in Section 3.5.2 a local description of the main algebraic sets built during the execution of LowRank.

3.5.1 Regularity of incidence varieties

We prove next Proposition 3.6. Both Zariski open sets $\mathscr{A}$ and $\mathscr{T}$ will be constructed by successive intersection of three non-empty Zariski open sets where respectively $\mathsf{G}_1, \mathsf{G}_2$ and G_3 hold.

Proof of G_1 *in Assertion 1.* Let $\sigma_r \subset \mathbb{M}_{m,m}(\mathbb{C})$ denote the set of $m \times m$ matrices of rank $\leq r$. By [28, Prop. 1.1] its singular locus is σ_{r-1}. For all $A \in \mathbb{M}_{m,m}^{n+1}(\mathbb{Q})$, the set $\mathcal{D}_r$ is the intersection of σ_r with the linear space $\mathcal{L} = A_0 + \langle A_1, \ldots, A_n \rangle$. By Bertini's theorem [62, Th. 17.16], if $\mathcal{L}$ is generic, the following holds:

$$\operatorname{sing} \mathcal{D}_r = \operatorname{sing} (\mathcal{L} \cap \sigma_r) = \mathcal{L} \cap \operatorname{sing} \sigma_r = \mathcal{L} \cap \sigma_{r-1} = \mathcal{D}_{r-1}.$$

We conclude that there exists a non-empty Zariski open set $\mathscr{A}_1 \subset \mathbb{C}^{m^2(n+1)}$ such that if $A \in \mathscr{A}_1$ then A satisfies Property G_1. *q.e.d.*

Proof of G_1 *in Assertion 2.* By hypothesis, that A satisfies G. In particular A satisfies G_1. By Sard's Lemma [139, Sec. 4.2], there exists a non-empty Zariski open set $\mathscr{T}_1 \subset \mathbb{C}$ such that if $t \in \mathscr{T}_1$, then a point in $\mathcal{D}_r \cap Z_{\mathbb{C}}(x_1 - t)$ is regular if and only if it is regular in $\mathcal{D}_r$. Then, for $t \in \mathscr{T}_1$, the $(n-1)-$variate liner matrix $\tilde{A}$ obtained by instantiating x_1 to t in A satisfies G_1. *q.e.d.*

Proof of G_2 *in Assertion 1.* Let $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$ and let $U \in \mathbb{M}_{m-r,m}(\mathbb{Q})$ be full-rank. We denote by $a_{\ell,i,j}$ the entry of the matrix A_ℓ at row i and column j, for $\ell = 0, 1, \ldots, n$, $i = 1, \ldots, m$ and $j = 1, \ldots, m$. We define the following map:

$$\begin{array}{rrcl} \varphi: & \mathbb{C}^n \times \mathbb{C}^{m(m-r)} \times \mathbb{M}^{n+1}_{m,m}(\mathbb{C}) & \longrightarrow & \mathbb{C}^{m(m-r)+(m-r)^2} \\ & (x, y, A) & \longmapsto & f(A, U, S) \end{array}$$

and, for a fixed $A \in \mathbb{M}^{n+1}_{m,m}(\mathbb{C})$, its section map

$$\begin{array}{rrcl} \varphi_A: & \mathbb{C}^n \times \mathbb{C}^{m(m-r)} & \longrightarrow & \mathbb{C}^{m(m-r)+(m-r)^2} \\ & (x, y) & \longmapsto & f(A, U, S). \end{array}$$

We divide the proof in two cases.

First case. Suppose that $\varphi^{-1}(0) = \emptyset$. This implies that, for any $A \in \mathbb{M}^{n+1}_{m,m}(\mathbb{C})$, $\mathcal{V}_r(A, U, S) = \emptyset$. By the Nullstellensatz (*cf.* [35, Ch. 8] or page 36), this implies that for any $A \in \mathbb{M}^{n+1}_{m,m}(\mathbb{C})$, $\langle f(A, U, S) \rangle = \langle 1 \rangle$ which is a radical ideal. In this case we conclude by defining $\mathscr{A}_2 = \mathbb{C}^{m^2(n+1)}$.

Second case. Suppose that $\varphi^{-1}(0) \neq \emptyset$. We prove that there exists a non-empty Zariski open set $\mathscr{A}_2 \subset \mathbb{M}^{n+1}_{m,m}(\mathbb{C})$ such that if $A \in \mathscr{A}_2$, the Jacobian matrix of $f(A, U, S)$ has maximal rank at each point of $\varphi_A^{-1}(0)$. As explained on page 80, this implies G_2.

We claim that 0 is a regular value of φ, *i.e.* at any point of the fiber $\varphi^{-1}(0)$ the Jacobian matrix associated to $f(A, U, S)$ (with respect to $a_{\ell,i,j}$, $x = (x_1, \ldots, x_n)$ and $y = (y_{1,1}, \ldots, y_{m,m-r})$) has maximal rank. By Thom's Algebraic Weak Transversality theorem (*cf.* [139, Sec. 4.2] or page 41) we conclude that there exists a non-empty Zariski open set $\mathscr{A}_2 \subset \mathbb{M}^{n+1}_{m,m}(\mathbb{C})$ such that, for every $A \in \mathscr{A}_2$, 0 is a regular value of the induced map φ_A, which implies the previous claim. We prove this new claim in the sequel.

Let $(x, y, A) \in \varphi^{-1}(0)$. Let us isolate the maximal minor of the Jacobian matrix Df obtained as follows:

- the $m(m-r) \times m^2$ block $\partial(A(x)Y(y))_{i,j} / \partial a_{0,k,\ell}$ of derivatives of entries of

$A(x)Y(y)$ with respect to $\{a_{0,k,\ell} : k, \ell = 1, \ldots, m\}$;

- the $(m-r)^2 \times m(m-r)$ block $\partial(UY(y)-S)/\partial y$ of derivatives of polynomial entries of $UY(y) - S$ with respect to variables y; remark also that these polynomials do not depend on a.

Up to permutations of rows and columns, the first block is a $m(m-r) \times m^2$ block-diagonal matrix, with m blocks of size $(m-r) \times m$ on the diagonal all equal to $Y(y)'$; the second block, up to re-ordering polynomials and variables, is a $(m-r)^2 \times m(m-r)$ matrix with $m-r$ blocks of size $(m-r) \times m$ on the diagonal all equal to the matrix U.

Since $(x, y, A) \in \varphi^{-1}(0)$, U and $Y(y)$ satisfy the matrix relation $UY(y) = S$ and S is full rank. So U and $Y(y)$ are full rank by the formula $\text{rank}\,(UY) \leq \min(\text{rank } U, \text{rank } Y)$. Moreover, the entries of $UY(y) - S$ do not depend on parameters $a_{0,i,j}$. Hence we can extract a square non-singular submatrix of the Jacobian matrix of order $m(m-r) + (m-r)^2 = (2m-r)(m-r)$, proving that it has row-rank $(2m-r)(m-r)$. *q.e.d.*

Proof of G_2 *in Assertion 2.* Suppose that $A \in \mathbb{M}^{n+1}_{m,m}(\mathbb{C})$ is such that that f satisfies Property G_2. Let

$$\begin{array}{rccc} \pi_1 : & \mathcal{V}_r & \to & \mathbb{C} \\ & (x, y) & \mapsto & x_1. \end{array}$$

Since $\mathcal{V}_r(A, u, s)$ is smooth and equidimensional of codimension $(2m-r)(m-r)$, by Sard's Lemma [139, Sec. 4.2] the critical values of π_1 are included in a hypersurface of $\mathbb{C}$. Let $\mathscr{T}_2 \subset \mathbb{C}$ be the complement of this hypersurface. Then, if $t \in \mathscr{T}_2$, one of the following facts hold:

- $\pi_1^{-1}(t) = \emptyset$. Hence $Z_{\mathbb{C}}(f_t) = \emptyset$ and by the Nullstellensatz $I(\pi_1^{-1}(t)) = \langle f_t \rangle = \langle 1 \rangle$, which is radical;
- $\pi_1^{-1}(t) \neq \emptyset$ and for all $(x, y) \in \pi_1^{-1}(t)$, (x, y) is not a critical point of the map π_1. So Df_t has full rank at each $(x, y) \in Z_{\mathbb{C}}(f_t)$, and by the Jacobian criterion (*cf.* [39, Th. 16.19] or Theorem 1.9) f_t defines a radical ideal and $Z_{\mathbb{C}}(f_t)$ is smooth and equidimensional of codimension $(2m-r)(m-r)+1$.

q.e.d.

Proof of G_3 *in Assertion 1.* This has been proved in Proposition 3.5. We denote by $\mathscr{A}_3$ the non-empty Zariski open set defined in Proposition 3.5. *q.e.d.*

Proof of G_3 *in Assertion 2.* Let $A \in \mathbb{M}^{n+1}_{m,m}(\mathbb{C})$ be such that $\mathcal{D}_p$ has codimension $(m-p)^2$ for some $0 \leq p \leq r$. Hence, by Bertini's Theorem [62, Th. 17.16] there exists a non-empty Zariski open set $\mathscr{T}^{(p)} \subset \mathbb{C}$ such that, if $t \in \mathscr{T} \cap \mathbb{Q}$, and $\tilde{A} = (A_0 + tA_1) + x_2A_2 + \cdots + x_nA_n$, hence $\tilde{\mathcal{D}}_p = \{x \in \mathbb{C}^{n-1} : \text{rank}\, \tilde{A}(x) \leq p\}$ has codimension $(m-p)^2$ or is empty. We conclude by defining $\mathscr{T}_3 = \bigcap_p \mathscr{T}^{(p)}$. *q.e.d.*

Conclusion of the proof. We end the proof by defining $\mathscr{A} = \mathscr{A}_1 \cap \mathscr{A}_2 \cap \mathscr{A}_3$ and $\mathscr{T} = \mathscr{T}_1 \cap \mathscr{T}_2 \cap \mathscr{T}_3$, which are non-empty and Zariski open. *q.e.d.*

3.5.2 Local analysis

Local description of the incidence variety

Let $A \in \mathbb{M}^{n+1}_{m,m}(\mathbb{Q})$ be a $n-$variate $m \times m$ linear matrix with coefficients in $\mathbb{Q}$, and let $r \leq m-1$. We show in Lemma 3.18 how to derive a local description of the incidence variety $\mathcal{V}_r(A, U, S)$. This actually relies on the strong structure of $\mathcal{D}_r$.

For $f \in \mathbb{Q}[x]$, denote by $\mathbb{Q}[x]_f$ the local ring of $\mathbb{Q}[x]$ at f. For $p \leq r$, let N be the upper-left $p \times p$ submatrix of A, so that A admits the following block partition:

$$A = \begin{pmatrix} N & Q \\ P' & R \end{pmatrix} \tag{3.2}$$

with $P, Q \in \mathbb{M}^{n+1}_{p,m-p}(\mathbb{Q})$ and $R \in \mathbb{M}^{n+1}_{m-p,m-p}(\mathbb{Q})$. The next Lemma 3.18 computes the equations of $\mathcal{V}_r$ in the local ring $\mathbb{Q}[x,y]_{\det N}$. We denote the Schur complement $R - P'N^{-1}Q$ of N in A by $\Sigma(N)$.

Lemma 3.18. *Let A, N, Q, P, R be as above, $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$ and $U \in \mathbb{M}_{m-r,r}(\mathbb{Q})$ be full-rank. Then there exist $\{q_{i,j}\}_{1 \leq i \leq p, 1 \leq j \leq m-r} \subset \mathbb{Q}[x]_{\det N}$ and $\{q'_{i,j}\}_{1 \leq i,j \leq m-p} \subset \mathbb{Q}[x]_{\det N}$ such that the locally closed set $\mathcal{V}_r \cap \{(x,y) : \det N(x) \neq 0\}$ is defined by the equations*

$$\begin{aligned} y_{i,j} - q_{i,1}y_{p+1,j} - \ldots - q_{i,m-p}y_{m,j} &= 0 \qquad i = 1,\ldots,p,\ j = 1,\ldots,m-r \\ q'_{i,1}y_{p+1,j} + \ldots + q'_{i,m-p}y_{m,j} &= 0 \qquad i = 1,\ldots,m-p,\ j = 1,\ldots,m-r \\ UY(y) - S &= 0. \end{aligned}$$

Proof. Denote by $Y^{(1)}$ and $Y^{(2)}$ the submatrices of $Y(y)$ containing respectively its first p rows and its last $m-p$ rows. Consider the block subdivision of A as in (3.2). We claim that, in $\mathbb{Q}[x,y]_{\det N}$, the $m(m-r)$ equations $A(x)Y(y) = 0$ can be read as the $m(m-r)$ equations:

$$\begin{pmatrix} \mathbb{I}_p Y^{(1)} + N^{-1}QY^{(2)} \\ \Sigma(N)Y^{(2)} \end{pmatrix} = 0$$

where $\Sigma(N) = R - P'N^{-1}Q$ is the Schur complement of N in A. Renaming the entries of $N^{-1}Q$ and $\Sigma(N)$ concludes the proof. To prove the claim, remark that since $\det N \neq 0$, $A(x)Y(y) = 0$ if and only if

$$\begin{pmatrix} \mathbb{I}_p & 0 \\ -P' & \mathbb{I}_{m-p} \end{pmatrix} \begin{pmatrix} N^{-1} & 0 \\ 0 & \mathbb{I}_{m-p} \end{pmatrix} \begin{pmatrix} N & Q \\ P' & R \end{pmatrix} Y(y) = 0.$$

q.e.d.

We conclude this section by observing that there is no loss of generality when choosing the block subdivision (3.2) to compute local equations, since this situation can be retrieved by applying elementary operations on the rows and columns of $A(x)$;

The rank at a critical point

In this section we study the expected rank of the linear matrix $A(x)$ when evaluated at a point $x \in \mathbb{C}^n$ that extends to a solution (x, y, z) of the Lagrange system $\ell(A \circ M, U, S, v)$. We conclude in Lemma 3.19 that up to genericity assumptions our algorithm always computes points of maximal rank r.

Let $N, P, Q, R, \Sigma(N)$ be as in the block subdivision 3.2. Let

$$\widetilde{A} = \begin{pmatrix} \mathbb{I}_p & N^{-1}Q \\ 0 & \Sigma(N) \end{pmatrix}.$$

Lemma 3.18 implies that the equations of $\mathcal{V}_r$ in the open set $\{(x, y) : \det N \neq 0\}$ read

$$\begin{aligned} \widetilde{A}Y(y) &= 0, \\ UY(y) - S &= 0. \end{aligned}$$

The entries of the above expressions are elements of the local ring $\mathbb{Q}[x]_{\det N}$. From Lemma 3.18 we deduce that one can eliminate $\{y_{i,j}\}_{1 \leq i \leq p, 1 \leq j \leq m-r}$ from the first group of equations $\widetilde{A}(x)Y(y) = 0$, since they can be expressed as polynomial functions of x and $\{y_{i,j}\}_{p+1 \leq i \leq m, 1 \leq j \leq m-r}$. That is, using the notations introduced in Lemma 3.18, we can express the entries of $Y^{(1)}$ as polynomials in x and in the entries of $Y^{(2)}$. Furthermore, since $UY(y) - S = 0$ consists in $(m-r)^2$ independent linear relations, one can eliminate $(m-r)^2$ among the $(m-p)(m-r)$ entries of $Y^{(2)}$. One can finally re-write $\Sigma(N)Y^{(2)} = 0$ as $(m-p)(m-r)$ relations in x and in the last $(r-p)(m-r)$ entries of $Y^{(2)}$.

We denote by F the resulting polynomial system. Let z be the vector $(z_1, \ldots, z_{(m-p)(m-r)})$ of Lagrange multipliers, and let

$$g = (g_1, \ldots, g_n) = z' D_x F - (w_1, \ldots, w_n).$$

The polynomial system (F, g) defines locally the critical points of the map $\pi_w \colon x \to w_1 x_1 + \ldots + w_n x_n$ restricted to $\mathcal{V}_r$. We deduce in Lemma 3.19 that, if $p \leq r-1$, and supposing that sets $\mathcal{D}_p$ have the expected dimension, the set of local critical points is generically empty.

Lemma 3.19. *Let $A \in \mathbb{M}^{n+1}_{m,m}(\mathbb{C}), 0 \leq r \leq m-1$ and suppose that (A, r) satisfies G_3. Let $p \leq r-1$ and let $g = (g_1, \ldots, g_n)$ be the polynomial system defined above. Then there exists a non-empty Zariski open set $\widetilde{\mathscr{W}} \subset \mathbb{C}^n$ such that if $w \in \widetilde{\mathscr{W}}$ then $g = 0$ has no solution.*

Proof. Let $C \subset \mathbb{C}^{2n+(r-p)(m-r)+(m-p)(m-r)}$ be the constructible set defined by $g = 0$ and by $\det N \neq 0$ and rank $A(x) = p$, and let $\overline{C}$ be its Zariski closure. Let $\pi_x \colon (x, y, z, w) \to x$ be the projection on the first n variables. The image $\pi_x(C)$ is included in $\mathcal{D}_p \subset \mathcal{D}_r$ and hence, since (A, r) satisfies G_3, it has dimension at most $n - (m-p)^2$. The fiber of π_x over a generic point $x \in \pi_x(C)$ is the graph of the polynomial function $w = z' D_x F$, and so it has codimension n and dimension

$(r-p)(m-r)+(m-p)(m-r) = (m-r)(m+r-2p)$. By the Theorem of the Dimension of Fibers [145, Sec. 6.3, Th. 7] one deduces that the dimension of C (and hence of $\overline{C}$) is at most $n-(m-p)^2+(m-r)(m+r-2p) = n-(r-p)^2$. Since $p \leq r-1$ then $\overline{C}$ has dimension at most $n-1$. We deduce that the projection of $\overline{C}$ onto the space $\mathbb{C}^n$ of w is a constructible set of dimension at most $n-1$, and consequently it is included in a proper hypersurface of $\mathbb{C}^n$. Defining $\widetilde{\mathscr{W}}$ as the complement of the hypersurface concludes the proof. *q.e.d.*

Local Lagrange systems

Let $\mathcal{V}_r = \mathcal{V}_r(A, U, S)$ be the incidence variety. Consider the restriction of the projection map $\pi_w \colon x \to w_1 x_1 + \ldots + w_n x_n$. We write below local equations for the Lagrange system encoding the critical points $\mathrm{crit}\,(\pi_w, \mathcal{V}_r)$ contained in some prescribed Zariski open subset of $\mathcal{V}_r$. Recall that our algorithm computes critical points with respect to the map $\pi_1 \colon x \to x_1$, which, in the aforementioned notation, equals π_{e_1} with $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^n$.

Suppose that $A \in \mathscr{A}$ (*cf.* Proposition 3.6), and let $c = (2m-r)(m-r)$. Hence $\mathcal{V}_r$ is smooth and equidimensional, of codimension c. Moreover, the set of critical points $\mathrm{crit}\,(\pi_w, \mathcal{V}_r)$ is the projection on the $(x, y)-$space of the solutions of

$$f(A, U, S), \qquad (g, h) = z' \left(\begin{array}{cc} D_x f & D_y f \\ w' & 0 \end{array} \right),$$

where $z = (z_1, \ldots, z_c, z_{c+1})$. Now, since $A \in \mathscr{A}$, at any solution (x, y, z) of this polynomial system, $z_{c+1} \neq 0$. Moreover, with the hypothesis $w \neq 0$, also $(z_1, \ldots, z_c) \neq 0$. Hence, to exclude possible trivial solutions coming from the homogeneity with respect to variables z, one introduces a linear relation $\sum_{i=1}^{c} v_i z_i - 1$ for some $v \in \mathbb{Q}^c$. This leads to the following system:

$$f = 0, \qquad g = 0, \qquad h = 0, \qquad \sum_{i=1}^{c} v_i z_i - 1 = 0. \tag{3.3}$$

Let N, P, Q, R as in the block subdivision of A in (3.2), with $p = r$. Lemma 3.18 implies that the local equations of $\mathcal{V}_r \cap \{(x, y) : \det N(x) \neq 0\}$ are

$$Y^{(1)} = -N^{-1} Q Y^{(2)}, \qquad \Sigma(N) Y^{(2)} = 0, \qquad U^{(1)} Y^{(1)} + U^{(2)} Y^{(2)} = S,$$

where $Y^{(1)}, Y^{(2)}$ is the row-subdivision of $Y(y)$ as in Lemma 3.18 and $U^{(1)}, U^{(2)}$ is the correspondent column-subdivision of U. We straightforwardly deduce the relation

$$S = U^{(1)}(-N^{-1} Q Y^{(2)}) + U^{(2)} Y^{(2)} = (-U^{(1)} N^{-1} Q + U^{(2)}) Y^{(2)}.$$

Since S is full-rank, then $Y^{(2)}$ and $-U^{(1)} N^{-1} Q + U^{(2)}$ are non-singular, and one deduces that:

- the second group of equations is $\Sigma(N) = 0$;

- the third group of equations is $Y^{(2)} = (-U^{(1)}N^{-1}Q + U^{(2)})^{-1}S$.

The entries of the Schur-complement $\Sigma(N)$ can be read in the local ring $\mathbb{Q}[x]_{\det N}$: by allowing division by $\det N$, they correspond exactly to the $(m-r)^2$ minors of $A(x)$ obtained as determinants of the $(r+1) \times (r+1)$ submatrices of $A(x)$ containing N (*cf.* the proof of [139, Prop. 3.2.7]). Since A satisfies G_1, the Jacobian $D_x(\Sigma(N)_{i,j})$ of the vector of entries of $\Sigma(N)$ has full-rank at each point x such that $\operatorname{rank} A(x) = r$.

We denote by $f' = (f'_1, \ldots, f'_c)$ the local equations represented by the entries of $Y^{(1)} + N^{-1}QY^{(2)}$, $\Sigma(N)$ and $Y^{(2)} - (-U^{(1)}N^{-1}Q + U^{(2)})^{-1}S$ in the local ring $\mathbb{Q}[x, y]_{\det N}$. The Jacobian matrix of f' has the following structure:

$$D(f') = [D_x(f') \;\; D_y(f')] = \left(\begin{array}{ccc} D_x[\Sigma(N)]_{i,j} & 0_{(m-r)^2 \times m(m-r)} & \\ & \mathbb{I}_{r(m-r)} & \star \\ \star & 0 & \mathbb{I}_{(m-r)^2} \end{array} \right).$$

Further, we consider the polynomials

$$(g'_1, \ldots, g'_n, h'_1, \ldots, h'_{m(m-r)}) = (z_1, \ldots, z_c, z_{c+1}) \left(\begin{array}{cc} D_x f' & D_y f' \\ w_1 \, \ldots \, w_n & 0 \end{array} \right).$$

Remark that by the particular structure of $D(f')$, polynomials in $h' = (h'_1, \ldots, h'_{m(m-r)})$ equal z_i, for $i = (m-r)^2 + 1, \ldots, c$, and hence variables $z_i, i = (m-r)^2 + 1, \ldots, c$ can be eliminated. So we finally deduce that the local equations of the Lagrange system (3.3) admit the following local equivalent formulation:

$$f' = 0, \qquad g' = 0, \qquad \sum_{i=1}^{(m-r)^2} v_i z_i - 1 = 0, \tag{3.4}$$

for a given $v \in \mathbb{Q}^{(m-r)^2}$. The projection of the solutions of this square system (with $n+c+1$ polynomials and $n+c+1$ variables) on the $(x, y)-$space contains the critical points of the restriction of π_w to the set $\mathcal{V}_r \cap \{(x, y) : \det N \neq 0\}$, where N is the upper-left $r \times r$ submatrix of $A(x)$. Analogous square systems can be obtained by localizing to a different $r \times r$ submatrix.

3.5.3 Finiteness of critical points

We first use the local analysis developed in the previous system to prove an intermediate result towards Proposition 3.7. The proof of Proposition 3.7 is given further.

An intermediate result

Denote by $\mathcal{W}_w = \mathcal{W}_w(A, U, S, v)$ the algebraic set defined by the polynomial system (3.3).

Lemma 3.20. *Let A satisfy* G*. There exist non-empty Zariski open sets $\mathscr{V} \subset \mathbb{C}^c$ and $\mathscr{W} \subset \mathbb{C}^n$ such that if $v \in \mathscr{V}$ and $w \in \mathscr{W}$, the following holds:*

1. *the set $\mathcal{W}_w(A, U, S, v)$ is finite and the Jacobian matrix of* (3.3) *has maximal rank at each point of $\mathcal{W}_w(A, U, S, v)$;*

2. *the projection of $\mathcal{W}_w(A, U, S, v)$ on the (x, y)-space contains the critical points of $\pi_w \colon (x, y) \to w'x$ restricted to $\mathcal{V}_r(A, U, S)$.*

Proof of Assertion 1. We first claim that we can reduce the polynomial equations defining $\mathcal{W}_w$ to the local formulation (3.4). Indeed, we can suppose that w lies in the non-empty Zariski open set $\widetilde{\mathscr{W}} \subset \mathbb{C}^n$ defined in Lemma 3.19. We deduce that, for such a choice, all solutions (x, y, z) to the system (3.3) are such that $\operatorname{rank} A(x) = r$. Consequently there exists a $r \times r$ submatrix N of $A(x)$ such that $\det N \neq 0$. We prove below that there exist non-empty Zariski open sets $\mathscr{V}'_N \subset \mathbb{C}^c$ and $\mathscr{W}_N \subset \mathbb{C}^n$ such that for $v \in \mathscr{V}'_N$ and $w \in \mathscr{W}_N$, the statement of Assertion (1) holds locally. Hence, to retrieve the global property, it is sufficient to define $\mathscr{V}'$ (resp. $\mathscr{W}$) as the finite intersection of sets $\mathscr{V}'_N$ (resp. $\mathscr{W}_N \cap \widetilde{\mathscr{W}}$), where N varies in the finite collection of $r \times r$ submatrices of A. Hence we prove below that sets $\mathscr{V}'_N$ and $\mathscr{W}_N$ exist.

Without loss of generality we choose N to be the upper-left $r \times r$ submatrix of $A(x)$. We define the map

$$\begin{array}{rccc} p: & \mathbb{C}^{n+c+1} \times \widetilde{\mathscr{W}} \times \mathbb{C}^{(m-r)^2} & \longrightarrow & \mathbb{C}^{n+c+1} \\ & (x, y, z, w, v) & \longmapsto & (f', g', \sum_{i=1}^{(m-r)^2} v_i z_i - 1) \end{array}$$

and its section map

$$\begin{array}{rccc} p_{v,w}: & \mathbb{C}^{n+c+1} & \longrightarrow & \mathbb{C}^{n+c+1} \\ & (x, y, z) & \longmapsto & (f', g', \sum_{i=1}^{(m-r)^2} v_i z_i - 1), \end{array}$$

for fixed $v \in \mathbb{C}^{(m-r)^2}$ and $w \in \widetilde{\mathscr{W}}$, where f' and g' have been defined in (3.4), page 103. As in the proof of Proposition 3.6, we consider a first trivial case. If $p^{-1}(0) = \emptyset$, then for all v, w, $p_{v,w}^{-1}(0) = \emptyset$, and the claim is proved by taking $\mathscr{V}' = \mathscr{V}'_N = \mathbb{C}^c$ and $\mathscr{W} = \mathscr{W}_N = \widetilde{\mathscr{W}}$ (see Lemma 3.19).

As in the proof of Proposition 3.6, for $(x, y, z, w, v) \in p^{-1}(0)$, we exhibit a non-singular maximal minor of Dp at (x, y, z, w, v), which proves that 0 is a regular value of p and that, by successively applying Thom's Weak Transversality Theorem and the Jacobian criterion, there exist $\mathscr{V}''_N \subset \mathbb{C}^{(m-r)^2}$ and $\mathscr{W}_N \subset \mathbb{C}^n$ such that if $v \in \mathscr{V}''_N \cap \mathbb{Q}^{(m-r)^2}$ and $w \in \mathscr{W}_N \cap \mathbb{Q}^n$, the set $\mathcal{W}_w(A, U, S, v) \cap \{(x, y, z) : \det N(x) \neq 0\}$ is empty or zero-dimensional. We deduce that the claim is true with $\mathscr{V}'_N = \mathscr{V}''_N \times \mathbb{C}^{c-(m-r)^2}$, which is also non-empty and Zariski open. We prove below this claim by exhibiting a non-singular submatrix of Dp.

Let us recall that A satisfies G. Hence the Jacobian matrix Df' has maximal rank at (x, y). Moreover, $z_{c+1} \neq 0$ and by $\sum v_i z_i - 1 = 0$ there exists $1 \leq \ell \leq$

$(m-r)^2$ such that $z_\ell \neq 0$. Then we consider the submatrix of Dp obtained by isolating:

- the non-singular submatrix of Df';
- the derivatives of $g_1, \ldots, g_n$ with respect to $w_1, \ldots, w_n$, giving the identity block $\mathbb{I}_n$;
- the derivative of $\sum v_i z_i - 1$ with respect to v_ℓ, which equals $z_\ell \neq 0$.

The previous blocks generate a submatrix of size $(n+c+1) \times (n+c+1)$ whose determinant does not vanish at (x, y, z, w, v). *q.e.d.*

Proof of Assertion 2. Suppose first that $\mathcal{W}_w(A, U, S, v) = \emptyset$ for all $w \in \widetilde{\mathscr{W}}$ and $v \in \mathbb{C}^c$. Let $w \in \widetilde{\mathscr{W}}$, and $(x, y) \in \text{crit}\,(\pi_w, \mathcal{V}_r)$. Since $\mathcal{V}_r$ is equidimensional and smooth, there exists $z \neq 0$ such that (x, y, z) verifies $z' Df = (w, 0)$. Since $z \leq 0$ there exists $v \in \mathbb{C}^c$ such that $v'z = 1$. Hence, we conclude $(x, y, (z, 1)) \in \mathcal{W}_w(A, U, S, v) \neq \emptyset$, which is a contradiction. Hence $\text{crit}\,(\pi_w, \mathcal{V}_r) = \emptyset$.

Let A satisfy G. Suppose that $Z_{\mathbb{C}}(p)$ is not empty, and that $w \in \mathscr{W}$ (defined in Assertion 1). By [139, Sec. 3.2], $\text{crit}\,(\pi_w, \mathcal{V}_r)$ is the image of the projection $\pi_{x,y}$ on x, y of the constructible set:

$$\mathcal{S} = \{(x, y, z) : f = g = h = 0,\ z \neq 0\}$$

where f, g, h have been defined in (3.3). As in the proof of Assumption 1, one can apply Thom's Weak Transversality Theorem to deduce that $\mathcal{S}$ has dimension at most 1. Moreover, by homogeneity with respect to variables z, for each $(x, y) \in \pi_{x,y}(\mathcal{S})$, the fiber $\pi_{x,y}^{-1}(x, y)$ has dimension 1 (indeed, if $(x, y, z) \in \pi_{x,y}^{-1}(x, y)$, then $(x, y, \lambda z) \in \pi_{x,y}^{-1}(x, y)$ for all $\lambda \neq 0$). By the Theorem on the Dimension of Fibers [145, Sec. 6.3, Th. 7], we deduce that $\pi_{x,y}(\mathcal{S})$ is finite. Moreover, for $(x, y) \in \pi_{x,y}(\mathcal{S})$, let $\mathscr{V}_{(x,y)} \subset \mathbb{C}^c$ be the non-empty Zariski open set such that if $v \in \mathscr{V}_{(x,y)}$, the hyperplane $\sum v_i z_i - 1 = 0$ intersects transversely $\pi_{x,y}^{-1}(x, y)$. Let $\mathscr{V}' \subset \mathbb{C}^c$ be the set defined in the proof of Assertion 1. By defining

$$\mathscr{V} = \mathscr{V}' \bigcap_{(x,y) \in \pi_{x,y}(\mathcal{S})} \mathscr{V}_{(x,y)}$$

one concludes the proof. Indeed, $\mathscr{V}$ is a finite intersection of non-empty Zariski open sets. *q.e.d.*

Proof of Proposition 3.7

We are finally able to deduce the proof of Proposition 3.7. The strategy is to translate the choice of a generic parameter $w \in \mathbb{C}^n$ (and hence, a generic linear function $\pi_w \colon x \to w'x$) to the choice of a change of variables $x \mapsto M\,x$ for some $M \in \mathrm{GL}_n(\mathbb{C})$.

Proof of Proposition 3.7. Let $\mathscr{V}, \mathscr{W}$ be the sets defined in Lemma 3.20. We denote by

$$\mathscr{M}_1 = \left\{ M \in \mathrm{GL}(n, \mathbb{C}) : \text{the first row of } M^{-1} \text{ lies in } \mathscr{W} \right\}.$$

$\mathscr{M}_1$ is non-empty and Zariski open since the entries of M^{-1} are rational functions of the entries of M. Let $v \in \mathscr{V}$. Denoted by $e'_1 = (1, 0, \ldots, 0) \in \mathbb{Q}^n$, remark that for any $M \in \mathscr{M}_1$ the following identity holds:

$$\begin{pmatrix} Df(A \circ M, U, S) \\ e'_1 \quad 0 \cdots 0 \end{pmatrix} = \begin{pmatrix} Df(A, U, S) \circ M \\ w' \quad 0 \cdots 0 \end{pmatrix} \begin{pmatrix} M & 0 \\ 0 & \mathbb{I}_{m(m-r)} \end{pmatrix}$$

where $Df(A, U, S) \circ M$ means that in all entries of $Df(A, U, S)$ we substitute $x \mapsto M\,x$.

We conclude that the set of solutions of the system

$$\begin{aligned} f(A, U, S) &= 0 \\ (z_1, \ldots, z_c) Df(A, U, S) + z_{c+1}(w', 0) &= 0 \\ v_1 z_1 + \cdots + v_c z_c - 1 &= 0 \end{aligned} \tag{3.5}$$

is the image by the map

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} M^{-1} & 0 & 0 \\ 0 & \mathbb{I}_{m(m-r)} & 0 \\ 0 & 0 & \mathbb{I}_{c+1} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

of the set $\mathcal{S}$ of solutions of the system

$$\begin{aligned} f(A \circ M, U, S) &= 0 \\ (z_1, \ldots, z_c) Df(A \circ M, U, S) + z_{c+1}(e'_1, 0) &= 0 \\ v_1 z_1 + \cdots + v_c z_c - 1 &= 0. \end{aligned} \tag{3.6}$$

Let π be the projection that forgets the last coordinate z_{c+1}. We deduce the equality $\pi(\mathcal{S}) = \mathscr{Z}(A \circ M, U, S, v)$, and that π is a bijection. Moreover, it is an isomorphism of affine algebraic varieties, since if $(x, y, z) \in \mathcal{S}$, then its z_{c+1}-coordinate is obtained by evaluating a polynomial at $(x, y, z_1 \ldots z_c)$. Thus, we deduce by Assertion 1 of Lemma 3.20 that:

- $\mathcal{S}$ and $\pi(\mathcal{S}) = \mathscr{Z}(A \circ M, U, S, v)$ are finite;
- the Jacobian matrix associated to (3.6) has maximal rank at any point of $\mathcal{S}$.

Since $\pi(\mathcal{S}) = \mathscr{Z}(A \circ M, U, S, v)$ and that π is an isomorphism, Assertion 1 of Proposition 3.7 follows. Assertion (2) is a straightforward consequence of Assertion 2 of Lemma 3.20. *q.e.d.*

3.5.4 Closure of projections

The final step is to prove Proposition 3.8 yielding a regularity result about projections and connected components of real algebraic sets $\mathcal{D}_r \cap \mathbb{R}^n$. The key strategy is to exploit general results proved in Chapter 2, Section 2.2. Indeed, for $\mathcal{Z} = \mathcal{D}_r$ and $M \in \mathrm{GL}_n(\mathbb{C})$, we recall that in Section 2.1.2 we built the recursive decomposition

$$\{\mathcal{O}_i(M^{-1}\mathcal{D}_r) : i = 0, \ldots, d\}$$

(with $d = \dim \mathcal{D}_r$) of $M^{-1}\mathcal{D}_r$ into algebraic sets.

Proof of Assertion 1 of Proposition 3.8. Let $\mathscr{M}_2 \subset \mathrm{GL}_n(\mathbb{C})$ be the non-empty Zariski open set defined in Proposition 2.4, page 67, for $\mathcal{Z} = \mathcal{D}_r$. We deduce that any $M \in \mathscr{M}_2$ verifies $\mathsf{P}(\mathcal{D}_r)$ (*cf.* Section 2.2.1 on page 65). Observe that for $M \in \mathrm{GL}_n(\mathbb{C})$, there is a natural bijective correspondence between the collection of connected components of $\mathcal{D}_r \cap \mathbb{R}^n$ and the one of $M^{-1}\mathcal{D}_r \cap \mathbb{R}^n$, which is given by $\mathcal{C} \leftrightarrow M^{-1}\mathcal{C}$.

Let us fix $M^{-1}\mathcal{C} \subset M^{-1}\mathcal{D}_r \cap \mathbb{R}^n$ connected component, with $M \in \mathscr{M}_2$. Consider the projection π_i restricted to $M^{-1}\mathcal{D}_r \cap \mathbb{R}^n$. Since $M \in \mathscr{M}_2$, by Proposition 2.2, the boundary of $\pi_i(M^{-1}\mathcal{C})$ is a subset of $\pi_i(\mathcal{O}_{i-1}(M^{-1}\mathcal{D}_r) \cap M^{-1}\mathcal{C})$ and hence it is contained in $\pi_i(M^{-1}\mathcal{C})$. This implies that $\pi_i(M^{-1}\mathcal{C})$ is closed and concludes Assertion 1. *q.e.d.*

Proof of Assertion 2 of Proposition 3.8. Let $M \in \mathscr{M}_2$ and let $t \in \mathbb{R}$ be in the boundary of $\pi_1(M^{-1}\mathcal{C})$. By Lemma 2.5, $\pi_1^{-1}(t) \cap M^{-1}\mathcal{C}$ is finite. For a fixed $x \in \pi_1^{-1}(t) \cap M^{-1}\mathcal{C}$, let $p \leq r$ be the rank of $A(x)$. The polynomial system $y \mapsto f(A, U, S)$ parametrized by U, S is hence linear in y and can be rewritten

$$\left(\begin{array}{c} A(x) \\ U \end{array}\right) Y(y) = \left(\begin{array}{c} 0_{m\times(m-r)} \\ S \end{array}\right). \tag{3.7}$$

System (3.7) is equivalent to $m - r$ linear systems of equations, whose unknowns are the entries of the columns of $Y(y)$ and whose constant terms are the columns of the right side matrix of (3.7). By linear algebra (apply the Rouché-Capelli Theorem [98] to each linear system) the equations (3.7) have a solution if and only if each column of the right side matrix lies in the space spanned by the columns of $A(x)$. This is possible if and only if

$$\mathrm{rank}\left(\begin{array}{c} A(x) \\ U \end{array}\right) = \mathrm{rank}\left(\begin{array}{cc} A(x) & 0_{m\times(m-r)} \\ U & S \end{array}\right).$$

Denote by:

- $r_{x,U}$, the number of rows of U that do not lie in the space spanned by the rows of $A(x)$, and by

- $r_{x,U,S}$, the number of rows of $(U \mid S)$ that do not lie in the space spanned by the rows of $(A(x) \mid 0)$.

Since S has full rank, we deduce that, for all U, $r_{x,U} \leq r_{x,U,S} = m - r$ (no line of S is the zero vector). Moreover, since x is fixed, there exists a non-empty Zariski open set $\mathscr{U}_{\mathcal{C},x}$ such that if $U \in \mathscr{U}_{\mathcal{C},x}$ then $r_{x,U} = m - r$ and

$$\operatorname{rank} \begin{pmatrix} A(x) \\ U \end{pmatrix} = p + r_{x,U} = p + m - r$$

and

$$\operatorname{rank} \begin{pmatrix} A(x) & 0_{m \times (m-r)} \\ U & S \end{pmatrix} = p + r_{x,U,S} = p + m - r.$$

This implies that, for $U \in \mathscr{U}_{\mathcal{C},x}$, the system $y \mapsto f(A, U, S)$ has at least one solution. One concludes by defining

$$\mathscr{U} = \bigcap_{\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n} \; \bigcap_{x \in \pi_1^{-1}(t) \cap \mathcal{C}} \mathscr{U}_{\mathcal{C},x}.$$

The set $\mathscr{U}$ is non-empty and Zariski open by the finiteness of the number of connected components of $\mathcal{D}_r \cap \mathbb{R}^n$ and of the set $\pi_1^{-1}(t) \cap \mathcal{C}$. *q.e.d.*

3.6 First experiments

We present in this section a first bunch of experimental data for the algorithm LowRank. We refer to Section 5.2.1 for a more deep discussion on the practical performances and to Section 5.1 for a description of the implementation.

(m, r, n)	PPC	LowRank	deg	MBB on deg	maxdeg	MBB on maxdeg
$(3, 2, 8)$	109	18	39	49	12	16
$(3, 2, 9)$	230	20	39	49	12	16
$(4, 2, 5)$	12.2	26	100	574	80	504
$(4, 2, 6)$	∞	593	276	2394	176	1820
$(4, 2, 7)$	∞	6684	532	6530	256	4136
$(4, 2, 8)$	∞	42868	818	12851	286	6321
$(4, 2, 9)$	∞	120801	1074	19403	286	6552
$(4, 3, 10)$	∞	303	284	367	84	108
$(4, 3, 11)$	∞	377	284	367	84	108
$(5, 2, 7)$	4.2	0.7	0	0	0	0
$(5, 2, 8)$	8	0.7	0	0	0	0
$(5, 2, 9)$	∞	903	175	5005	175	5005
$(6, 3, 8)$	∞	2	0	0	0	0
$(6, 5, 4)$	∞	8643	726	1106	540	825

Table 3.1 – Timings and degrees for dense linear matrices (*cf.* Table 5.1, page 148)

We have implemented our algorithm under MAPLE and compared it with the functions of the real algebraic geometry library RAGLIB [135] implemented by Safey El Din. Table 3.1 contains some data. In column LowRank we report on timings (expressed in seconds) of our algorithm with input randomly generated linear matrices, and allowing the expected rank to vary; in column PPC we give the timings of the execution of the function PointsPerComponents of RAGLIB: this function computes at least one point in each connected component of the input algebraic set, implementing the best known state-of-the-art exact algorithms [137]. The input of PointsPerComponents is the list of minors of size $r+1$ of A.

We recall that LowRank is a recursive algorithm. At the end of each recursive subroutine LowRankRec, a rational parametrization is computed (*cf.* step 4 in the formal description of this subroutine on page 83). After each recursive step the first variable x_1 is eliminated by considering the intersection of $\mathcal{D}_r \cap \mathbb{R}^n$ with a generic fiber of the projection $\pi_1 \colon x \to x_1$.

- The symbol ∞ means that the computation did not end after 48 hours.
- The column deg contains the degree of the output representation parametrizing the union of all finite sets computed during the recursive subroutines. That is, the output of LowRank is a rational parametrization of degree deg, whose solutions meet every connected component of $\mathcal{D}_r \cap \mathbb{R}^n$. We put a 0 in the table when the output of LowRank is the empty list.
- The column maxdeg contains the maximum of the degrees of the parametrizations computed in the recursive routines LowRankRec at step 4 (*cf.* page 83). As for deg, when the empty list is returned, we put a 0 in the table.
- The columns "MBB on deg" and "MBB on maxdeg" contain the multilinear Bézout bounds (on deg and maxdeg, respectively) computed in Corollary 3.11 and Proposition 3.10, respectively.

As clarified by Table 5.1, the values of deg, maxdeg and of their multilinear bounds, stabilize when m, r are fixed and n grows. Moreover, the bounds are sharp when no solutions are computed (that is, when the real algebraic set $\mathcal{D}_r \cap \mathbb{R}^n$ is empty).

Chapter 4

Exact algorithms for linear matrix inequalities

Linear matrix inequalities (LMI) define an important class of algebraic constraints that appear in many contexts, mainly in various formulations of polynomial optimization problems, and in control theory, where such inequalities arise, *e.g.*, in Lyapunov stability theory for linear systems. The main interest concerning LMI is deciding whether the associated feasible set is empty or not, and, in the positive case, to compute one of its elements. Moreover, it is of tremendous importance to design algorithms proving the emptiness when it occurs. In this chapter we design an exact algorithm solving this decision problem. The key result is a theorem for spectrahedra, allowing to reduce the emptiness problem to a suitable version of the real root finding problem for low rank loci of symmetric linear matrices.

The results of this chapter are part or generalizations of [69, 71].

4.1 Introduction

4.1.1 Problem statement

Let $A_0, A_1, \ldots, A_n \in \mathbb{S}_m(\mathbb{Q})$ be symmetric matrices of size m with entries in the field $\mathbb{Q}$. Consistently with the notation introduced for linear matrices in Chapter 3, we denote by $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$ the pencil generated by matrices A_i. For any $x \in \mathbb{R}^n$, the evaluated matrix $A(x)$ is also symmetric.

The central object of this chapter is the semi-algebraic set containing the points $x \in \mathbb{R}^n$ where the eigenvalues of the evaluated matrix $A(x)$ are all non-negative, that is the associated spectrahedron

$$\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}.$$

This chapter investigates the following problem.

Problem 4 (Emptiness of spectrahedra)**.** *Design an efficient and exact algorithm whose input is* $A = (A_0, A_1, \ldots, A_n) \in \mathbb{S}_m^{n+1}(\mathbb{Q})$*, and whose output is either*

(A) the empty list, if and only if $\mathscr{S} = \emptyset$*, or*

(B) a rational parametrization encoding a finite subset of $\mathbb{R}^n$ *meeting the set* $\mathscr{S}$ *in at least one point* x^**.*

In order to make the goal which is hidden into Problem 4 explicit, we remark that our approach for solving Problem 4 will lead to answer the following questions, for a given symmetric pencil A:

- Is the associated spectrahedron $\mathscr{S}$ empty?
- If $\mathscr{S} \neq \emptyset$, how can we compute a point $x^* \in \mathscr{S}$?
- If $\mathscr{S} \neq \emptyset$, what is the minimum rank attained by $A(x)$ on $\mathscr{S}$?

4.1.2 Main result and chapter outline

Strategy and main result

The strategy to attack Problem 4 relies on a result (discussed and proved in Section 4.2) which relates the emptiness problem for spectrahedra and that of symmetric low rank loci. It can be summarized as follows.

Step 1. Suppose that $\mathscr{S}$ is not empty, and that $r(A)$ is the minimum rank attained by $A(x)$ on $\mathscr{S}$. We prove further in Theorem 4.4 that $\mathscr{S}$ contains at least one of the connected components of $\mathcal{D}_r \cap \mathbb{R}^n$ where $r = r(A)$.

Step 2. Hence computing one point on $\mathscr{S}$ can be performed by computing one point per connected component on the set $\mathcal{D}_r \cap \mathbb{R}^n$, where $r = r(A)$. This can be done by designing a variant of algorithm LowRank described in Chapter 3, dedicated to symmetric pencils.

The main result of this chapter is stated by the following two theorems. The proof of the first one is obtained by merging the proof of correctness of the algorithm given in Theorem 4.14, page 127, and the estimate of its complexity, which is given by Theorem 4.19, page 137.

Similarly to the algorithm LowRank of Chapter 3, this new algorithm is exact and probabilistic. Given as input a symmetric pencil $A(x)$, it builds a flag of incidence varieties $\mathcal{V}_p, p = 0, \ldots, r(A)$, as in Chapter 3. The algorithm assume as hypotheses that the incidence varieties are regular (smooth and equidimensional) and that the defining polynomials generate a radical ideal.

Theorem 4.1 (Main Theorem 4)**.** *Suppose that for* $0 \leq r \leq m-1$*, the incidence variety* $\mathcal{V}_r$ *is smooth and equidimensional and that its defining polynomial system generates a radical ideal. There is a probabilistic algorithm that takes* A *as input and returns:*

1. either the empty list, if and only if $\mathscr{S} = \emptyset$, or

2. the coordinates of a vector x^* such that $A(x^*) = 0$, if and only if the linear system $A(x) = 0$ has a solution, or

3. a rational parametrization $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \in \mathbb{Q}[t]^{n+2}$ such that there exists $t^* \in \mathbb{R}$ with $q_{n+1}(t^*) = 0$ and:

- $A(q_1(t^*)/q_0(t^*), \ldots, q_n(t^*)/q_0(t^*)) \succeq 0$ *and*
- $\operatorname{rank} A(q_1(t^*)/q_0(t^*), \ldots, q_n(t^*)/q_0(t^*)) = r(A)$.

The number of arithmetic operations performed are in

$$\tilde{\mathcal{O}}\left(n \sum_{r \leq m-1} \binom{m}{r} (n + p_r + r(m-r))^7 \binom{p_r + n}{n}^6\right) \quad \text{if } \mathscr{S} \text{ is empty}$$

$$\tilde{\mathcal{O}}\left(n \sum_{r \leq r(A)} \binom{m}{r} (n + p_r + r(m-r))^7 \binom{p_r + n}{n}^6\right) \quad \text{if } \mathscr{S} \text{ is not empty,}$$

with $p_r = (m-r)(m+r+1)/2$. If $\mathscr{S}$ is not empty, the degree of q is in

$$\mathcal{O}\left(\binom{m}{r(A)} \binom{p_{r(A)} + n}{n}^3\right).$$

In Section 4.6 we present a dedicated variant for linear Hankel matrices. This variant leads to the following special complexity estimates.

Theorem 4.2 (Main Theorem 5)**.** *Let A be a $n-$variate linear Hankel matrix of size m, and let $r \leq m-1$. Suppose that for $0 \leq p \leq r$, the incidence variety $\mathcal{V}_p$ is smooth and equidimensional and that its defining polynomial system generates a radical ideal. Suppose that, for $0 \leq p \leq r$, $\mathcal{D}_p$ is empty or has the expected dimension $n - 2m + 2p + 1$.*

There exists a probabilistic algorithm such that, with input (A, r), returns in output a rational parametrization $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \subset \mathbb{Q}[t]$ of degree at most

$$\binom{2m - r - 1}{r} + \sum_{k=2m-2r}^{n} \sum_{p=0}^{r} d(m, k, p)$$

with $d(m, k, p) \in \mathcal{O}(\binom{k+2m-p-1}{k}^3)$, within

$$\mathcal{O}\left(\sum_{p=0}^{r} pn(2m-p)(pn(2m-p)(n+2m)^2 + (n+2m)^4) d(m,n,p)^2\right)$$

arithmetic operations over $\mathbb{Q}$. The set defined by q contains at least one point in each connected component of the real algebraic set $\mathcal{D}_r \cap \mathbb{R}^n$.

We remark here that both Theorem 4.1 and 4.2 lead to complexities respectively for the emptiness problem for spectrahedra and for the real root finding problem for Hankel linear matrices, that are essentially quadratic on the degree of the output rational parametrization. Moreover, for the special subclasses of these problems when the size m of the matrix is fixed, the complexity is polynomial in n, the number of variables.

Outline

Section 4.2 contains a geometric fact about spectrahedra on which the whole procedure strongly relies. The main algorithm together with its subroutines is described in Section 4.4, while its complexity is estimated in Section 4.5. Finally, in Section 4.6 we describe a dedicated variant of such an algorithm for Hankel matrices. Numerical tests and examples of the applications of the algorithms of Theorems 4.1 and 4.2 are in Chapter 5, Section 5.3.

4.2 The smallest rank on a spectrahedron

We first present a geometric result about spectrahedra: it relates the geometry of linear matrix inequalities to the rank stratification of the defining symmetric pencil $A(x)$.

Notation 4.3. *If $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$ is not empty, we denote the minimum rank attained by $A(x)$ on $\mathscr{S}$ by $r(A)$.*

We recall that $\mathbb{S}_m(\mathbb{Q})$ denotes the vector space of symmetric $m \times m$ matrices with entries in $\mathbb{Q}$, and that a symmetric linear matrix is an element of the $(n+1)-$fold Cartesian product $\mathbb{S}_m^{n+1}(\mathbb{Q})$ of $\mathbb{S}_m(\mathbb{Q})$. We believe that the following fact is known to the community of researchers working on real algebraic geometry and semidefinite optimization; however, we did not find any explicit reference in the literature, and hence we prefer to give our original contribution.

Theorem 4.4 (Main Theorem 3)**.** *Suppose that $\mathscr{S} \neq \emptyset$, and let $r = r(A)$. Let $\mathcal{C}$ be a connected component of $\mathcal{D}_r \cap \mathbb{R}^n$ such that $\mathcal{C} \cap \mathscr{S} \neq \emptyset$. Then $\mathcal{C} \subset \mathscr{S}$ (and, hence, $\mathcal{C} \subset (\mathcal{D}_r \setminus \mathcal{D}_{r-1}) \cap \mathbb{R}^n$).*

Proof. By assumption, at all points of $\mathscr{S}$, the matrix A has rank at least $r = r(A)$ and there exists a point in $\mathscr{S}$ where the rank of A is exactly r. We consider the vector function

$$e = (e_1, \ldots, e_m) : \quad \mathbb{R}^n \quad \longrightarrow \quad \mathbb{R}^m$$

where $e_1(x) \leq \ldots \leq e_m(x)$ are the ordered eigenvalues of $A(x)$.

Let $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$ be the given connected component such that $\mathcal{C} \cap \mathscr{S} \neq \emptyset$, and let $x \in \mathcal{C} \cap \mathscr{S}$. One has rank $A(x) = r$ and

$$e_1(x) = \ldots = e_{m-r}(x) = 0 < e_{m-r+1}(x) \leq \ldots \leq e_m(x).$$

Suppose that there exists $y \in \mathcal{C}$ such that $y \notin \mathscr{S}$; that is $\operatorname{rank} A(y) \leq r$ but $A(y)$ is not positive semi-definite. In particular, one eigenvalue of $A(y)$ is strictly negative.

Let $g\colon [0,1] \to \mathcal{C}$ be a continuous semi-algebraic map such that $g(0) = x$ and $g(1) = y$. This map exists since $\mathcal{C}$ is a connected component of an algebraic set. The image $g([0,1])$ is compact and semi-algebraic. Let

$$T = \{t \in [0,1] : g(t) \in \mathscr{S}\} = g^{-1}(g([0,1]) \cap \mathscr{S}).$$

Since g is continuous, $T \subset [0,1]$ is closed. So it is a finite union of closed intervals. Since $0 \in T$ (in fact, $g(0) = x \in \mathscr{S}$) there exists $t_0 \in [0,1]$ and $N \in \mathbb{N}$ such that $[0,t_0] \in T$ and for all $p \geq N$, $t_0 + \frac{1}{p} \notin T$.

One gets that $g(t_0) = x' \in \mathscr{S}$ and that for all $p \geq N$, $g(t_0 + \frac{1}{p}) = x'_p \notin \mathscr{S}$. By definition, $x', x'_p \in \mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$ for all $p \geq N$, and since $x' \in \mathscr{S}$, we get $\operatorname{rank} A(x') = r$ and $\operatorname{rank} A(x'_p) \leq r$ for all $p \geq N$. We also get that $\operatorname{rank} A(g(t)) = r$ for all $t \in [0, t_0]$. We finally have $x'_p \to x'$ when $p \to +\infty$, since g is continuous. There exists a map

$$\varphi\colon \Big\{p \in \mathbb{N} : p \geq N\Big\} \to \{0, 1, \ldots, r\}$$

which assigns to p the index of eigenvalue-function among $e_1, \ldots, e_m$ corresponding to the maximum strictly negative eigenvalue of $A(x'_p)$, if it exists; otherwise it assigns 0. Remark that since $\operatorname{rank} A(x'_p) \leq r$ for all p, then $0 \leq \varphi(p) \leq r$ for all p, and the map is well defined. In other words, the eigenvalues of $A(x'_p)$ are

$$\begin{aligned}
e_1(x'_p) \leq \ldots \leq e_{\varphi(p)}(x'_p) < 0 \\
0 = e_{\varphi(p)+1}(x'_p) = \ldots = \lambda_{\varphi(p)+m-r}(x'_p) \\
0 \leq e_{\varphi(p)+m-r+1}(x'_p) \leq \ldots \leq e_m(x'_p),
\end{aligned}$$

for $p \geq N$. Since the sequence $\{\varphi(p)\}_{p \geq N}$ is bounded, up to taking a subsequence, it admits at least a limit point by Bolzano-Weierstrass Theorem [2, Th. 3.4.8]. Since it is an integer-value sequence, this limit point is an integer number. Moreover, if $0 \leq \ell \leq r$ is a limit point, and $\{p_j\}_{j \in \mathbb{N}}$ a subsequence such that $\varphi(p_j) \to \ell$, then we claim that there exists an integer N' such that $\varphi(p_j) = \varphi(p_{j+1}) = \ell$ for all $j \geq N'$ (which means that $j \mapsto \varphi(p_j)$ is constant for $j \geq N'$): this holds since the map φ takes only integer values.

Suppose that there exists a limit point $\ell > 0$ (strictly positive), and let $\{p_j\}_{j \in \mathbb{N}}$ and N' be as above. One obtains that $\varphi(p_j) \to \ell$ and that this sequence is constant for $j \geq N'$. Hence, the zero eigenvalues of $A(x'_{p_j})$ are

$$0 = \lambda_{\ell+1}(x'_{p_j}) = \ldots = \lambda_{\ell+m-r}(x'_{p_j}),$$

for all $j \geq N'$. Since $x'_{p_j} \to x'$ and $e_1, \ldots, e_m$ are continuous functions, we obtain that

$$\begin{aligned} e_1(x') \leq \ldots \leq e_\ell(x') \leq 0, \\ 0 = e_{\ell+1}(x') = \ldots = e_{\ell+m-r}(x'), \\ 0 \leq e_{\ell+m-r+1}(x') \leq \ldots \leq e_m(x'). \end{aligned}$$

Since $A(x') \succeq 0$, one gets $0 = e_1(x') = \ldots = e_{\ell+m-r}(x')$, that is $A(x')$ has at least $\ell + m - r > m - r$ zero eigenvalues. This implies that $\operatorname{rank} A(x') \leq r - 1$, which is a contradiction, since we assumed $x' \in \mathscr{S}$ and that r is the minimum rank attained by A on $\mathscr{S}$.

We deduce that 0 is the unique limit point of φ, hence φ converges to 0. We already showed that in this case $\varphi(p) = 0$ for $p \geq N''$, for some $N'' \in \mathbb{N}$. This means in particular that for $p \geq N''$, the number of strictly negative eigenvalues of $A(x'_p) = A(g(t_0 + \frac{1}{p}))$ is zero, that is the matrix A is positive semidefinite at any point in

$$\{x'_p : p \geq N''\}.$$

So this set is included in $\mathscr{S}$, which contradicts our assumptions. We conclude that the set $\mathcal{C} \setminus \mathscr{S}$ is empty, that is $\mathcal{C} \subset \mathscr{S}$. By the minimality of the rank r in $\{\operatorname{rank} A(x) : x \in \mathscr{S}\}$, one deduces that $\mathcal{C} \subset (\mathcal{D}_r \setminus \mathcal{D}_{r-1}) \cap \mathbb{R}^n$. *q.e.d.*

The previous theorem allows to deduce the following dichotomy on the input symmetric linear pencil $A(x)$:

- Either $\mathscr{S} = \emptyset$, or
- $\mathscr{S}$ contains one connected component of $\mathcal{D}_{r(A)}$.

Consequently, the problem of deciding the emptiness and computing sample finite subsets of spectrahedra is reduced to a dedicated variant of the real root finding problem (already addressed for generic matrices in Chapter 3) for symmetric pencils. In particular, by Theorem 4.4, in order to compute one point on $\mathscr{S}$ one only has to solve the following problem: for $r = r(A)$, compute a rational parametrization whose solution set meets every connected component $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$ such that $\mathcal{C} \cap \mathcal{D}_{r-1} = \emptyset$.

4.3 Examples of spectrahedra

We present some examples of spectrahedra with explicit pictures. In Section 5.3 we will apply the algorithm of this chapter on these examples.

Example 4.5. *We consider the* 3×3 *linear matrix*

$$A(x) = \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_3 \\ x_2 & x_3 & 1 \end{pmatrix}.$$

The convex region $\{x \in \mathbb{R}^3 : A(x) \succeq 0\}$ is the Cayley spectrahedron, pictured in Figure 4.1 with its algebraic boundary.

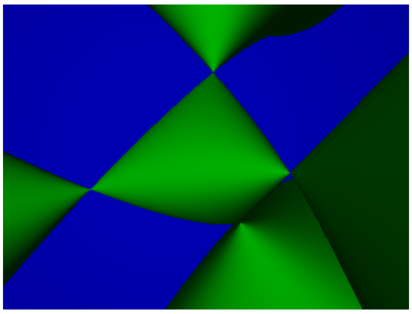


Figure 4.1 – The Cayley spectrahedron

This spectrahedron appears in the semidefinite relaxation of the well-known MAXCUT *problem in graph theory, cf. for example [102]. The four points delimitating the convex region are singular points of the hypersurface $\{x \in \mathbb{R}^3 : \det A(x) = 0\}$ and correspond to the unique four rank-1 matrices in the pencil $A(x)$. In this case, one can verify directly the correctness of Theorem 4.4, since the boundary of $\mathscr{S}$ contains all four connected components of $\mathcal{D}_1 \cap \mathbb{R}^3$.*

Example 4.6. *Let*

$$A(x) = \begin{pmatrix} 1 & x_1 & 0 & x_1 \\ x_1 & 1 & x_2 & 0 \\ 0 & x_2 & 1 & x_3 \\ x_1 & 0 & x_3 & 1 \end{pmatrix}.$$

The spectrahedron $S = \{x \in \mathbb{R}^3 : A(x) \succeq 0\}$ is known as the pillow, see also [23, Sec. 5.1.1]. As for Figure 4.1, it is pictured in Figure 4.2 with the help of the raytracing software POVRAY, *freely distributed and donwloadable at* http://www.povray.org.

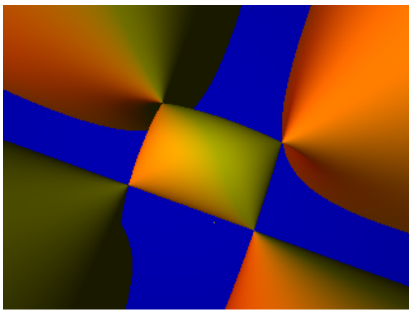


Figure 4.2 – The pillow and its algebraic boundary

The Zariski closure of its boundary is the real trace of the complex hypersur-

face defined by the vanishing of

$$\det A(x) = 1 - x_3^2 - x_2^2 - 2x_1^2 + x_1^2 x_3^2 - 2x_1^2 x_2 x_3 + x_1^2 x_2^2.$$

As is clear from Figure 4.2, the determinantal hypersurface consists of four branches arising from the convex set S. The boundary of S contains 4 singular points of the determinantal hypersurface, where $A(x)$ has rank 2. Their coordinates can be found by computing a Gröbner basis of the ideal generated by the 3×3 minors of A, namely $\{2x_1^2 - 1, 2x_3^2 - 1, x_2 + x_3\}$. In particular, these four points are contained in the hyperplane $x_2 + x_3 = 0$.

Example 4.7. *We consider the following matrix:*

$$A(x) = \begin{pmatrix} 1 & 0 & x_1 & 0 & -\frac{3}{2} - x_2 & x_3 \\ 0 & -2x_1 & \frac{1}{2} & x_2 & -2 - x_4 & -x_5 \\ x_1 & \frac{1}{2} & 1 & x_4 & 0 & x_6 \\ 0 & x_2 & x_4 & -2x_3 + 2 & x_5 & \frac{1}{2} \\ -\frac{3}{2} - x_2 & -2 - x_4 & 0 & x_5 & -2x_6 & \frac{1}{2} \\ x_3 & -x_5 & x_6 & \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix}.$$

This is the Gram matrix of the trivariate polynomial

$$f(u_1, u_2, u_3) = u_1^4 + u_1 u_2^3 + u_2^4 - 3u_1^2 u_2 u_3 - 4u_1 u_2^2 u_3 + 2u_1^2 u_3^2 + u_1 u_3^3 + u_2 u_3^3 + u_3^4,$$

that is $f = v' A(x) v$ where $x = (x_1, x_2, x_3, x_4, x_5, x_6)$ is any real vector and

$$v = (u_1^2, u_1 u_2, u_2^2, u_1 u_3, u_2 u_3, u_3^2)$$

is the monomial basis of the vector space of polynomials of degree 2 in u_1, u_2, u_3.

The polynomial f is nonnegative and hence, since it is homogeneous of degree 4 in 3 variables, by Hilbert's theorem (cf. [83]) it is a sum of at most three squares of polynomials in $\mathbb{R}[u_1, u_2, u_3]$. Scheiderer proved in [142] that f does not admit a sum-of-squares decomposition in $\mathbb{Q}[u_1, u_2, u_3]$, that is, the summands in the sos-decomposition cannot be chosen to have rational coefficients, answering a question of Sturmfels. We will compute the smallest rank of Scheiderer's spectrahedron in Section 5.3.5.

4.4 The algorithm

In this section we describe the algorithm SolveLMI, taking as input a symmetric linear matrix $A(x)$ and returning as output a point $x \in \mathbb{R}^n$ such that $A(x)$ is positive semi-definite, if such a point exists; otherwise it detects the emptiness of the spectrahedron $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$. The coordinates of this feasible point can be given as the solution of a system of linear equations (in case the pencil $A(x)$ contains the null matrix) or they are encoded by a rational univariate parametrization.

The formal description of SolveLMI is shown further in Section 4.4.3 and the analysis of its correctness is contained in Section 4.4.4.

4.4.1 Expected dimension of symmetric low rank loci

We first compute the dimension of algebraic sets $\mathcal{D}_r$, for $r = 0, \ldots, m-1$ and when A is a generic symmetric pencil. This can be done according to classical results about the dimension of the set of singular symmetric matrices and of their strata depending on the rank. We review in the next lemma this classical proof.

Lemma 4.8. *There exists a non-empty Zariski open subset $\mathscr{A} \subset \mathbb{S}_m^{n+1}(\mathbb{C})$ such that, if $A \in \mathscr{A} \cap \mathbb{S}_m^{n+1}(\mathbb{Q})$, for all $r = 0, \ldots, m-1$, the set $\mathcal{D}_r$ is either empty or it has dimension $n - \binom{m-r+1}{2}$.*

Proof. Let $\tilde{x}$ denote the vector of $m(m+1)/2$ variables $x_{i,j}, 1 \leq i \leq j \leq m$, and let $X \in \mathbb{S}_m(\mathbb{Q}[\tilde{x}])$ be the symmetric matrix with entries $x_{i,j}$.

Let $\mathcal{Z} = Z_{\mathbb{C}}(\text{minors}\,(r+1, X)) \subset \mathbb{C}^{m(m+1)/2}$. Let $\mathbb{G}(m-r, m)$ be the Grassmannian of $(m-r)-$planes in $\mathbb{C}^m$: it is an affine variety of dimension $r(m-r)$ (*cf.* [62, Lec. 6]). Let

$$\mathcal{I} = \{(X, \mathcal{H}) \in \mathbb{C}^{m(m+1)/2-1} \times \mathbb{G}(m-r, m) : \mathcal{H} \subset \ker(X)\}.$$

Let π_1 and π_2 be the projections of $\mathcal{I}$ respectively onto the first and the second factor. Then π_2 maps $\mathcal{I}$ surjectively onto $\mathbb{G}(m-r, m)$, and for $\mathcal{H} \in \mathbb{G}(m-r, m)$, then $\dim \pi_2^{-1}(\mathcal{H}) = r(r+1)/2$. To check this last dimension count, suppose without loss of generality that $\mathcal{H}$ is generated by the first $m-r$ vectors of the standard basis: then $\pi_2^{-1}(\mathcal{H})$ is the set of symmetric matrices such that the first $m-r$ columns and, hence, the first $m-r$ rows, are zero.

We deduce by the Theorem on the Dimension of Fibers [145, Sec. 6.3, Th. 7] that $\mathcal{I}$ is irreducible of dimension $r(m-r) + r(r+1)/2$. Hence $\mathcal{Z} = \pi_1(\mathcal{I})$ is irreducible, of dimension $r(m-r) + r(r+1)/2$ (and codimension $\binom{m-r+1}{2}$) since any fiber of π_1 is finite. We conclude that $\mathcal{D}_r$ has the claimed dimension by applying Bertini's theorem as in the proof of Proposition 3.5, page 77. *q.e.d.*

We call the value of the dimension of $\mathcal{D}_r$ in Lemma 4.8 the expected dimension of $\mathcal{D}_r$.

4.4.2 Notation

We start with a description of the algebraic sets built during SolveLMI. These sets share some features with those presented in Chapter 3 for algorithm LowRank. We also have to build an incidence variety and to prove that this set is smooth and equidimensional when the input linear matrix is generic.

We emphasize how the intrinsic structure induced by the symmetry of $A(x)$ generates explicit relations among the generators of such polynomial systems, and how some of them can be eliminated.

Incidence varieties for symmetric pencils

Let $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$ be an $n-$variate $m \times m$ symmetric linear matrix, and let $0 \leq r \leq m-1$. As in Chapter 3, Section 3.3.2, we introduce lifting variables $y = (y_{i,j})_{1 \leq i \leq m, 1 \leq j \leq m-r}$ and we build an algebraic set whose projection on the $x-$space is contained in the algebraic set $\mathcal{D}_r$. Let

$$Y(y) = \left(\begin{array}{ccc} y_{1,1} & \cdots & y_{1,m-r} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ y_{m,1} & \cdots & y_{m,m-r} \end{array} \right).$$

For $\iota = \{i_1, \ldots, i_{m-r}\} \subset \{1, \ldots, m\}$, with $\#\iota = m-r$, we denote by Y_ι the $(m-r) \times (m-r)$ sub-matrix of $Y(y)$ obtained by isolating the rows indexed by ι. There are $\binom{m}{r}$ such sub-matrices. For any choice of indices $\iota = \{i_1, \ldots, i_{m-r}\}$ and for any matrix $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$, we define the set

$$\mathcal{V}_r(A, \iota, S) = \{(x, y) \in \mathbb{C}^n \times \mathbb{C}^{m(m-r)} : A(x)Y(y) = 0, Y_\iota - S = 0\}.$$

Since S has full rank, the previous polynomial relations imply that Y has full rank and that the projection of $\mathcal{V}_r$ over the $x-$space is by definition contained in the low rank locus $\mathcal{D}_r$. Often, we will have $S = \mathbb{I}_{m-r}$, and, in this case, we simplify the notation by denoting $\mathcal{V}_r(A, \iota, \mathbb{I}_{m-r})$ by $\mathcal{V}_r(A, \iota)$, and $f(A, \iota, \mathbb{I}_{m-r})$ by $f(A, \iota)$. We denote by $f(A, \iota, S)$, or simply by f, the polynomial system defining $\mathcal{V}_r(A, \iota, S)$. For $M \in \mathrm{GL}_n(\mathbb{C})$ we denote by $f(A \circ M, \iota, S)$ the entries of $A(M\,x)Y(y)$ and $Y_\iota - S$, and by $\mathcal{V}_r(A \circ M, \iota, S)$ its zero set. We also denote by $U_\iota \in \mathbb{M}_{m-r,m}(\mathbb{Q})$ the full rank matrix whose entries are in $\{0, 1\}$, and such that $U_\iota Y(y) = Y_\iota$. By simplicity we call U_ι the boolean matrix with index ι.

We finally remark the similarity between the polynomial system

$$A(x)Y(y) = 0$$

and the so-called *complementarity conditions* for the solutions of a couple of primal-dual semidefinite program, see for example [117, Th. 3]. The difference is that, in our case, the special size of $Y(y)$ and the affine constraint $Y_\iota = S$ force a rank condition on $Y(y)$ and hence on $A(x)$.

Eliminating redundancies

The polynomial system defining $\mathcal{V}_r(A, \iota, S)$ contains redundancies induced by polynomial relations between its generators. These relations can be eliminated to obtain a minimal polynomial system defining the incidence variety, and allowing to compute the codimension of $\mathcal{V}_r$.

Lemma 4.9. *Let $M \in \mathrm{GL}_n, (\mathbb{C})$. Let $\iota \subset \{1, \ldots, m\}$, with $\#\iota = m-r$, and $S \in \mathrm{GL}_{m-r}, (\mathbb{Q})$. Let $A \in \mathbb{S}_m^{n+1}(\mathbb{Q})$, and let $f \in \mathbb{Q}[x, y]^{m(m-r)+(m-r)^2}$ be the* *polynomial system defining $\mathcal{V}_r$. Then we can construct a system $f_{red} \subset f$ of length $m(m-r)+\binom{m-r+1}{2}$ such that $\langle f_{red}\rangle = \langle f\rangle$.*

Proof. In order to simplify notations and without loss of generality we suppose $M = \mathbb{I}_n$, $S = \mathbb{I}_{m-r}$ and $\iota = \{1,\dots,m-r\}$. We substitute $Y_\iota = \mathbb{I}_{m-r}$ in $A(x)Y(y)$, and we denote by $g_{i,j}$ the $(i,j)-$th entry of the resulting matrix. We denote by f_{red} the following system:

$$f_{red} = (g_{i,j} \text{ for } i \leq j,\ Y_\iota - \mathbb{I}_{m-r})\,.$$

We claim that for $1 \leq i \neq j \leq m-r$, then

$$g_{i,j} \equiv g_{j,i} \mod \langle g_{k,\ell},\, k > m-r\rangle\,,$$

which implies that f_{red} verifies the statement. Let $a_{i,j}$ denote the $(i,j)-$th entry of $A(x)$. Let $i<j$ and write

$$g_{i,j} = a_{i,j} + \sum_{\ell=m-r+1}^{m} a_{i,\ell}y_{\ell,j} \text{ and } g_{j,i} = a_{j,i} + \sum_{\ell=m-r+1}^{m} a_{j,\ell}y_{\ell,i}.$$

We deduce that $g_{i,j} - g_{j,i} = \sum_{\ell=m-r+1}^{m} a_{i,\ell}y_{\ell,j} - a_{j,\ell}y_{\ell,i}$ since A is symmetric. Also, modulo the ideal $\langle g_{k,\ell},\, k > m-r\rangle$, and for $\ell \geq m-r+1$, one can explicit $a_{i,\ell}$ and $a_{j,\ell}$, by using polynomial relations $g_{\ell,i}=0$ and $g_{\ell,j}=0$, as follows:

$$\begin{aligned} g_{i,j} - g_{j,i} &\equiv \sum_{\ell=m-r+1}^{m}\left(-\sum_{t=m-r+1}^{m} a_{\ell,t}y_{t,i}y_{\ell,j} + \sum_{t=m-r+1}^{m} a_{\ell,t}y_{t,j}y_{\ell,i}\right) \equiv \\ &\equiv \sum_{\ell,t=m-r+1}^{m} a_{\ell,t}\left(-y_{t,i}y_{\ell,j} + y_{t,j}y_{\ell,i}\right) \equiv 0 \mod \langle g_{k,\ell},\, k > m-r\rangle\,. \end{aligned}$$

This last congruence concludes the proof. *q.e.d.*

Example 4.10. *We explicitly write down the redundancies shown in Lemma 4.9 for a simple example. We consider a 3×3 symmetric matrix of unknowns, and the kernel corresponding to the configuration $\{1,2\} \subset \{1,2,3\}$. Hence, let*

$$\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \\ f_{31} & f_{32} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ y_{31} & y_{32} \end{pmatrix}.$$

We consider the classes of f_{12}, f_{12} in the quotient ring $\mathbb{Q}[x] \big/ \langle f_{31}, f_{32}\rangle$, deducing the following linear relation:

$$f_{12} - f_{21} = y_{32}x_{13} - y_{31}x_{23} \equiv -y_{31}x_{33}y_{32} + y_{32}x_{33}y_{31} = 0.$$

Lagrange systems

Let $f(A, \iota, S)$ be the polynomial system defining $\mathcal{V}_r(A, \iota, S)$. We set

$$c = m(m-r) + \binom{m-r+1}{2} \qquad \text{and} \qquad e = \binom{m-r}{2},$$

so that $\mathcal{V}_r \subset \mathbb{C}^{c+e}$ and $c = \# f_{red}$ (*cf.* Lemma 4.9 on page 120). We define, for a given $M \in \mathrm{GL}_n(\mathbb{C})$, the polynomial system $\ell = \ell(A \circ M, \iota, S)$, given by the coordinates of the map

$$\begin{array}{rccc} \ell : & \mathbb{C}^n \times \mathbb{C}^{m(m-r)} \times \mathbb{C}^{c+e} & \longrightarrow & \mathbb{C}^{n+m(m-r)+c+e} \\ & (x, y, z) & \longmapsto & (f, z' Df - (e_1', 0)), \end{array}$$

where $f = f(A \circ M, \iota, S)$ and $e_1 \in \mathbb{Q}^n$ is the first element of the standard basis. We define also $\mathcal{Z}(A \circ M, \iota, S) = Z_{\mathbb{C}}(\ell(A \circ M, \iota, S))$. When $S = \mathbb{I}_{m-r}$, we omit it in the previous notations.

4.4.3 Real root finding for symmetric linear matrices

Before giving a detailed formal description of SolveLMI, we describe its main subroutine LowRankSym, which is a variant for symmetric pencils of the algorithm of Chapter 3.

Let A be a symmetric linear pencil, and let $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$ be the associated spectrahedron. We recall that by Theorem 4.4, the set $\mathscr{S}$ (actually, its Euclidean boundary) contains at least one connected component of the real set $\mathcal{D}_r \cap \mathbb{R}^n$, where r minimizes the rank of A over $\mathscr{S}$. Moreover, as remarked at the end of Section 4.2, Theorem 4.4 implies that it is sufficient to compute a finite set intersecting each connected component of $\mathcal{D}_r \cap \mathbb{R}^n$ which does not meet $\mathcal{D}_{r-1} \cap \mathbb{R}^n$.

Genericity properties

We define the following properties for a symmetric linear matrix $A \in \mathbb{S}_m^{n+1}(\mathbb{Q})$:

- Property P_1. We say that A satisfies P_1 if, for all $\iota \subset \{1, \ldots, m\}$, with $\#\iota = m - r$, and for all $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$, the incidence variety $\mathcal{V}_r(A, \iota, S)$ is either empty or smooth and equidimensional. We will always suppose $S = \mathbb{I}_{m-r}$ without loss of generality.

- Property P_2. We say that A satisfies P_2 if, for all r such that $n < \binom{m-r+1}{2}$, the algebraic set $\mathcal{D}_r$ has the expected dimension. By Lemma 4.8, this means that $\mathcal{D}_r = \emptyset$. Property P_2 holds generically in $\mathbb{S}_m^{n+1}(\mathbb{Q})$, as shown by Lemma 4.8.

We also define the following properties for a polynomial system $f \subset \mathbb{Q}[x]$ and a Zariski open set $\mathscr{O} \subset \mathbb{C}^n$:

- Property Q. Suppose that $f \subset \mathbb{Q}[x]$ generates a radical ideal and that it defines an algebraic set of codimension c, and let $\mathscr{O} \subset \mathbb{C}^n$ be a Zariski open set. We say that f satisfies Q in $\mathscr{O}$, if the rank of Df is c in $Z_{\mathbb{C}}(\langle f \rangle) \cap \mathscr{O}$.

Algorithm

We give next the formal description of the main body of algorithm $\mathsf{LowRankSym}$, taking as input A satisfying Properties P_1 and P_2, and $r \in \{0, \ldots, m-1\}$, and returning a sample subset of $\mathcal{D}_r \cap \mathbb{R}^n$. Consistently with the analogous algorithm for generic linear matrices presented in Chapter 3, we prefer to stress on the recursive feature by presenting, first, its preamble.

$\mathsf{LowRankSym}(A, r)$

Input: A symmetric $n-$variate linear matrix $A(x)$ of size m, encoded by the $m(m+1)(n+1)/2$ rational entries of $A_0, A_1, \ldots, A_n$, and an integer $1 \leq r \leq m-1$;

Output: Either the empty list [ ], if and only if $\mathcal{D}_r \cap \mathbb{R}^n = \emptyset$, or an error message stating that the genericity assumptions are not satisfied, or a rational parametrization $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \in \mathbb{Q}[t]^{n+2}$, such that for every connected component $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$, with $\mathcal{C} \cap \mathcal{D}_{r-1} = \emptyset$, there exists $t^* \in Z_{\mathbb{R}}(q_{n+1})$ with $(q_1(t^*)/q_0(t^*), \ldots, q_n(t^*)/q_0(t^*)) \in \mathcal{C}$.

Procedure:

1. if $n < \binom{m-r+1}{2}$ then return [ ];
2. for $\iota \subset \{1, \ldots, m\}$ with $\#\iota = m - r$ do
   - if $\mathsf{IsReg}((A, \iota)) =$ `false` then return("the input is not generic");
3. return($\mathsf{LowRankSymRec}(A, r)$).

The algorithm first excludes the case when the expected (complex) dimension of $\mathcal{D}_r$ is negative. In this case it returns immediately the empty list. Otherwise, it checks that the genericity assumptions are satisfied for the variety $\mathcal{V}_r(A, \iota)$, where ι runs over the subsets of $\{1, \ldots, m\}$ of cardinality $m - r$. This is done with the subroutine IsReg, *cf.* its definition on page 81.

If the incidence varieties are all regular, a recursive subroutine $\mathsf{LowRankSymRec}$ is called. We use subroutines Image, Union and Lift manipulating rational parametrizations, already defined in Chapter 3, Section 3.3.3, on page 81.

LowRankSymRec(A, r)

Procedure:

1. choose $M \in \mathrm{GL}_n(\mathbb{Q})$;
2. $q \leftarrow [\,]$; for $\iota \in \{1, \ldots, m\}$ with $\#\iota = m - r$ do
   - $q_\iota \leftarrow \mathsf{Image}(\mathsf{RatParProj}(\ell(A \circ M, \iota)), M^{-1})$;
   - $q \leftarrow \mathsf{Union}(q, q_\iota)$;
3. choose $t \in \mathbb{Q}$; $A \leftarrow (A_0 + tA_1, A_2, \ldots, A_n)$;
4. $q' \leftarrow \mathsf{Lift}(\mathsf{LowRankSymRec}(A, r), t)$;
5. return($\mathsf{Union}(q, q')$).

The routine RatParProj computes a rational parametrization of the projection of the Lagrange system $\ell(A \circ M, \iota)$ over the $x-$space. The correctness theorem of the main algorithm SolveLMI (shown in Theorem 4.14) relies on the correctness of its main sub-algorithm LowRankSym.

4.4.4 Main algorithm: description

The input is a symmetric $n-$variate linear matrix $A(x)$ of size m, that is the $m(m+1)(n+1)/2$ entries of $A_0, A_1, \ldots, A_n$, and the expected output is one of the following three alternatives:

- the empty list;
- a vector $x^* = (x_1^*, \ldots, x_n^*)$ such that $A(x^*) = 0$;
- a rational parametrization $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \in \mathbb{Q}[t]^{n+2}$, such that there exists $t^* \in Z_{\mathbb{R}}(q_{n+1})$ with

$$A(q_1(t^*)/q_0(t^*), \ldots, q_n(t^*)/q_0(t^*)) \succeq 0.$$

The different subroutines of SolveLMI are described next:

- SolveLinear: with input linear equations $A(x) = 0$, it returns the empty list iff $Z_{\mathbb{C}}(A(x))$ is empty, otherwise it returns x^* such that $A(x^*) = 0$;
- LowRankSym: with input a symmetric linear matrix A and an integer r, it returns either the empty list (if and only if $\mathcal{D}_r \cap \mathbb{R}^n = \emptyset$) or a rational parametrization q whose solutions meet each connected component $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$ such that $\mathcal{C} \cap \mathcal{D}_{r-1} = \emptyset$: its formal description is given in Section 4.4.3;
- CheckLMI: with input a symmetric linear matrix A and a rational parametrization $q = (q_0, q_1, \ldots, q_n, q_{n+1})$, it returns `true` if there exists $t^* \in Z_{\mathbb{R}}(q_{n+1})$ such that $A(q_1(t^*)/q_0(t^*), \ldots, q_n(t^*)/q_0(t^*)) \succeq 0$, and `false` otherwise.

SolveLMI(A)

Input: A symmetric $n-$variate linear matrix $A(x)$ of size m, encoded by the $m(m+1)(n+1)/2$ rational entries of $A_0, A_1, \ldots, A_n$;

Output: The empty list [ ] if and only if $\{x \in \mathbb{R}^n : A(x) \succeq 0\}$ is empty; or an error message stating that genericity assumptions are not satisfied, or, otherwise, either a vector $x^* = (x_1^*, \ldots, x_n^*)$ such that $A(x^*) = 0$, or a rational parametrization $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \in \mathbb{Q}[t]^{n+2}$, such that there exists $t^* \in Z_{\mathbb{R}}(q_{n+1})$ with $A(q_1(t^*)/q_0(t^*), \ldots, q_n(t^*)/q_0(t^*)) \succeq 0$.

Procedure:

1. $x^* \leftarrow \mathsf{SolveLinear}(A = 0)$; if $x^* \neq [\ ]$ then return(x^*);
2. for r from 1 to $m-1$ do:
   - $q \leftarrow \mathsf{LowRankSym}(A, r)$;
   - if $q =$ "the input is not generic" then return (q);
   - if $q \neq [\ ]$ then $b \leftarrow \mathsf{CheckLMI}(A, q)$;
   - if $b = \texttt{true}$ then return(q);
3. return([ ], "the spectrahedron is empty").

The formal description of SolveLMI is given on page 125.

4.4.5 Main algorithm: correctness

We prove that algorithm SolveLMI returns a correct output if genericity properties on input data and on random parameters chosen during its execution are satisfied. We write down a correctness proof in Theorem 4.14; it relies on some preliminary results that are described before. The proofs of these intermediate results are given in Section 4.4.6.

The genericity of Property P_2 has been proved in Lemma 4.8. Remark that, following the path of the proof of Proposition 3.6, page 84, one can easily check that, if a linear matrix $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$ satisfies P_2, and if $t \in \mathbb{Q}$ is generic, then the $(n-1)-$variate linear matrix $A_0 + tA_1 + x_2 A_2 + \cdots + x_n A_n$ also satisfies P_2. This property is technical and it is only needed to use induction in the proof of correctness (Theorem 4.14). In practical but also degenerate situations where P_2 does not hold, the algorithm returns the correct output. We show one example in Section 5.3.5 where P_2 does not hold.

The first result is a regularity theorem for the incidence varieties. We focus on property P_1 for the input matrix A. Its statement and the proof strategy are similar to Proposition 3.6, page 84 of Chapter 3.

Proposition 4.11. *Let $m, n, r \in \mathbb{N}$ with $0 \leq r \leq m-1$.*

1. *There exists a non-empty Zariski-open set $\mathscr{A} \subset \mathbb{S}_m^{n+1}(\mathbb{C})$ such that if $A \in \mathscr{A} \cap \mathbb{S}_m^{n+1}(\mathbb{Q})$, then A satisfies P_1;*

2. *if A satisfies P_1, there exists a non-empty Zariski open set $\mathscr{T} \subset \mathbb{C}$ such that if $t \in \mathscr{T} \cap \mathbb{Q}$, the matrix $A_0 + tA_1 + x_2A_2 + \cdots + x_nA_n$ satisfies P_1.*

The second proposition computes the dimension of Lagrange systems defined in Section 4.4.2. We show that the projection of their solution set over the $x-$space is finite and that this set meets the critical points of the restriction of the map $\pi_1 \colon x \to x_1$ to the incidence variety, similarly to Proposition 3.7.

Proposition 4.12. *Let $A \in \mathbb{S}_m^{n+1}(\mathbb{Q})$ satisfy P_1 and $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$. Then there exists a non-empty Zariski open set $\mathscr{M}_1 \subset \mathrm{GL}_n(\mathbb{C})$ such that, if $M \in \mathscr{M}_1 \cap \mathbb{M}_{n,n}(\mathbb{Q})$, for all $\iota \subset \{1, \ldots, m-r\}$ of cardinality $m-r$, the following holds:*

1. *The set $\mathcal{Z}(A \circ M, \iota, S)$ satisfies Q in $\{(x, y, z) : \operatorname{rank} A(M\,x) = r\}$;*

2. *the projection of $\mathcal{Z}(A \circ M, \iota, S) \cap \{(x, y, z) : \operatorname{rank} A(M\,x) = r\}$ on the $x-$space is empty or finite;*

3. *the projection of $\mathcal{Z}(A \circ M, \iota, S) \cap \{(x, y, z) : \operatorname{rank} A(M\,x) = r\}$ on (x, y) contains the set of critical points of the restriction of π_1 to $\mathcal{V}_r(A \circ M, \iota, S) \cap \{(x, y) : \operatorname{rank} A(M\,x) = r\}$.*

Finally, we show that, after a generic linear change of variables, closure properties of the projection maps restricted to $\mathcal{D}_r$. Also, when we restrict our original problem to the connected components of $\mathcal{D}_r \cap \mathbb{R}^n$ not meeting $\mathcal{D}_{r-1}$, the next proposition shows that it is sufficient to compute critical points on the incidence variety $\mathcal{V}_r$.

Proposition 4.13. *Let $\mathscr{A} \subset \mathbb{S}_m^{n+1}(\mathbb{C})$ be the set defined in Proposition 4.11 and let $A \in \mathscr{A}$. Let $d = \dim \mathcal{D}_r$. There exists a non-empty Zariski open set $\mathscr{M}_2 \subset \mathrm{GL}_n(\mathbb{C})$ such that if $M \in \mathscr{M}_2 \cap \mathbb{M}_{n,n}(\mathbb{Q})$, for any connected component $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$, the following holds:*

1. *for $i = 1, \ldots, d$, $\pi_i(M^{-1}\mathcal{C})$ is closed; further, for $t \in \mathbb{R}$ lying on the boundary of $\pi_1(M^{-1}\mathcal{C})$, then $\pi_1^{-1}(t) \cap M^{-1}\mathcal{C}$ is finite;*

2. *let t lie on the boundary of $\pi_1(M^{-1}\mathcal{C})$: for $x \in \pi_1^{-1}(t) \cap M^{-1}\mathcal{C}$, with $\operatorname{rank} A(M\,x) = r$, there exists $\iota \subset \{1, \ldots, m\}$ and $(x, y) \in \mathcal{V}_r(A \circ M, \iota)$ such that $\pi_1(x, y) = t$.*

As in Section 3.3.4 for algorithm $\mathsf{LowRank}$, to prove correctness of $\mathsf{SolveLMI}$ we must state some genericity hypothesis. We say that H holds if: input data A and all parameters generated by $\mathsf{SolveLMI}$ belong to the Zariski open sets defined in Proposition 4.11, 4.12 and 4.13, and A satisfies Property P_2.

We recall that $\mathbb{S}_{m,m}^{n+1}(\mathbb{Q})$ is the set of vectors $(A_0, A_1, \ldots, A_n)$ such that A_i is a symmetric matrix with entries in $\mathbb{Q}$, for $i = 0, \ldots, n$.

Theorem 4.14. *Let $A \in \mathbb{S}^{n+1}_{m,m}(\mathbb{Q})$ be the input symmetric linear matrix, and suppose that hypothesis* H *is satisfied. Let $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$ be the associated spectrahedron. Then two alternatives can hold:*

1. *$\mathscr{S} = \emptyset$: hence the output of* SolveLMI *with input A is the empty list;*

2. *$\mathscr{S} \neq \emptyset$: hence the output of* SolveLMI *with input A is either a vector x^* such that $A(x^*) = 0$, if such a point exists; or a rational parametrization $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \in \mathbb{Q}[t]^{n+2}$ such that there exists $t^* \in Z_{\mathbb{R}}(q_{n+1})$ with:*

   - *$A(q_1(t^*)/q_0(t^*), \ldots, q_n(t^*)/q_0(t^*)) \succeq 0$ and*
   - *$\operatorname{rank} A(q_1(t^*)/q_0(t^*), \ldots, q_n(t^*)/q_0(t^*)) = r(A)$.*

Proof. Suppose that the linear system defined by the entries of $A(x)$ has at least one solution. Hence, the routine SolveLinear with input A returns a vector x^* such that $A(x^*) = 0$. Since the zero matrix is positive semi-definite, we deduce that $x^* \in \mathscr{S} \neq \emptyset$ and that the rank of A attains its minimum on $\mathscr{S}$ at x^*. We deduce that, if $A(x) = 0$ has at least one solution, the algorithm returns a correct output.

Suppose now that $A(x)$ has rank greater than or equal to one on $\mathscr{S}$. We claim that the sub-routine LowRankSym is correct, in the following sense: with input the symmetric linear matrix $A(x)$ of size m and any $1 \leq r \leq m-1$, such that H holds, the output of LowRankSym(A, r) is a rational parametrization whose solutions meet each connected component $\mathcal{C}$ of $\mathcal{D}_r$ such that $\mathcal{C} \cap \mathcal{D}_{r-1} = \emptyset$.

We assume for the moment this claim and consider two possible alternatives:

1. $\mathscr{S} = \emptyset$. Consequently, CheckLMI outputs `false` at each iteration of Step 2 in SolveLMI. Hence the output of SolveLMI is the empty list, and correctness follows.

2. $\mathscr{S} \neq \emptyset$. Let $r \geq 1$ be the minimum rank attained by $A(x)$ on $\mathscr{S}$. Denote by $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$ a connected component such that $\mathcal{C} \cap \mathscr{S} \neq \emptyset$. By Theorem 4.4, we deduce that $\mathcal{C} \subset \mathscr{S}$, and that $\mathcal{C} \cap \mathcal{D}_{r-1} = \emptyset$, by the minimality of r. Let q be the output of LowRankSym at Step 2 of SolveLMI. The correctness hypothesis on LowRankSym implies that q defines a finite set whose solutions meet $\mathcal{C}$, hence $\mathscr{S}$. Consequently, the sub-routine CheckLMI returns `true` at Step 2, and hence the algorithm stops returning the correct output q.

We end the proof by showing that LowRankSym is correct. This is straightforwardly implied by the correctness of the recursive subroutine LowRankSymRec, which is proved below by using induction on the number of variables n.

For $n < \binom{m-r+1}{2}$, since H holds, the algebraic set $\mathcal{D}_r$ is empty, and hence LowRankSym returns the correct answer: [ ], the empty list.

Let $n \geq \binom{m-r+1}{2}$, and let (A, r) be the input. The induction hypothesis implies that for any $(n-1)-$variate symmetric linear matrix $\tilde{A}$ satisfying P_1, then $\mathsf{LowRankSymRec}$ with input $(\tilde{A}, r)$ returns a rational parametrization of a set meeting each connected component $\mathcal{C} \subset \tilde{\mathcal{D}}_r$ such that $\mathcal{C} \cap \tilde{\mathcal{D}}_{r-1} = \emptyset$, with $\tilde{\mathcal{D}}_r = \{x \in \mathbb{R}^{n-1} : \operatorname{rank} \tilde{A}(x) \leq r\}$.

Let $\mathcal{C} \subset \mathcal{D}_r$ be a connected component with $\mathcal{C} \cap \mathcal{D}_{r-1} = \emptyset$, and let M be the matrix chosen at Step 1. Hence, since H holds, by Proposition 4.13 the set $\pi_1(M^{-1}\mathcal{C})$ is closed.

First case. Suppose first that $\pi_1(M^{-1}\mathcal{C}) = \mathbb{R}$, let $t \in \mathbb{Q}$ be the rational number chosen at Step 3, and let $\tilde{A} = (A_0 + tA_1, A_2, \ldots, A_n) \in \mathbb{S}_m^n(\mathbb{Q})$. We deduce that $\pi_1^{-1}(t) \cap M^{-1}\mathcal{C} \neq \emptyset$ is the union of some connected components of the algebraic set $\tilde{\mathcal{D}}_r = \{x \in \mathbb{R}^{n-1} : \operatorname{rank} \tilde{A}(x) \leq r\}$ not meeting $\tilde{\mathcal{D}}_{r-1}$. Also, since A satisfies G, so does $A \circ M$; by Proposition 4.11, for all $\iota \subset \{1, \ldots, m\}$, for $f = f(A \circ M, \iota)$, the polynomial system $(f, x_1 - t)$ generates a smooth equidimensional algebraic set. By the induction assumption, $\mathsf{LowRankSymRec}$ with input $(\tilde{A}, r)$ returns at least one point in each connected component $\tilde{\mathcal{C}} \subset \tilde{\mathcal{D}}_r$ not meeting $\tilde{\mathcal{D}}_{r-1}$, hence one point in $\mathcal{C}$ by applying the subroutine Lift at Step 4. Correctness follows.

Second case. Otherwise, $\pi_1(M^{-1}\mathcal{C}) \neq \mathbb{R}$ and, since it is a closed set, its boundary is non-empty. Let t belong to the boundary of $\pi_1(M^{-1}\mathcal{C})$, and suppose w.l.o.g. that $\pi_1(M^{-1}\mathcal{C}) \subset [t, +\infty)$. Hence t is the minimum of the restriction of the map π_1 to $M^{-1}\mathcal{C}$. By Proposition 4.13, the set $\pi_1^{-1}(t) \cap M^{-1}\mathcal{C} \neq \emptyset$ is finite, and for all x in this set, $\operatorname{rank} A(M\,x) = r$. Fix $x \in \pi_1^{-1}(t) \cap M^{-1}\mathcal{C}$. By Proposition 4.13, there exists ι and $y \in \mathbb{C}^{m(m-r)}$ such that $(x, y) \in \mathcal{V}_r(A \circ M, \iota)$. Also, by Proposition 4.11, the set $\mathcal{V}_r(A \circ M, \iota)$ is smooth and equidimensional. As in the proof of Theorem 3.9, one deduces that (x, y) is a critical point of the restriction of π_1 to $\mathcal{V}_r(A \circ M, \iota)$ and that there exists z such that $(x, y, z) \in \mathcal{Z}(A \circ M, \iota)$. Hence, at Step 2, the routine $\mathsf{LowRankSymRec}$ outputs a rational parametrization q_ι, among whose solutions the vector x lies. *q.e.d.*

4.4.6 Proof of intermediate results

Proof of Proposition 4.11

Proof of Assertion 1. Suppose w.l.o.g. that $M = \mathbb{I}_n$ and $S = \mathbb{I}_{m-r}$. For $\iota \subset \{1, \ldots, m\}$ of cardinality $m-r$, let f_{red} be the polynomial system given by Lemma 4.9. We prove that there exists a non-empty Zariski open set $\mathscr{A}_\iota \subset \mathbb{S}_m^{n+1}(\mathbb{C})$ such that, if $A \in \mathscr{A}_\iota \cap \mathbb{S}_m^{n+1}(\mathbb{Q})$, f_{red} generates a radical ideal and $Z_{\mathbb{C}}(f_{red})$ is empty or equidimensional, of codimension the length of f_{red}, that is $m(m-r) + \binom{m-r+1}{2}$.

We conclude that, for $A \in \mathscr{A}_\iota$, A satisfies P_1. Then, we conclude by defining $\mathscr{A} = \cap_\iota \mathscr{A}_\iota$, non-empty and Zariski open.

Suppose w.l.o.g. that $\iota = \{1, \ldots, m-r\}$. We consider the map

$$\begin{array}{rccc} p: & \mathbb{C}^{n+m(m-r)} \times \mathbb{S}_m^{n+1}(\mathbb{C}) & \longrightarrow & \mathbb{C}^{m(m-r)+\binom{m-r+1}{2}} \\ & (x, y, A) & \longmapsto & f_{red} \end{array}$$

and, for a fixed $A \in \mathbb{S}_m^{n+1}(\mathbb{C})$, its section map

$$\begin{array}{rccc} p_A: & \mathbb{C}^{n+m(m-r)} & \longrightarrow & \mathbb{C}^{m(m-r)+\binom{m-r+1}{2}} \\ & (x,y) & \longmapsto & p(x,y,A). \end{array}$$

Remark that, for any A, $Z_{\mathbb{C}}(p_A)$ equals $\mathcal{V}_r(A,\iota)$.

Suppose $p^{-1}(0) = \emptyset$: this implies that, for all $A \in \mathbb{S}_m^{n+1}(\mathbb{C})$, $Z_{\mathbb{C}}(f_{red}) = \mathcal{V}_r(A,\iota) = \emptyset$, that is A satisfies P_1 for all A in $\mathscr{A}_\iota = \mathbb{S}_m^{n+1}(\mathbb{C})$.

If $p^{-1}(0) \neq \emptyset$, we prove below that 0 is a regular value of p, and we conclude on the existence of the Zariski open set $\mathscr{A}_\iota$ by Thom's Weak Transversality Theorem [139, Sec. 4.2] and by the Jacobian criterion (*cf.* [39, Th. 16.19] or Theorem 1.9), as in the proof of Proposition 3.6.

Let Dp be the Jacobian matrix of p: it contains the derivatives of polynomials in f_{red} with respect to variables x, y, A. We recall that A is a short-hand notation for the vector of symmetric matrices $(A_0, A_1, \ldots, A_n) \in \mathbb{S}_m^{n+1}(\mathbb{C})$; we denote by $a_{\ell,i,j}$ the variable encoding the $(i,j)-$th entry of the matrix A_ℓ. We isolate the columns of Dp corresponding to:

- the derivatives with respect to variables $\{a_{0,i,j} : i \leq m-r \text{ or } j \leq m-r\}$;
- the derivatives with respect to variables $y_{i,j}$ such that $i \in \iota$.

Let $(x,y,A) \in p^{-1}(0)$, and consider the evaluation of Dp at (x,y,A). The above columns contain the following non-singular blocks:

- the derivatives w.r.t. $\{a_{0,i,j} : i \leq m-r \text{ or } j \leq m-r\}$ of the entries of $A(x)Y(y)$ after reduction, that is $\mathbb{I}_{(m-r)(m+r+1)/2}$;
- the derivatives w.r.t. $\{y_{i,j} : i \in \iota\}$ of polynomials in $Y_\iota - \mathbb{I}_{m-r}$, that is $\mathbb{I}_{(m-r)^2}$.

Hence, the above columns define a maximal non-singular sub-matrix of Dp at (x,y,A), of size $m(m-r)+\binom{m-r+1}{2} = \#f_{red} = c$ (*cf.* Section 4.4.2). Indeed, the entries of $Y_\iota - \mathbb{I}_{m-r}$ do not depend on variables $a_{0,i,j}$. Since $(x,y,A) \in p^{-1}(0)$ is arbitrary, we deduce that 0 is a regular value of p, and we conclude. *q.e.d.*

Proof of Assertion 2. Fix $\iota \subset \{1,\ldots,m\}$ with $\#\iota = m-r$. Since A satisfies P_1, $\mathcal{V}_r(A,\iota)$ is either empty or smooth and equidimensional of codimension $m(m-r)+\binom{m-r+1}{2}$. Suppose first that $\mathcal{V}_r = \emptyset$. Hence for all $t \in \mathbb{C}$, $\mathcal{V}_r \cap \{x_1 - t = 0\} = \emptyset$, and we conclude by defining $\mathscr{T} = \mathbb{C}$. Otherwise, consider the restriction of the projection map $\pi_1 : (x,y) \to x_1$ to $\mathcal{V}_r$. By Sard's Lemma [139, Sec. 4.2], the set of critical values of the restriction of π_1 to $\mathcal{V}_r$ is included in a finite subset $\mathcal{H} \subset \mathbb{C}$. We deduce that, for $t \in \mathscr{T} = \mathbb{C} \setminus \mathcal{H}$, the matrix $A_0 + tA_1 + x_2A_2 + \cdots + x_nA_n$ satisfies P_1. *q.e.d.*

Proof of Proposition 4.12

We recall that for a given symmetric pencil $A \in \mathbb{S}_m^{n+1}(\mathbb{Q})$, $S \in \mathrm{GL}_{m-r}(\mathbb{C})$ and for $\iota \subset \{1, \ldots, m\}$ of cardinality $m - r$, we have denoted by $f = f(A, \iota, S)$ the polynomial system defining $\mathcal{V}_r(A, \iota, S)$. We set

$$c = m(m-r) + \binom{m-r+1}{2} \qquad \text{and} \qquad e = \binom{m-r}{2}.$$

Then f has length $c+e = m(m-r)+(m-r)^2$, and e is the number of redundancies that are eliminated by Lemma 4.9. By this lemma and by Proposition 4.11, we deduce that, for $A \in \mathscr{A}$:

- there exists $f_{red} \subset f$ of length c, such that $Z_{\mathbb{C}}(f_{red}) = Z_{\mathbb{C}}(f) = \mathcal{V}_r$;
- $\mathcal{V}_r$ is smooth and equidimensional of codimension c.

In particular, the rank of Df is constantly equal to c if evaluated along a point in $\mathcal{V}_r$.

Let $A(x)$ be a symmetric linear matrix, and consider the locally closed set

$$\widehat{\mathcal{D}_r} = \{x \in \mathbb{C}^n : \operatorname{rank} A(x) = r\}.$$

The set $\widehat{\mathcal{D}_r}$ is given by the union of sets $\mathcal{D}_r \cap \{x \in \mathbb{C}^n : \det N(x) \neq 0\}$ where N runs over all $r \times r$ sub-matrices of $A(x)$.

Fix S and ι as above. Let N be the upper left sub-matrix $r \times r$ sub-matrix of $A(x)$, and consider the corresponding block division of A as in (3.2), on page 100. We do not take any advantage of the symmetry of N, so that this block-subdivision is without loss of generality.

Let $\mathbb{Q}[x, y]_{\det N}$ be the local ring obtained by localizing $\mathbb{Q}[x, y]$ at $\langle \det N \rangle$. Let $Y^{(1)}$ (resp. $Y^{(2)}$) be the matrix obtained by isolating the first r (resp. the last $m - r$) rows of $Y(y)$. Hence, Lemma 3.18, page 100, gives local equations of $\mathcal{V}_r$ in $\{(x, y) : \det N(x) \neq 0\}$. These are given by

$$Y^{(1)} + N^{-1}PY^{(2)} = 0, \qquad \Sigma(N)Y^{(2)} = 0, \qquad Y_\iota - S = 0, \tag{4.1}$$

where $\Sigma(N) = R - P'N^{-1}P$ is the Schur complement of N in A.

Let $w \in \mathbb{C}^n$ be a non-zero vector and consider the projection map induced by w

$$\pi_w \colon (x_1, \ldots, x_n, y) \mapsto w_1x_1 + \cdots + w_nx_n.$$

For $A \in \mathscr{A}$ (given by Proposition 4.11), for all ι and S as above, the critical points of the restriction of π_w to $\mathcal{V}_r(A, \iota, S)$ are encoded by the polynomial system

$$f(A, \iota, S), \qquad (g, h) = z' \begin{pmatrix} Df \\ D\pi_w \end{pmatrix} = z' \begin{pmatrix} D_xf & D_yf \\ w' & 0 \end{pmatrix}, \tag{4.2}$$

where $z = (z_1, \ldots, z_{c+e}, 1)$ is a vector of Lagrange multipliers. Indeed, equations induced by (g, h) imply that the vector w is normal to the tangent space of $\mathcal{V}_r$ to (x, y).

We compute the dimension and prove regularity properties of the intersection of $Z_{\mathbb{C}}(f, g, h)$ with the Zariski open set $\{(x, y, z) : \operatorname{rank} A(x) = r\}$.

Lemma 4.15. *Let $S \in \mathrm{GL}_{m-r}(\mathbb{Q})$, and let $A \in \mathbb{S}_m^{n+1}(\mathbb{Q})$ satisfy P_1. There exists a non-empty Zariski open set $\mathscr{W} \subset \mathbb{C}^n$ such that, if $w \in \mathscr{W}$, and ι is as above, the following holds:*

1. *the system (f, g, h) in (4.2) satisfies Q in $\{(x, y, z) : \operatorname{rank} A(x) = r\}$;*
2. *the projection of $Z_{\mathbb{C}}(f, g, h) \cap \{(x, y, z) : \operatorname{rank} A(x) = r\}$ in the $x-$space is empty or finite;*
3. *the projection of $Z_{\mathbb{C}}(f, g, h) \cap \{(x, y, z) : \operatorname{rank} A(x) = r\}$ in the space of x, y, contains the critical points of the restriction of π_w to the locally closed set $\mathcal{V}_r \cap \{(x, y) : \operatorname{rank} A(x) = r\}$.*

Proof of Assertion 1. The strategy relies on applying Thom Weak Transversality Theorem and Jacobian criterion, as in the proof of Lemma 3.20, page 103.

We prove below the following claim: given a $r \times r$ sub-matrix N of $A(x)$, there exists $\mathscr{W}_N \subset \mathbb{C}^n$ such that for $w \in \mathscr{W}_N$, (f, g, h) satisfies Q in $\{(x, y, z) : \det N \neq 0\}$. We straightforwardly deduce Assertion 1 by defining $\mathscr{W} = \bigcap_N \mathscr{W}_N$, where N runs over all $r \times r$ sub-matrices of $A(x)$.

Let $U_\iota \in \mathbb{C}^{(m-r) \times m}$ be the boolean matrix such that $U_\iota Y(y) = Y_\iota$, and let $U_\iota = (U_\iota^{(1)} \mid U_\iota^{(2)})$ be the subdivision with $U_\iota^{(1)} \in \mathbb{C}^{(m-r) \times r}$ and $U_\iota^{(2)} \in \mathbb{C}^{(m-r) \times (m-r)}$. We deduce from Lemma 4.1 the following local equations for $\mathcal{V}_r$:

$$Y^{(1)} + N^{-1} P Y^{(2)} = 0, \qquad \Sigma(N) Y^{(2)} = 0, \qquad U_\iota Y(y) - S = 0.$$

Similarly to Lemma 3.20, we deduce that the above equations are equivalent to

$$\Sigma(N) = 0, \qquad Y^{(1)} + N^{-1} P Y^{(2)} = 0, \qquad Y^{(2)} - (U_\iota^{(2)} - U_\iota^{(1)} N^{-1} P)^{-1} S = 0,$$

in the local ring $\mathbb{Q}[x, y]_{\det N}$. We collect the above equations in a system $\tilde{f}$, of length $c + e$. Hence, the Jacobian matrix of $\tilde{f}$ is

$$D\tilde{f} = \left(\begin{array}{ccc} D_x[\Sigma(N)]_{i,j} & 0_{(m-r)^2 \times m(m-r)} & \\ & \mathbb{I}_{r(m-r)} & \star \\ \star & 0 & \mathbb{I}_{(m-r)^2} \end{array} \right).$$

By hypothesis, the rank of $D\tilde{f}$ is constant and equal to c if evaluated at $(x, y) \in Z_{\mathbb{C}}(\tilde{f}) = \mathcal{V}_r(A, \iota, S) \cap \{(x, y) : \det N \neq 0\}$. We similarly define

$$(\tilde{g}, \tilde{h}) = z' \left(\begin{array}{cc} D\tilde{f} & \\ w' & 0 \end{array} \right)$$

with $z = (z_1, \ldots, z_{c+e}, 1)$. The structure of $D\tilde{f}$ implies that polynomial $\tilde{h}_i$ reads $z_{(m-r)^2+i}$, for $i = 1, \ldots, m(m-r)$, and hence it can be eliminated, together with the corresponding variables $z_{(m-r)^2+i}$. Hence, one can consider the equivalent equations $(\tilde{f}, \tilde{g}, \tilde{h})$ where the last $m(m-r)$ variables z do not appear in $\tilde{g}$.

Let us define the map

$$\begin{array}{rrcl} p: & \mathbb{C}^{n+c+e+m(m-r)} \times \mathbb{C}^n & \longrightarrow & \mathbb{C}^{n+c+e+m(m-r)} \\ & (x, y, z, w) & \longmapsto & (\tilde{f}, \tilde{g}, \tilde{h}) \end{array}$$

and, for $w \in \mathbb{C}^n$, its section map $p_w\colon (x, y, z) \mapsto p(x, y, z, w)$. In the last part of this proof, we show that 0 is a regular value of the map p, and we conclude.

We first exclude the trivial situation $p^{-1}(0) = \emptyset$, by defining in this case $\mathscr{W}_N = \mathbb{C}^n$.

Otherwise, let $(x, y, z, w) \in p^{-1}(0)$. We first observe that polynomials in $\tilde{f}$ just depend on variables x and y, hence their contribution in the Jacobian matrix Dp at (x, y, z, w) is the block $D\tilde{f}$, whose rank is c, since $(x, y) \in \mathcal{V}_r$. Hence, we deduce that the rank of Dp at (x, y, z, w) is at most $n + c + m(m-r)$. Further, by isolating the columns corresponding to

- the derivatives w.r.t. x, y,
- the derivatives w.r.t. $w_1, \ldots, w_n$, and
- the derivatives w.r.t. $z_{(m-r)^2+i}, i = 1, \ldots, m(m-r)$,

one obtains a $(n + c + e + m(m-r)) \times (2n + 2m(m-r))$ sub-matrix of Dp with rank $n + c + m(m-r)$. *q.e.d.*

Proof of Assertion 2. From Assertion 1 we deduce that the locally closed set

$$\mathcal{E} = Z_{\mathbb{C}}(f, g, h) \cap \{(x, y, z) : \operatorname{rank} A(x) = r\}$$

is empty or $e-$equidimensional. If it is empty, we are done. Suppose that it is $e-$equidimensional. Consider the projection map

$$\begin{array}{rrcl} \pi_x: & \mathbb{C}^{n+m(m-r)+c+e} & \longrightarrow & \mathbb{C}^n \\ & (x, y, z) & \longmapsto & x \end{array}$$

and its restriction to $\mathcal{E}$. Let $x^* \in \pi_x(\mathcal{E})$. Then $\operatorname{rank} A(x^*) = r$ and there exists a unique $y \in \mathbb{C}^{m(m-r)}$ such that $f(x^*, y) = 0$. Hence the fiber $\pi_x^{-1}(x^*)$ is isomorphic to the linear space defined by

$$\Big\{(z_1, \ldots, z_{c+e}) : (z_1, \ldots, z_{c+e})Df = (w', 0)\Big\}.$$

Since the rank of Df is c, one deduces that $\pi_x^{-1}(x^*)$ is a linear space of dimension e, and by the Theorem on the Dimension of Fibers [145, Sec. 6.3, Theorem 7] we deduce that $\pi_x(\mathcal{E})$ has dimension 0. *q.e.d.*

Proof of Assertion 3. Since the set $\mathcal{V}_r \cap \{(x,y) : \operatorname{rank} A(x) = r\}$ is smooth and equidimensional, by [139, Lem. 3.2.1], for $w \neq 0$, the set $\operatorname{crit}(\pi_w, \mathcal{V}_r)$ is the set of points $(x,y) \in \mathcal{V}_r$ such that the matrix

$$D(f, \pi_w) = \begin{pmatrix} Df \\ D\pi_w \end{pmatrix}$$

has a rank $\leq c$. In particular there exists $z = (z_1, \ldots, z_{c+e}, z_{c+e+1}) \neq 0$, such that $z' D(f, \pi_w) = 0$. One can exclude that $z_{c+e+1} = 0$, since this implies that Df has a non-zero vector in the left kernel, which contradicts the fact that $A \in \mathscr{A}$. Hence w.l.o.g. we deduce that $z_{c+e+1} = 1$ and we conclude. *q.e.d.*

Proof of Proposition 4.12. Define $\mathscr{M}_1$ as the set of matrices $M \in \mathrm{GL}_n(\mathbb{C})$ such that the first row of M^{-1} is contained in the set $\mathscr{W}$ defined in Lemma 4.15. The proof of all assertions follows from Lemma 4.15 since, for $M \in \mathscr{M}_1$, one gets

$$\begin{pmatrix} Df(A \circ M, \iota, S) \\ e_1' \quad 0 \cdots 0 \end{pmatrix} = \begin{pmatrix} Df(A, U, S) \circ M \\ w' \quad 0 \cdots 0 \end{pmatrix} \begin{pmatrix} M & 0 \\ 0 & \mathbb{I}_{m(m-r)} \end{pmatrix}, \tag{4.3}$$

where w' is the first row of M^{-1}. Indeed, for $z = (z_1, \ldots, z_{c+e})$, we deduce from the previous relation that the set of solutions to the equations

$$f(A, \iota, S) = 0, \qquad z' Df(A, \iota, S) = (w', 0) \tag{4.4}$$

is the image of the set of solutions of

$$f(A \circ M, \iota, S) = 0, \qquad z' Df(A \circ M, \iota, S) = (e_1', 0) \tag{4.5}$$

by the linear map

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} M^{-1} & 0 & 0 \\ 0 & \mathbb{I}_{m(m-r)} & 0 \\ 0 & 0 & \mathbb{I}_{c+e} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

This last fact is straightforward since from (4.3) we deduce that system (4.5) is equivalent to

$$f(A \circ M, \iota, S) = 0, \qquad z' \left(Df(A, \iota, S) \circ M \right) = (w', 0).$$

Hence the three assertions of Proposition 4.12 are straightforwardly deduced by Lemma 4.15. *q.e.d.*

Proof of Proposition 4.13

Proof of Assertion 1. We denote by $\mathscr{M}_2 \subset \mathrm{GL}_n(\mathbb{C})$ the non-empty Zariski open set defined in Proposition 2.4, page 67, for $\mathcal{Z} = \mathcal{D}_r$. Hence, for $M \in \mathscr{M}_2$, we deduce that M satisfies $\mathsf{P}(\mathcal{D}_r)$ (*cf.* page 65), and consequently, by Proposition 2.2, page 65, that $\mathsf{Q}(M^{-1}\mathcal{D}_r)$ holds. We deduce that for $i = 1, \ldots, d$, and for any connected component $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$, the boundary of $\pi_i(M^{-1}\mathcal{C})$ is contained in $\pi_i(\mathcal{O}_{i-1}(M^{-1}\mathcal{D}_r) \cap M^{-1}\mathcal{C}) \subset \pi_i(M^{-1}\mathcal{C})$, and hence $\pi_i(M^{-1}\mathcal{C})$ is closed.

Moreover, let $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$ be a connected component and let $t \in \mathbb{R}$ be in the boundary of $\pi_1(M^{-1}\mathcal{C})$. Hence Proposition 2.5, page 67, implies that $\pi_1^{-1}(t) \cap M^{-1}\mathcal{C}$ is finite. *q.e.d.*

Proof of Assertion 2. Let $M \in \mathscr{M}_2$. Consider the open set

$$\mathscr{O} = \{(x, y) \in \mathbb{C}^{n+m(m-r)} : \operatorname{rank} A(M\,x) = r, \operatorname{rank} Y(y) = m - r\}.$$

Its projection $\pi_x(\mathscr{O})$ on the $x-$space is the locally closed set

$$M^{-1}\widehat{\mathcal{D}}_r = \{x \in \mathbb{C}^n \;: \operatorname{rank} A(M\,x) = r\} = M^{-1}(\mathcal{D}_r \setminus \mathcal{D}_{r-1}).$$

We consider the restriction of polynomial equations in $A(M\,x)Y(y) = 0$ to $\mathscr{O}$. By definition of $\mathscr{O}$, we can split the locally closed set $\mathscr{O} \cap Z_{\mathbb{C}}(A(M\,x)Y(y))$ into the union

$$\mathscr{O} \cap Z_{\mathbb{C}}(A(M\,x)Y(y)) = \bigcup_{\substack{\iota \subset \{1,\ldots,m\} \\ \#\iota = m-r}} \Big(\mathscr{O}_\iota \cap Z_{\mathbb{C}}(A(M\,x)Y(y))\Big).$$

with $\mathscr{O}_\iota = \{(x, y) : \det Y_\iota \neq 0\}$.

Let $\mathcal{C} \subset \mathcal{D}_r \cap \mathbb{R}^n$ be a connected component. Let t lie in the frontier of $\pi_1(M^{-1}\mathcal{C})$, and $x \in \pi_1^{-1}(t) \cap M^{-1}\mathcal{C}$ with $\operatorname{rank} A(M\,x) = r$. Hence there exists $\iota \subset \{1, \ldots, m\}$ such that x lies in the projection of $\mathcal{V}_r(A \circ M, \iota)$ on the $x-$space. Hence there exists y such that $\pi_x(x, y) = x$ and hence such that $\pi_1(x, y) = t$. *q.e.d.*

4.5 Complexity analysis

In order to estimate the complexity of the whole procedure SolveLMI we suppose, as in Chapter 3 for LowRank, that all genericity assumptions are satisfied, and consequently that no subroutine return error messages. We do not estimate the complexity of IsReg.

4.5.1 Degree bounds

We first provide a bound on the degree of the rational parametrizations. The reasoning is similar to that of Section 3.4.1, and it uses Multilinear Bézout bounds.

Proposition 4.16. *Let $A \in \mathbb{S}_m^{n+1}$ be the input of* SolveLMI*. Let $p_r = (m - r)(m + r + 1)/2$. If* H *holds, for all $\iota \subset \{1, \ldots, m\}$, the degree of the rational* *parametrization q returned by* LowRankSymRec *at Step 2 is bounded above by*

$$\theta(m,n,r) = \sum_{k \in \mathcal{G}_{m,n,r}} \binom{p_r}{n-k} \binom{n-1}{k+p_r-1-r(m-r)} \binom{r(m-r)}{k},$$

with $\mathcal{G}_{m,n,r} = \{k \,:\, \max\{0, n-p_r\} \leq k \leq \min\{n-\binom{m-r+1}{2}, r(m-r)\}\}$*. Moreover, for all* m,n,r*,* $\theta(m,n,r)$ *is bounded above by* $\binom{p_r+n}{n}^3$*.*

Proof. The proof is similar to that of Proposition 3.10 and Proposition 3.12. We simplify the polynomial system defining the incidence variety to a system of p_r bilinear equations with respect to variables $x = (x_1, \ldots, x_n)$ and $y = (y_{m-r+1,1}, \ldots, y_{m,m-r})$. Indeed, by Lemma 4.9, the incidence variety is defined by $Y_\iota - S = 0$ and by $m(m-r) - e = p$ entries of $A(x)Y(y)$, where $e = \binom{m-r}{2}$ (*cf.* Section 4.4.6); hence we just eliminate equations $Y_\iota - S = 0$ and the variables corresponding to the entries of Y_ι. Consequently, the Lagrange system can be also simplified, similarly to the proof of Proposition 3.10.

The bound $\theta(m,n,r)$, by [139, Ch. 11], is the coefficient of the monomial $s_x^n s_y^{r(m-r)} s_z^{p_r-1}$ in the expansion of

$$(s_x+s_y)^{p_r}(s_y+s_z)^{n-1}(s_x+s_z)^{r(m-r)}.$$

The estimate $\theta(m,n,r) \leq \binom{p_r+n}{n}^3$ can be obtained by applying the formula given in the proof of Proposition 3.12. *q.e.d.*

We straightforwardly deduce the following global estimate.

Corollary 4.17. *Let* $A \in \mathbb{S}_m^{n+1}$ *be the input of* SolveLMI*, and suppose that* $\mathscr{S}$ *is not empty. Let* $r(A)$ *be the minimum rank attained by* $A(x)$ *on* $\mathscr{S}$*. Let* $\theta(m,n,r)$ *be the bound computed in Proposition 4.16. If* H *holds, the sum of the degrees of the rational parametrizations computed during* SolveLMI *is bounded above by*

$$\sum_{r \leq r(A)} \binom{m}{r} \theta(m,n,r) \in \mathcal{O}\left(\max_{r \leq r(A)} \binom{m}{r} \binom{\frac{m^2+m}{2}+n}{n}^3 \right).$$

The degree of the rational parametrization whose solutions intersect $\mathscr{S}$ *is*

$$\binom{m}{r(A)} \theta(m,n,r(A)) \in \mathcal{O}\left(\binom{m}{r(A)} \binom{p_{r(A)}+n}{n}^3 \right).$$

Proof. We recall that, by Proposition 4.16, for any $\iota \subset \{1, \ldots, m\}$ of cardinality $m-r$, the degree of the rational parametrization returned by LowRankSymRec at Step 2 is bounded above by $\theta(m,n,r)$. The proof follows since:

1. the number of subsets $\iota \subset \{1, \ldots, m\}$ of cardinality $m-r$ is $\binom{m}{m-r} = \binom{m}{r}$;

2. SolveLMI stops when r reaches $r(A)$.

q.e.d.

4.5.2 Complexity estimates

We first summarize the the complexity estimates of the main subroutines: SolveLinear, CheckLMI,Project, Lift, Image and Union. Further, we analyze the main subroutine LowRankSym and give a Complexity Theorem for SolveLMI.

Complexity of minor subroutines

We first provide complexity estimates for all minor subroutines:

- The subroutine SolveLinear computes, if it exists, a solution of the linear system $A(x) = 0$. It can be essentially performed by Gaussian elimination. The complexity of solving a linear system of at most n variables is hence in $\mathcal{O}(n^3)$ arithmetic operations.

- The subroutine CheckLMI. This subroutine can be represented as follows. Let $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \subset \mathbb{Q}[t]$ be the rational parametrization in the input of CheckLMI, and let $A(x)$ be the symmetric pencil. The spectrahedron $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$ is the semi-algebraic set defined, *e.g.*, by the list of coefficients of the characteristic polynomial

$$p(s; x) = \det(s\,\mathbb{I}_m + A(x)) = s^m + p_1(x)s^{m-1} + \cdots + p_{m-1}(x)s + p_m(x).$$

  That is $\mathscr{S} = \{x \in \mathbb{R}^n : p_i(x) \geq 0, \forall\, i = 1, \ldots, m\}$. Now, we make the substitution $x_i \leftarrow q_i(t)/q_0(t)$ in $A(x)$ and compute the coefficients of $p(s, x(t))$, that are rational functions of the variable t. Hence CheckLMI boil down to deciding on the sign of m univariate rational functions over the finite set defined by $q_{n+1}(t) = 0$. We deduce that the complexity of CheckLMI is polynomial in m and on the degree of q_{n+1} (that is, on the degree of q).

- The complexities of Project, Lift, Image and Union have been given in Proposition 3.16, page 96. One can derive similar bounds by substituting $\delta(m, n, r)$ with the bound $\theta(m, n, r)$ computed in Proposition 4.16.

Complexity of the main subroutine and of the whole algorithm

As for algorithm LowRank described in Chapter 3, the complexity of LowRankSym can be retrieved by computing the complexity of the recursive subroutine LowRankSymRec, which also strictly depends on the computation of the rational parametrization.

We recall that for symmetric pencils, the simplified Lagrange system (*cf.* the proof of Proposition 4.16) contains p_r polynomials of multidegree bounded by $(1, 1, 0)$, $n - 1$ polynomials of multidegree bounded by $(0, 1, 1)$ and $r(m - r)$ polynomials of multidegree bounded by $(1, 0, 1)$. Since the reasoning of Section

3.4.2 on page 93 holds, we can use similarly the algorithm in [88] to build a geometric resolution of the Lagrange system.

Following *mutatis mutandis* the proof of Proposition 3.15 and of technical Lemma 3.14, one can similarly obtain the following complexity estimate for the subroutine RatParProj.

Proposition 4.18. *Let $A \in \mathbb{S}_m^{n+1}(\mathbb{Q})$ be the input of* SolveLMI *and $0 \leq r \leq m-1$. Let $\theta(m, n, r)$ be the bound defined in Proposition 4.16. Let $p_r = (m-r)(m+r+1)/2$. Then Step 2 of* LowRankSymRec *returns a rational parametrization within*

$$\mathcal{O}\tilde{}\left(\binom{m}{r}(n + p_r + r(m-r))^7 \, \theta(m, n, r)^2\right)$$

arithmetic operations over $\mathbb{Q}$.

Proof. Let ℓ be the simplified Lagrange system representing the input of RatParProj. We follow the same path as for the proof of Lemma 3.14, to compute a bound for the degree e' of the homotopy curve 3.1, just substituting p_r to $m(m-r)$. We deduce the following bound:

$$e' \in \mathcal{O}\left((n + p_r + r(m-r)) \min\{n, p_r\} \, \theta(m, n, r)\right).$$

We deduce the claimed complexity result by applying the estimate in [88, Prop. 6.1] (*cf.* page 93), and by recalling that there are $\binom{m}{r}$ many subsets of $\{1, \ldots, m\}$ of cardinality $m - r$. *q.e.d.*

We straightforwardly deduce the following complexity estimate for SolveLMI.

Theorem 4.19. *Let $A \in \mathbb{S}_m^{n+1}(\mathbb{Q})$ be the input symmetric pencil and suppose that* H *holds. Let $r(A)$ be the minimum rank attained by A on $\mathscr{S}$, if $\mathscr{S}$ is not empty, and let $r(A) = m - 1$ otherwise. Then algorithm* SolveLMI *runs within*

$$\mathcal{O}\tilde{}\left(n \sum_{r \leq r(A)} \binom{m}{r} (n + p_r + r(m-r))^7 \, \theta(m, n, r)^2\right)$$

arithmetic operations over $\mathbb{Q}$.

Proof. The proof follows since:

- From Proposition 4.18, we deduce that LowRankSymRec runs essentially within $\mathcal{O}\tilde{}(\binom{m}{r}(n + p_r + r(m-r))^7 \, \theta(m, n, r)^2)$ arithmetic operations;
- there are at most n recursive calls of LowRankSymRec in LowRankSym;
- SolveLMI stops when r reaches $r(A)$ if $\mathscr{S} \neq 0$, otherwise it stops when $r = m - 1$;
- the cost of subroutines SolveLinear, CheckLMI, Project, Lift, Image and Union is negligible.

q.e.d.

The complexity class in Theorem 4.19 is expressed directly as a function of the bounds $\theta(m,n,r)$ computed in Proposition 4.16. Remark, finally, that this value can be simplified by recalling that $\theta(m,n,r) \leq \binom{p_r+n}{n}^3$.

4.6 Exploiting structure: Hankel pencils

The algorithm SolveLMI deals with the problem of deciding the emptiness of spectrahedra, that is of sets given by positivity conditions on symmetric pencils. Typically, the interest is to design dedicated algorithms that potentially exploit relations among the entries of the matrix.

In this section we discuss a dedicated variant of the main subroutine LowRankSym for Hankel matrices. The main idea is to exploit the relations between the polynomial equations defining the incidence variety, arising from the Hankel structure of the linear matrix, and is detailed in the next two paragraphs. For more detailed results about the real root finding problem for Hankel matrices with rank defect we refer to the work [69], which is published in the proceedings of ISSAC 2015, Bath UK.

Hankel linear pencils are interesting in the literature, as shown by the following example extracted from [141].

Example 4.20. *Let $n \in \mathbb{N}$, and consider the Hankel pencil*

$$A(x) = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_n \\ x_1 & x_2 & x_3 & \cdots & x_{n+1} \\ x_2 & x_3 & x_4 & \cdots & x_{n+2} \\ \vdots & \vdots & \vdots & & \vdots \\ x_n & x_{n+1} & x_{n+2} & \cdots & x_{2n} \end{pmatrix},$$

subject to the affine constraint

$$1 - \sum_{j=0}^{n} \binom{n}{j} x_{2j} = 0.$$

The even moment curve (cf. [141, Sec. 5.2]) is the intersection of the real determinantal variety $\mathcal{D}_1 \cap \mathbb{R}^{2n+1}$, and of the associated spectrahedron $\mathscr{S} = \{x \in \mathbb{R}^{2n+1} : A(x) \succeq 0\}$. This curve is defined by the 2×2 minors of the matrix A.

The convex hull of the even moment curve is exactly the spectrahedron $\mathscr{S}$. The set $\mathscr{S}$ is, in particular, an orbitope [141], called the universal Carathéodory orbitope. We will apply our algorithm with the variant for Hankel matrices, to this example, in Section 5.3.3.

We remark here that the smallest rank attained by A on $\mathscr{S}$ is 1, and hence that $\mathscr{S}$ is the convex hull of its points of minimal rank. This situation is not typical in the geometry of spectrahedra, and in general to obtain the whole convex *set it is not sufficient to consider just the minimal rank point. In the work [84], the spectrahedra generated by rank 1 matrices have been studied.*

Kernel of Hankel matrices

For $\{h_1, \ldots, h_{2m-1}\} \subset \mathbb{Q}$, we recall that the matrix $H = (h_{i+j-1})_{1 \le i,j \le m} \in \mathbb{S}_m(\mathbb{Q})$ with constant skew-diagonals, is called a Hankel matrix. We denote by $\mathcal{H} \subset \mathbb{S}_m(\mathbb{Q})$ the $(2m-1)-$dimensional linear space of $m \times m$ Hankel matrices.

The structure of a Hankel matrix induces a structure on its kernel. By [64, Th. 5.1], if H is a Hankel matrix of rank at most p, then there exists a non-zero vector $y = (y_1, \ldots, y_{p+1}) \in \mathbb{Q}^{p+1}$ such that the columns of the $m \times (m-p)$ matrix

$$Y(y) = \begin{pmatrix} y_1 & 0 & \ldots & 0 \\ y_2 & y_1 & \ddots & \vdots \\ \vdots & y_2 & \ddots & 0 \\ y_{p+1} & \vdots & \ddots & y_1 \\ 0 & y_{p+1} & & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & y_{p+1} \end{pmatrix} \in \mathbb{M}_{m,m-p}(\mathbb{Q}[y])$$

generate a $(m-p)-$dimensional subspace of the kernel of H. For example, for $m = 5$ and $p = 2$ the matrix reads

$$Y(y) = \begin{pmatrix} y_1 & 0 & 0 \\ y_2 & y_1 & 0 \\ y_3 & y_2 & y_1 \\ 0 & y_3 & y_2 \\ 0 & 0 & y_3 \end{pmatrix}.$$

Hence a condition under which the matrix H has rank at most p is that $HY(y) = 0$ has a solution for some $y \neq 0$. Remark that also the product $HY(y)$ is Hankel and that the product $HY(y)$ can be written as a matrix-vector product $\tilde{H}y$, with $\tilde{H}$ a given rectangular Hankel matrix. Indeed, if H is the Hankel matrix defined by the vector $(h_1, \ldots, h_{2m-1}) \in \mathbb{Q}^{2m-1}$ as above, $H\,Y(y)$ is a rectangular Hankel matrix, of size $m \times (m-p)$, the set of whose entries coincide with the set of entries of the matrix-vector product

$$\tilde{H}y = \begin{pmatrix} h_1 & \ldots & h_{p+1} \\ \vdots & & \vdots \\ h_{2m-p-1} & \ldots & h_{2m-1} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_{p+1} \end{pmatrix}.$$

Incidence varieties for Hankel linear matrices

Suppose that the generating symmetric matrices $A_0, A_1, \ldots, A_n$ belong to $\mathcal{H}$. Hence, the associated symmetric pencil $A(x)$ is also a Hankel matrix for any $x \in \mathbb{R}^n$. Let $r \leq m-1$ be the maximum admissible rank for $A(x)$.

For $p \leq r$, let $\tilde{A}(x)$ be the associated rectangular matrix defined in the previous section. From [34, Cor. 2.2] one deduces that, for $p' \leq p$, the ideals $\langle \text{minors}\,(p'+1, A(x)) \rangle$ and $\langle \text{minors}\,(p'+1, \tilde{A}(x)) \rangle$ coincide. One deduces that $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$ satisfies $\text{rank}\, A(x) = p'$ if and only if it satisfies $\text{rank}\, \tilde{A}(x) = p'$.

Consider the algebraic set

$$\mathcal{D}_r = \{x \in \mathbb{C}^n : \text{rank}\, A(x) \leq r\}.$$

We can define a flag of incidence varieties along the set $\mathcal{D}_r$, dedicated to Hankel pencils, as follows. For $p \leq r$, we define

$$\mathcal{V}_p = \{(x, y) \in \mathbb{C}^n \times \mathbb{C}^{p+1} : \tilde{A}(x)y = 0, u'y - 1 = 0\},$$

for a given $u \in \mathbb{Q}^{p+1}$. The set $\mathcal{V}_p$ encodes the fact that the rank of $A(x)$ is at most p, for some $p \leq r$, and hence its projection on the $x-$space is contained in $\mathcal{D}_p \subset \mathcal{D}_r$.

Description of the variant and main theorem

The algorithm dedicated to Hankel pencils then proceeds similarly to LowRank-Sym, and relies on the following properties (that can be proved with the same techniques as for Proposition 4.11, 4.12 and 4.13).

1. Up to genericity assumptions on the Hankel matrices $A_0, A_1, \ldots, A_n$, all algebraic sets $\mathcal{V}_p$ are smooth and equidimensional, since the polynomial system f containing $u'y - 1$ and the set defined by $\tilde{A}(x)y = 0$ is smooth and equidimensional. Also, the ideal generated by f is generically radical. Moreover, up to genericity assumptions, the algebraic sets $\mathcal{D}_p$ have are empty or have the expected dimension $n - 2m + 2p + 1$.
2. Up to a generic change of variables $x \mapsto M\,x$, the image of any connected component $\mathcal{C} \subset \mathcal{D}_p \cap \mathbb{R}^n$ by the projection $\pi_1 \colon x \to x_1$ is closed (this can be directly deduced by applying Proposition 2.5).
3. The set of critical points of the restriction of π_1 to $\mathcal{D}_p \setminus \mathcal{D}_{p-1} \cap \mathbb{R}^n$ is finite. Then, since $\mathcal{D}_r = \cup_{1 \leq p \leq r} \mathcal{D}_p \setminus \mathcal{D}_{p-1}$, one can compute one point per connected component of $\mathcal{D}_r$ by computing one point in each connected component $\mathcal{C} \subset \mathcal{D}_p \cap \mathbb{R}^n$ such that $\mathcal{C} \cap \mathcal{D}_{p-1} = \emptyset$ as for LowRankSym, and iterating on p.

The previous three facts can be proved using Thom's Weak Transversality Theorem and Sard's Lemma as in Chapter 3 and 4. Also, the Lagrange systems encoding the critical points of the restriction of linear projections to the incidence varieties $\mathcal{V}_p$ have a bilinear structure, and their degree is bounded by multilinear Bézout bounds. Using the symbolic homotopy algorithm in [88] we conclude Main Theorem 5, that is the following complexity estimate for the special case of linear Hankel matrices. We restate below the main theorem.

Let A be a $n-$variate linear Hankel matrix of size m, and let $r \leq m-1$. Let f be the polynomial system defining the incidence variety $\mathcal{V}_r$. Suppose that the ideal $\langle f \rangle$ is radical, that $\mathcal{V}_r$ is smooth and equidimensional and that, for $0 \leq p \leq r$, $\mathcal{D}_p$ is empty or has the expected dimension $n-2m+2p+1$.

There exists a probabilistic algorithm such that, with input (A, r), returns in output a rational parametrization $q = (q_0, q_1, \ldots, q_n, q_{n+1}) \subset \mathbb{Q}[t]$ of degree at most

$$\binom{2m-r-1}{r} + \sum_{k=2m-2r}^{n} \sum_{p=0}^{r} \delta(m,k,p)$$

with $\delta(m,k,p) \in \mathcal{O}(\binom{k+2m-p-1}{k}^3)$, within

$$\mathcal{O}\left(\sum_{p=0}^{r} pn(2m-p)(pn(2m-p)(n+2m)^2 + (n+2m)^4)\delta(m,n,p)^2\right)$$

arithmetic operations over $\mathbb{Q}$. The set parametrized by q contains at least one point in each connected component of the real algebraic set $\mathcal{D}_r \cap \mathbb{R}^n$.

We provide next, in Chapter 5, Section 5.3.3, results of numerical experiments of an implementation of this speed-up for Hankel pencils.

4.7 First experiments

We report on some numerical experiments. A description of the implementation, which is under MAPLE, is given in the next chapter, in Section 5.1.

(m,r,n)	LowRankSym	deg	maxdeg	(m,r,n)	LowRankSym	deg	maxdeg
$(3,2,7)$	16	13	6	$(5,2,5)$	1	0	0
$(3,2,8)$	17	13	6	$(5,2,6)$	5	35	35
$(3,2,9)$	18	13	6	$(5,3,4)$	1592	110	90
$(4,2,3)$	2	10	10	$(5,4,5)$	858	105	40
$(4,2,4)$	9	40	30	$(6,3,5)$	2	0	0
$(4,2,5)$	29	82	42	$(6,3,6)$	704	112	112
$(4,3,6)$	24	40	16	$(6,4,3)$	9	35	35
$(4,3,7)$	26	40	16	$(6,5,3)$	591	116	80

Table 4.1 – Timings and degrees for dense symmetric linear matrices

Table 4.1 contains a first group of tests on randomly generated symmetric linear matrices. We recall that m is the size of the matrix, r is the expected rank and n is the number of variables. A complete list of timings and values for the output degree is given in Table 5.5, Section 5.3.1, page 153.

In the column LowRankSym we report the timing in seconds; the column deg contains the sum of the degrees of the rational parametrizations returned in output, and column maxdeg the maximum of such degrees. With deg $= 0$ and maxdeg $= 0$ we mean that the corresponding real determinantal variety $\mathcal{D}_r \cap \mathbb{R}^n$ is empty (hence the algorithm returns the empty list). The values of maxdeg corresponds to the algebraic degree of certain semidefinite programs, and we refer to Section 5.3.1 for a more complete discussion.

Chapter 5

Software description and numerical experiments

In Chapter 3 and 4 we have presented two exact algorithms. The goal is to solve emptiness problems for real algebraic or semi-algebraic sets with determinantal structure. The correctness of these algorithms relies on results of Chapter 2 and bounds for their complexity have been discussed and provided explicitly. Our next goal is to report results of experiments to reflect the complexity estimates.

This section contains a general documentation of a MAPLE library implemented during this PhD thesis. The name of this library is SPECTRA, and will be made available in September 2015. It strongly relies on results described in Chapter 2, 3 and 4, and implements the aforementioned algorithms. We conclude our manuscript by discussing numerical results on dense matrices and on interesting examples of the literature.

5.1 SPECTRA: a library for real algebraic geometry and semidefinite optimization

The algorithms described in Chapter 3 and 4 can be implemented in any computer algebra system working in exact arithmetic performing computations using Gröbner bases and, possibly, computing rational parametrizations of zero-dimensional ideals. Examples of these software systems are MAPLE [109], MACAULAY2 [107] or MAGMA [108].

We present next our first implementation in the computer algebra system MAPLE. Its name is SPECTRA and it takes the form of a collection of scripts containing the functions implementing algorithms LowRank (*cf.* Chapter 3), LowRankSym (containing the dedicated variant for Hankel matrices) and SolveLMI (*cf.* Chapter 4).

The scripts contained in SPECTRA have to be understood as a first step towards a more general library containing implementations of algorithms using symbolic computation techniques for a large class of problems in real algebraic geometry and optimization. Indeed, the acronym of SPECTRA reads:

Semidefinite
Programming and
Exact
Computation
Towards
Real
Algebra.

The library SPECTRA can be loaded under MAPLE. It relies on the software FGB for computing Gröbner bases, developed by J.-C. Faugère [43]. The current version of FGB is available at: http://www-polsys.lip6.fr/~jcf/Software/FGb/index.html. The first step for using the scripts contained in SPECTRA is to load the main file in a MAPLE worksheet as follows:

```
read "main.mpl";
```

This will automatically load the two instrumental scripts:

- `proc.mpl`: basic procedures for computing with algebraic or semi-algebraic sets;
- `lowrank.mpl`: implementations of the algorithms described in Chapter 3 and 4.

Below, we briefly describe the implementations of algorithms LowRankSym and SolveLMI. Algorithm LowRank is similarly implemented.

Implementation of LowRankSym

The procedure implementing this subroutine is named the same: `LowRankSym`. It takes as input a couple (A, r) and returns a collection of rational parametrizations the union of whose solutions meets all connected components of $\mathcal{D}_r \cap \mathbb{R}^n$.

We list below the main aspects of such implementation:

- The computation of Gröbner bases is done using FGB. We use the implementations of the algorithms [41] and [42]. Given a list of polynomials `list` on variables `vars`, with `fgb_gbasis(list,0,vars,[])` one computes a Gröbner bases of the ideal generated by `list`; similarly `fgb_gbasis_elim` provides Gröbner bases of elimination ideals. We use this last command to eliminate Lagrange multipliers and kernel variables from Lagrange systems, and to compute a Gröbner basis of the obtained ideal.
- The rational parametrizations are computed by the function `fgb_matrixn`. This implements the new efficient versions of FGLM algorithm [48] obtained in [49, 47].

- In order to simplify computations, we do not introduce changes of variables. According to Lemma 4.15, one can equivalently compute critical points of projections over generic lines in the space of the set $\mathcal{D}_r$. This is what our implementation does in practice.

Implementation of SolveLMI

The expected input of the function `SolveLMI` is a symmetric linear matrix A. Then:

- It first solves the linear system $A(x) = 0$. This is done by performing Gaussian operations on the rows of A.
- Then it iterates the subroutine `LowRankSym` for ranks $r = 1, \ldots, m-1$. Any rational parametrization computed by subroutines of `LowRankSym` is the input of a routine `CheckLMI` checking whether one of the solutions lies on the spectrahedron. This is done in practice by:
  - applying the substitution $x_i = \frac{q_i(t)}{q_0(t)}$ in the linear matrix;
  - isolating all solutions t^* of the equation $q_{n+1}(t) = 0$;
  - for each solution t^* of $q_{n+1}(t) = 0$, computing the coefficients of the polynomial $s \mapsto \det(A(x(t^*)) + s\mathbb{I}_m)$ and checking numerically whether these are all positive.

A first example to run

Let us illustrate how SPECTRA works on a simple example, the spectrahedron $\mathscr{S}$ of Example 1.31, page 52 of Chapter 1. The real trace of its algebraic boundary is a cubic which factorizes into the product of a conic and a line intersecting into two points. At these points the defining matrix has rank 1 and is positive semidefinite.

We call function `SolveLMI` in our MAPLE worksheet with input the matrix A of Example 1.31, as follows:

```
read "main.mpl";
A := [[1 - x1, x2, 0], [x2, 1 + x1, 0], [0, 0, -x2]];
SolveLMI(A);
```

The algorithm first verifies easily that $A(x) = 0$ has no solution by solving the associated linear system. Then it runs LowRankSym with input (A, r) for $r = 1, 2$. We analyze below the output messages for $r = 1$.

For $r = 1$ there are three possible kernel configurations since $\binom{3}{1} = 3$. We denote these configurations by $[1, 2], [2, 3], [1, 3]$ where $[i, j]$ means that we are considering the set $\{i, j\} \subset \{1, 2, 3\}$. Hence at Step 2 three rational parametrizations are computed. As remarked in Example 1.31, the two points where A has rank 1 are $(-1, 0)$ and $(1, 0)$, and their kernels are the linear spaces $\langle e_2, e_3 \rangle$ and $\langle e_1, e_3 \rangle$ respectively. Consequently, after the first step, the output is the empty list:

```
----------------------------------------------
Kernel Configuration : [1, 2]
Assumptions : satisfied
----------------------------------------------
Number of Variables : 2
Degree of the Parametrization : 0
Time for Grobner Bases Computation : 2.50
Time for Rational Parametrization : 0.00
Fiber for Reconstruction Next Step : -9*x-96*y
Real Solutions : 0
Time for Isolation of Real Solutions : 0.00
Output : Empty list []
----------------------------------------------
Number of Variables : 1
Degree of the Parametrization : 0
Time for Grobner Bases Computation : 4.58
Time for Rational Parametrization : 0.00
Fiber for Reconstruction Next Step : 75*y
Real Solutions : 0
Time for Isolation of Real Solutions : 0.00
Output : Empty list []
----------------------------------------------
```

Conversely, the remaining configurations allow to compute and isolate the coordinates of the two points, according to their kernel structure:

```
----------------------------------------------
Kernel Configuration : [1, 3]
Assumptions : satisfied
----------------------------------------------
Number of Variables : 2
Degree of the Parametrization : 1
Time for Grobner Bases Computation : 2.44
Time for Rational Parametrization : 0.00
Fiber for Reconstruction Next Step : -16*x+39*y
Real Solutions : 1
Time for Isolation of Real Solutions : 0.00
Output : [[x = [1., 1.], y = [0., 0.]]]
----------------------------------------------
Number of Variables : 1
Degree of the Parametrization : 0
Time for Grobner Bases Computation : 4.54
Time for Rational Parametrization : 0.00
Fiber for Reconstruction Next Step : 93*y
Real Solutions : 0
Time for Isolation of Real Solutions : 0.00
Output : Empty list []
----------------------------------------------
```

```
---------------------------------------------
Kernel Configuration : [2, 3]
Assumptions : satisfied
---------------------------------------------
Number of Variables : 2
Degree of the Parametrization : 1
Time for Grobner Bases Computation : 2.46
Time for Rational Parametrization : 0.00
Fiber for Reconstruction Next Step : 67*x-83*y
```

```
Real Solutions : 1
Time for Isolation of Real Solutions : 0.00
Output : [[x = [-1., -1.], y = [0., 0.]]]
--------------------------------------------
Number of Variables : 1
Degree of the Parametrization : 0
Time for Grobner Bases Computation : 4.64
Time for Rational Parametrization : 0.00
Fiber for Reconstruction Next Step : 93*y
Real Solutions : 0
Time for Isolation of Real Solutions : 0.00
Output : Empty list []
--------------------------------------------
```

One can also have a report on the computational timings of all steps: computation of Gröbner bases, computation of parametrizations and isolation of real solutions. Also, we display after `Fiber for Reconstruction Next Step` the random linear function that is used to extract critical points. This expression can be used to recover the eliminated variables from partial solutions computed (if any) at the successive steps.

For any configuration, the timing of these computations was around 7 seconds. For this tests and for all the experiments of this chapter, we have used the same machine, with the following characteristics:

Intel(R) Xeon(R) CPU E7540@2.00GHz 256 Gb of RAM.

5.2 Numerical experiments for Chapter 3

Now, we discuss numerical tests on randomly generated and on practical examples of linear matrices.

The arithmetic complexity of our algorithms has been computed, for LowRank, in Sections 3.3.4 and 3.4.2 of Chapter 3. We have also been interested in testing whether this complexity estimate can be made effective by a performing implementation.

While the complexity of the main routine RatPar is computed in Section 3.4.2 referring to the symbolic homotopy algorithm in [88], the practical computation of rational parametrizations is done using Gröbner bases. To compute Gröbner bases we use the engine FGb developed by J.-C. Faugere [43].

The first series of tests to evaluate the performance of LowRank is done with input random linear matrices. We comment this randomness in the next section, reporting on the obtained results. Further, in Section 5.2.2 we apply LowRank to two interesting examples of the literature.

5.2.1 Tests on generic input data

As stated in the Main Theorem of Chapter 3, page 73, our algorithm LowRank is probabilistic, since its correctness depends on the choice of random linear changes of variables. Theorem 3.9 ensures that when both input data and parameters are generic, the algorithm returns the correct and expected output.

(m, r, n)	RAG	SPECTRA	deg	maxdeg	(m, r, n)	RAG	SPECTRA	deg	maxdeg
$(3, 2, 2)$	0.2	6	9	6	$(5, 2, 3)$	0.9	0.5	0	0
$(3, 2, 3)$	0.3	7.5	21	12	$(5, 2, 4)$	1	0.5	0	0
$(3, 2, 4)$	0.9	9.5	33	12	$(5, 2, 5)$	1.6	0.5	0	0
$(3, 2, 5)$	5.1	13.5	39	12	$(5, 2, 6)$	3	0.6	0	0
$(3, 2, 6)$	15.5	15	39	12	$(5, 2, 7)$	4.2	0.7	0	0
$(3, 2, 7)$	31	16.5	39	12	$(5, 2, 8)$	8	0.7	0	0
$(3, 2, 8)$	109	18	39	12	$(5, 2, 9)$	∞	903	175	175
$(3, 2, 9)$	230	20	39	12	$(5, 3, 2)$	0.4	0.5	0	0
$(4, 2, 2)$	0.2	0.5	0	0	$(5, 3, 3)$	0.5	0.5	0	0
$(4, 2, 3)$	0.3	0.5	0	0	$(5, 3, 4)$	43	22	50	50
$(4, 2, 4)$	2.2	2.5	20	20	$(5, 3, 5)$	∞	5963	350	300
$(4, 2, 5)$	12.2	26	100	80	$(5, 4, 2)$	0.5	125	25	20
$(4, 2, 6)$	∞	593	276	176	$(5, 4, 3)$	10	167	105	80
$(4, 2, 7)$	∞	6684	532	256	$(5, 4, 4)$	∞	561	325	220
$(4, 2, 8)$	∞	42868	818	286	$(5, 4, 5)$	∞	5574	755	430
$(4, 2, 9)$	∞	120801	1074	286	$(6, 3, 3)$	4	1	0	0
$(4, 3, 3)$	1	8	52	36	$(6, 3, 4)$	140	1	0	0
$(4, 3, 4)$	590	18	120	68	$(6, 3, 5)$	∞	1	0	0
$(4, 3, 5)$	∞	56	204	84	$(6, 3, 6)$	∞	2	0	0
$(4, 3, 6)$	∞	114	264	84	$(6, 3, 7)$	∞	2	0	0
$(4, 3, 7)$	∞	124	284	84	$(6, 3, 8)$	∞	2	0	0
$(4, 3, 8)$	∞	124	284	84	$(6, 4, 2)$	0.6	40	0	0
$(4, 3, 9)$	∞	295	284	84	$(6, 4, 3)$	1	64	0	0
$(4, 3, 10)$	∞	303	284	84	$(6, 4, 4)$	341	300	105	105
$(4, 3, 11)$	∞	377	284	84	$(6, 5, 3)$	95	276	186	150
$(5, 2, 2)$	0.6	0.5	0	0	$(6, 5, 4)$	∞	8643	726	540

Table 5.1 – Timings and degrees for dense linear matrices

Regarding randomness of rational data (input or parameters), it is necessary here to make a clarification. To obtain a rational number, we generate a couple of integers with uniform distribution in a fixed interval $[-N, N]$ where the integer N is fixed *a priori*. Consequently, once N is given, the set of our possible inputs is finite, and one cannot guarantee that the instances lie inside the admissible Zariski open set defined by Propositions 3.6, 3.7 and 3.8. In any case, we implemented a subroutine to control that the assumptions on the algebraic sets built during LowRank (regularity of incidence varieties and finiteness of Lagrange systems) are satisfied. Since performing this control step can be done with Gröbner bases (*cf.* Section 1.2.2), this can be done efficiently using FGB.

We compare our timings (reported in column "SPECTRA") with the function PointsPerComponents, reported in column RAG, which is implemented in the real algebraic geometry library RAGLIB, *cf.* [135]. Finally, we give the degree of the output rational parametrization (column "deg") and the maximum degree of partial rational parametrizations (column "maxdeg"). We make the following remarks about Table 5.1. With deg $= 0$ and maxdeg $= 0$ we mean that the corresponding real determinantal variety $\mathcal{D}_r \cap \mathbb{R}^n$ is empty (hence the algorithm returns the empty list). The symbol ∞ means the computation did not stop after 4 days of computation.

1. We first observe that our algorithm allows to tackle examples that are out

of reach for RAGLIB, and we conclude that our implementation reflects the complexity gain.

2. The growth in terms of timings with respect to n seems to respect the correspondent growth in terms of degrees of output parametrizations. Remarkably, for $(m, n, r) = (4, 2, 9)$ the algorithm runs approximately 34 hours but is able to return a rational parametrization of degree 1074, or for $(m, n, r) = (4, 2, 7)$, after less than 3 hours, it returns polynomials of degree 532.

3. Accordingly to the related Multilinear Bézout Bounds computed in Section 3.4.1, the degrees of rational parametrizations stabilize when n grows, since when $n > m^2 - r^2$ and the input is generic, LowRank does not compute critical points at first calls. This fact is remarkable, since:

   - a natural geometric invariant associated to $\mathcal{D}_r$, its degree as a complex algebraic set, does not depend on the dimension n of the affine section (as stated by Proposition 3.5, *cf.* also [10, Ch. II, § 4]);
   - an algebraic invariant naturally associated to the output size (the degree) is constant in n, consistently with the above mentioned geometric invariant.

5.2.2 Examples

This final section reports numerical results of interesting examples and shows how our algorithm can be used to tackle different problems of the literature involving linear matrices.

Joint probability matrices

We come back to Example 3.2, page 74, concerning joint probability matrices. We recall that these are $m \times m$ full matrices of variables

$$A(x) = \begin{pmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & & \vdots \\ x_{m,1} & \cdots & x_{m,m} \end{pmatrix},$$

subject to the linear constraint

$$x_{1,1} = 1 - \sum_{(i,j) \neq (1,1)} x_{i,j}.$$

The entries $x_{i,j} = p(S = i, R = j)$ encode the joint probabilities of two discrete random variables. Eliminating the positivity constraint and allowing matrix A to have bounded rank, one considers the algebraic set

$$\mathcal{D}_r \cap \mathbb{R}^{m^2} = \left\{ x \in \mathbb{R}^{m^2} : \operatorname{rank} A(x) \leq r, \, \sum x_{i,j} = 1 \right\}.$$

Testing LowRank on such matrices yield the data of Table 5.2.

m	r	time (s)	deg	maxdeg
2	1	11.33	5	2
3	2	37	35	12
3	1	35	36	12
4	3	2530	244	78
4	2	12829	1311	283

Table 5.2 – Timings and degrees for joint probability matrices

The table shows the timings (in seconds) of our algorithm with input the linear matrix A with the above linear constraint on its entries, jointly with the degrees of output parametrizations. The column deg contains the degree of the output parametrization, while maxdeg the maximum degree with respect to the recursive steps.

Remarkably, our algorithm can tackle a 4×4 linear matrix with 15 variables, and with expected rank 3, and solve it in less than 40 minutes, with an output representation of degree 244. When two rank defects are considered, the algorithm returns a rational parametrization of degree 1311 after approximately 4 hours of computation.

Sylvester matrices

The second problem deals with Sylvester matrices of Example 3.3 on page 75:

$$A = \begin{pmatrix} f_0 & f_1 & f_2 & \dots & f_{d_1} & 0 & \dots & 0 \\ 0 & f_0 & f_1 & f_2 & \dots & f_{d_1} & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & f_0 & f_1 & f_2 & \dots & f_{d_1} \\ g_0 & g_1 & \dots & g_{d_2} & 0 & \dots & \dots & 0 \\ 0 & g_0 & g_1 & \dots & g_{d_2} & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & g_0 & g_1 & \dots & \dots & g_{d_2} \end{pmatrix}$$

Suppose that the coefficients of f and g admit uncertainties represented by $n-$variate affine polynomials with coefficients in $\mathbb{Q}$, that is $f_i, g_j \in \mathbb{Q}[x]$ with $x = (x_1, \ldots, x_n)$. This situation is quite common when dealing with data arising from practical experiments. By substituting each f_i and g_j in A with the corresponding affine polynomials, one gets a linear matrix of size $d_1 + d_2$ in n variables, with Sylvester structure for any evaluation of the vector x.

Hence, we tested LowRank with input a Sylvester linear matrix $A(x)$, whose rational coefficients have been chosen randomly with respect to a predetermined probability distribution (*cf.* discussion in Section 5.2.1). Table 5.3 shows the obtained results in terms of timings and of degrees of output polynomials. The title of the columns are the same as for the previous tables.

d_1	d_2	n	r	reg	time (s)	deg	maxdeg
2	1	4	1	2.2	16	1	1
2	1	5	1	6	23	2	1
2	1	6	1	12	27	2	1
2	1	7	1	15	35	2	1
2	1	4	2	✗	17	11	5
2	1	5	2	✗	23	12	5
2	1	6	2	✗	30	12	5
2	1	7	2	✗	36	12	5
2	2	3	1	0.1	14	0	0
2	2	4	1	0.3	18	0	0
2	2	5	1	0.3	23	0	0
2	2	6	1	∞	28	1	1
2	2	3	2	14	12	7	4
2	2	4	2	120	20	10	4
2	2	5	2	619	24	10	4
2	2	3	3	✗	15	21	9
2	2	4	3	✗	18	24	9
2	2	5	3	✗	24	24	9
3	2	2	1	0.8	9	0	0
3	2	3	1	1.44	12	0	0
3	2	4	1	2.82	17	0	0
3	2	3	2	3056	14	1	1
3	2	4	2	∞	0	1	1
3	2	5	2	∞	0	1	1
3	1	3	3	✗	12	15	7
3	1	4	3	✗	17	15	7
3	1	4	2	75	17	1	1
3	1	5	2	190	24	1	1

Table 5.3 – Timings and degrees for dense Sylvester linear matrices

In the column `reg` we write ✗ if regularity assumptions are not satisfied, otherwise the time needed to check these assumptions, or the symbol ∞ if the computation did not end after 48 hours. We remark here that, in the special cases $r = d_1 + d_2 - 1$, randomness of input rational numbers does not imply the sufficient conditions for the correctness of the algorithm.

However, this does not represent an obstacle to run the algorithm and to extract a finite subset of the determinantal variety associated to $A(x)$, that is to the discriminant variety for couples (f, g) of polynomials of degree bounded by (d_1, d_2), when $r = d_1+d_2-1$. A certificate establishing whether the output finite set meets or not every connected component (when the genericity assumptions are not satisfied) is hard to imagine, and so far there are no exact algorithm solving this task. We conjecture that for Sylvester matrices the generic assumptions that we have imposed for generic linear matrices (that is smoothness and equidimensionality of an incidence variety) are actually too restrictive.

We finally remark that, as in the special case of Hankel matrices described in Chapter 4, Section 4.2, one could design dedicated variants of algorithm LowRank for structured matrices as those of Sylvester type, exploiting the structure of input polynomials induced by the special structure of the matrix A.

Hurwitz matrices and stability of complex polynomials

We recall that in Example 3.4 we have associated to a univariate polynomial $f(z) = f_0 z^m + f_1 z^{m-1} + \ldots + f_{m-1} z + f_m$ of complex variable z, of degree at most m, its Hurwitz matrix

$$A = \begin{pmatrix} f_1 & f_3 & f_5 & \cdots & 0 \\ f_0 & f_2 & f_4 & & \vdots \\ 0 & f_1 & f_3 & & \vdots \\ \vdots & 0 & \vdots & \ddots & 0 \\ 0 & \vdots & & & f_m \end{pmatrix}.$$

We suppose, as in the previous example, that the coefficients of this polynomial are not known but can be expressed as an affine combination with given rational coefficients, and we are interested in the locus where A has fixed rank defects.

m	n	r	`reg`	`time (s)`	`deg`	`maxdeg`
2	2	1	✗	7.25	1	1
2	3	1	✗	12.4	3	2
2	4	1	✗	18.5	3	2
2	5	1	✗	22	3	2
2	6	1	✗	28	3	2
3	2	2	✗	7	6	3
3	3	2	✗	12	8	3
3	4	2	✗	17.5	9	3
3	5	2	✗	23	9	3
3	6	2	✗	28	9	3
3	2	1	0.53	7	1	1
3	3	1	1	13	1	1
3	4	1	✗	17	2	1
3	5	1	✗	24	2	1
3	6	1	✗	27	2	1
4	2	3	✗	7	11	7
4	3	3	✗	12	16	7
4	4	3	✗	18	18	7
4	5	3	✗	23	19	7
4	6	3	✗	28	19	7
4	2	2	2.2	7	3	3
4	3	2	✗	12	5	3
4	4	2	✗	19	7	3
4	5	2	✗	24	8	3
4	6	2	✗	30	8	3
4	2	1	0.2	7	0	0
4	3	1	0.2	13	0	0
4	4	1	0.2	17	0	0
4	5	1	1079	19	1	1
4	6	1	4348	29	1	1

Table 5.4 – Timings and degrees for dense Hurwitz linear matrices

Table 5.4 contains data about the output of LowRank with input such generic affine sections of low rank loci of Hurwitz matrices. We use the same name for the columns of Table 5.4 as for the tables of the previous sections.

We recall that when genericity assumptions are not satisfied, we cannot guarantee that the set defined by the output rational parametrization meets every connected component of $\mathcal{D}_r \cap \mathbb{R}^n$. As for Sylvester matrices, a perspective here is to design a dedicated variant of LowRank for linear matrices with Hurwitz structure, such that the assumptions on the input matrix $A(x)$ guaranteeing the correctness of the output results are satisfied in a non-empty Zariski open set of the set of linear Hurwitz matrices.

5.3 Numerical experiments for Chapter 4

5.3.1 Random spectrahedra

As for algorithm LowRank, we have tested its variant for symmetric matrices on generic instances.

(m, r, n)	SPECTRA	deg	maxdeg	(m, r, n)	SPECTRA	deg	maxdeg
$(3, 2, 2)$	8	9	6	$(5, 2, 3)$	0	0	0
$(3, 2, 3)$	11	13	6	$(5, 2, 4)$	1	0	0
$(3, 2, 4)$	13	13	6	$(5, 2, 5)$	1	0	0
$(3, 2, 5)$	14	13	6	$(5, 2, 6)$	5	35	35
$(3, 2, 6)$	15	13	6	$(5, 2, 7)$	25856	175	140
$(3, 2, 7)$	16	13	6	$(5, 3, 2)$	1	0	0
$(3, 2, 8)$	17	13	6	$(5, 3, 3)$	3	20	20
$(3, 2, 9)$	18	13	6	$(5, 3, 4)$	1592	110	90
$(4, 2, 2)$	0	0	0	$(5, 3, 5)$	16809	317	207
$(4, 2, 3)$	2	10	10	$(5, 4, 2)$	7	25	20
$(4, 2, 4)$	9	40	30	$(5, 4, 3)$	42	65	40
$(4, 2, 5)$	29	82	42	$(5, 4, 4)$	42	65	40
$(4, 2, 6)$	71	112	42	$(5, 4, 5)$	858	105	40
$(4, 2, 7)$	103	122	42	$(5, 4, 6)$	∞	-	-
$(4, 2, 8)$	106	122	42	$(5, 4, 7)$	∞	-	-
$(4, 2, 9)$	106	122	42	$(6, 3, 3)$	0	0	0
$(4, 3, 3)$	10	32	16	$(6, 3, 4)$	1	0	0
$(4, 3, 4)$	21	40	16	$(6, 3, 5)$	2	0	0
$(4, 3, 5)$	22	40	16	$(6, 3, 6)$	704	112	112
$(4, 3, 6)$	24	40	16	$(6, 3, 7)$	∞	-	-
$(4, 3, 7)$	26	40	16	$(6, 3, 8)$	∞	-	-
$(4, 3, 8)$	27	40	16	$(6, 4, 2)$	1	0	0
$(4, 3, 9)$	28	40	16	$(6, 4, 3)$	9	35	35
$(4, 3, 10)$	29	40	16	$(6, 4, 4)$	∞	-	-
$(4, 3, 11)$	30	40	16	$(6, 5, 3)$	591	116	80
$(5, 2, 2)$	0	0	0	$(6, 5, 4)$	∞	-	-

Table 5.5 – Timings and degrees for dense symmetric linear matrices

This gives an estimate of the time needed to perform the whole procedure SolveLMI, since LowRankSym is its main procedure. One can compare these values with the timings of PointsPerComponents in Table 5.1 since this function does not take reasonable advantage from the symmetry of the input pencil.

As for the previous tables, we report on the degree of the output parametrizations in column deg, and we put deg $= 0$ if the empty list is returned. We remark that the values of maximum degrees reported in column maxdeg coincide with the algebraic degree of the associated semidefinite program [117]. Indeed, one can compare data of Table 2 in [117] with data of Table 5.5. We recall that the main subroutine LowRankSym is recursive, and that at each step it eliminates one of the $x-$variables.

Let us consider, for example, the case $m = 4$ and $r = 2$ (this corresponds to case $n = 4$, $r = 2$ of Table 2 in [117]). Respectively for $n = 3, 4$ and 5 ($m = 3, 4, 5$ for Table 2 in [117]) we find the following degrees: 10, 30 and 42, which correspond to the algebraic degrees of generic semidefinite programs with the same parameters. In this thesis we did not provide an exact formula for the total degree of the output parametrization (but we obtained multilinear Bézout bounds). We conjecture that this degree is given as a weighted sum of algebraic degrees corresponding to partial generic semidefinite programs. Precisely:

Conjecture 5.1. *Let $\delta(k, m, r)$ be the algebraic degree of a generic semidefinite program with parameters k, m, r as computed in [117]. If property* H *holds (cf. page 127) then the sum of the degrees of the rational parametrizations computed during* SolveLMI *is given by the formula*

$$\sum_{r=1}^{r(A)} \binom{m}{r} \sum_{k=p_r-r(m-r)}^{\min(n,p_r+r(m-r))} \delta(k, m, r),$$

where $p_r = (m - r)(m + r + 1)/2$.

For completeness, we also report in Table 5.6, for each triple of parameters m, n, r, the degree of the parametrization containing the critical points of expected rank exactly r (in column deg) together with the bound $\theta(m, n, r)$ as computed in Proposition 4.16, page 134.

When the algorithm does not compute critical points (that is, when the Lagrange system generates the empty set) we put deg $= 0$. We remark here that both the degree and the bound are constant and equal to 0 if n is large enough. This explains why in column deg of Table 5.5 the sequence of degrees stabilizes in n once m, r are fixed. Similar behaviors appear, for example, when computing the Euclidean Distance degree (EDdegree) of determinantal varieties [37, 120]. In [120, Table 1], the authors report on the EDdegree of determinantal hypersurfaces generated by linear matrices $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$: for generic weights in the distance function, and when the codimension of the vector space generated by $A_1, \ldots, A_n$ is small (for us, when n is big, since matrices A_i are randomly generated, hence independent for $n \leq \binom{m+1}{2} = \dim \mathbb{S}_m(\mathbb{Q})$) the EDdegree is constant. Analogous comparisons can be done with results in [120, Example 4] and [120, Corollary 3.5].

(m,r,n)	deg	$\theta(m,n,r)$	(m,r,n)	deg	$\theta(m,n,r)$
$(3,2,2)$	6	9	$(4,3,9)$	0	0
$(3,2,3)$	4	16	$(5,2,5)$	0	0
$(3,2,4)$	0	15	$(5,2,6)$	35	924
$(3,2,5)$	0	6	$(5,2,7)$	140	10296
$(3,2,6)$	0	0	$(5,3,3)$	20	84
$(4,2,3)$	10	35	$(5,3,4)$	90	882
$(4,2,4)$	30	245	$(5,4,2)$	20	30
$(4,2,5)$	42	896	$(5,4,3)$	40	120
$(4,2,6)$	30	2100	$(5,4,4)$	40	325
$(4,2,7)$	10	3340	$(5,4,5)$	16	606
$(4,2,8)$	0	3619	$(6,3,3)$	0	0
$(4,2,9)$	0	2576	$(6,3,4)$	0	0
$(4,2,12)$	0	0	$(6,3,5)$	0	0
$(4,3,3)$	16	52	$(6,3,6)$	112	5005
$(4,3,4)$	8	95	$(6,4,2)$	0	0
$(4,3,7)$	0	20	$(6,4,3)$	35	165
$(4,3,8)$	0	0	$(6,5,3)$	80	230

Table 5.6 – Degrees and bounds for rational parametrizations

5.3.2 Toy examples

The pillow

Let

$$A(x) = \begin{pmatrix} 1 & x_1 & 0 & x_1 \\ x_1 & 1 & x_2 & 0 \\ 0 & x_2 & 1 & x_3 \\ x_1 & 0 & x_3 & 1 \end{pmatrix}$$

be the linear matrix of Example 4.6, page 117, and let $S = \{x \in \mathbb{R}^3 \mid A(x) \succeq 0\}$ its associated spectrahedron (*the pillow*).

We tested LowRank with input $(A, 2)$. We obtain a rational representation $q = (q_0, q_1, q_2, q_3, q_4)$ of degree 4 (with only real roots) parametrizing the four points. By isolating the 4 roots of q_4, one gets rational approximations of the singular points. We give one of these points for illustration:

$$\begin{aligned} x_1 &\in [-\tfrac{6521908912666475339}{9223372036854775808}, -\tfrac{13043817825332644843}{18446744073709551616}] \approx -\sqrt{2}/2 \\ x_2 &\in [\tfrac{26087635650665343561}{36893488147419103232}, \tfrac{6521908912666428733}{9223372036854775808}] \approx \sqrt{2}/2 \\ x_3 &\in [-\tfrac{6521908912666412349}{9223372036854775808}, -\tfrac{13043817825332731855}{18446744073709551616}] \approx -\sqrt{2}/2. \end{aligned}$$

We observe, finally, a typical output in terms of the degree of the rational parametrizations and the number of real solutions. Details are given in Table 5.7.

	partial deg	deg	real	$A \succeq 0$
$r = 1$	0 0 0	0	0	0
$r = 2$	4 0 0	4	4	4
$r = 3$	6 8 4	18	14	6

Table 5.7 – Degrees for the pillow

In column partial deg we report on the degrees of the rational parametrizations at the 3 recursive steps of LowRankSymRec. The sum of this values is reported in the column deg and it corresponds to the degree of the parametrization returned by the whole procedure (as before, deg $= 0$ means that no solutions are computed). In real we report the number of real solutions, and in $A \succeq 0$ the number of these solutions lying in the spectrahedron $\mathscr{S}$.

Remarkably, with $r = 3$ the algorithm allows to compute smooth points of the determinantal variety that lie over the spectrahedron: indeed, among the 6 points computed at first step, 4 are the degenerate points with rank 2 (underlying the special case of symmetric matrices, the same behavior does not hold for generic matrices) and the other two points are typically symmetric with respect to the origin, for example:

$$\begin{pmatrix} -0.5256010100 \\ 0.2250169442 \\ -0.8442037391 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0.5256010100 \\ -0.2250169442 \\ 0.8442037391 \end{pmatrix}.$$

The Cayley cubic

We consider the 3×3 linear matrix

$$A(x) = \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_3 \\ x_2 & x_3 & 1 \end{pmatrix}.$$

The convex region $\{x \in \mathbb{R}^3 \mid A(x) \succeq 0\}$ is the Cayley spectrahedron, and is pictured in Figure 4.1 in Example 4.5

We run LowRankSym with input (A, r) with $r = 2$ and $r = 1$ (the case $r = 0$ is trivial since $A(x)$ is always non-zero and hence $\mathcal{D}_0$ is empty). In both cases, the algorithm first verifies that the genericity assumptions are satisfied.

Let us first analyze the case $r = 2$. For any kernel configuration, LowRankSym runs 3 recursive steps. Its output is a rational parametrization of degree 14 with 12 real solutions and 2 complex solutions. We give below details of each recursive call. At the first step, a rational parametrization of degree 5 is returned, with the following 5 real solutions:

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 18.285118452 \\ 164.322822823 \\ 4.552268485 \end{pmatrix} \right\}.$$

The coordinates of the fifth point are approximated to 9 certified digits and such an approximation can be computed by isolating the coordinates in rational intervals:

$$\begin{aligned} x_1 &\in [\tfrac{21081306277346124211}{1152921504606846976}, \tfrac{21081306277346754459}{1152921504606846976}] \approx 18.285118452 \\ x_2 &\in [\tfrac{5920353629066611305}{36028797018963968}, \tfrac{23681414516266799197}{144115188075855872}] \approx 164.322822823 \\ x_3 &\in [\tfrac{10496816461511385723}{2305843009213693952}, \tfrac{2624204115377866059}{576460752303423488}] \approx 4.552268485, \end{aligned}$$

Remark that it also computes the 4 singular points of $\mathcal{D}_2$, where the rank of A is 1. At the second (resp. third) recursive call, it returns a rational parametrization of degree 6 (resp. of degree 3) with 4 (resp. 3) real solutions. Here it is worth to compare the values $4, 6, 3$ for the degrees of the output parametrization, with the Euclidean distance degree of the Cayley determinant, computed in [37, Ex. 5.7]. Indeed, the euclidean distance degree of the variety $\mathcal{D}_2 = \{x \in \mathbb{C}^3 : \det A(x) = 0\}$ is exactly $4 + 6 + 3 = 13$. Hence also in special examples we recover the behavior of Conjecture 5.1.

In the case $r = 1$, the algorithm returns a rational parametrization of degree 4 which encodes the 4 singular points of $\mathcal{D}_2 \cap \mathbb{R}^3$, that is $\mathcal{D}_1 \cap \mathbb{R}^3$.

We finally remark that the above results are typical, in the sense that the 4 singular points contained in $\mathcal{D}_1 \cap \mathbb{R}^3$ are always computed at the first recursion step, both in case $r = 2$ and $r = 1$. Conversely, the coordinates of the other real solutions depend on the choice of random parameters (while their number is constant). Moreover, all computations end after a few seconds.

5.3.3 Hankel spectrahedra

Random tests

We have described in Section 4.6 a variant of the subroutine LowRankSym when the input is a Hankel matrix. We give the corresponding results in Table 5.8. The input are randomly generated Hankel linear matrices, and as in the previous section we compare timings with RAGlib. All genericity assumptions are satisfied by the inputs of our tests. The symbol ∞ here means that the computation did not end after 1 day, while `matbig` means that the maximum limitation in terms of size of matrices occurring in the Gröbner bases computation has been reached.

(m,r,n)	RAG	SPECTRA	deg	maxdeg	(m,r,n)	RAG	SPECTRA	deg	maxdeg
$(3,2,3)$	0.6	10	21	12	$(5,3,3)$	2	2	20	20
$(3,2,4)$	2	13	33	12	$(5,3,4)$	202	18	110	90
$(3,2,5)$	7	20	39	12	$(5,3,5)$	∞	583	338	228
$(3,2,6)$	13	21	39	12	$(5,3,6)$	∞	6544	698	360
$(3,2,7)$	20	21	39	12	$(5,3,7)$	∞	28081	1058	360
$(3,2,8)$	53	21	39	12	$(5,3,8)$	∞	∞	-	-
$(4,2,3)$	2	2.5	10	10	$(5,4,2)$	1	5	25	20
$(4,2,4)$	43	6.5	40	30	$(5,4,3)$	48	30	105	80
$(4,2,5)$	56575	18	88	48	$(5,4,4)$	8713	885	325	220
$(4,2,6)$	∞	35	128	48	$(5,4,5)$	∞	15537	755	430
$(4,2,7)$	∞	46	143	48	$(5,4,6)$	∞	77962	1335	580
$(4,2,8)$	∞	74	143	48	$(6,2,7)$	∞	6	36	36
$(4,3,2)$	0.3	8	16	12	$(6,2,8)$	∞	`matbig`	-	-
$(4,3,3)$	3	11	36	52	$(6,3,5)$	∞	10	56	56
$(4,3,4)$	54	31	120	68	$(6,3,6)$	∞	809	336	280
$(4,3,5)$	341	112	204	84	$(6,3,7)$	∞	49684	1032	696
$(4,3,6)$	480	215	264	84	$(6,3,8)$	∞	`matbig`	-	-
$(4,3,7)$	528	324	264	84	$(6,4,3)$	3	5	35	35
$(4,3,8)$	2638	375	264	84	$(6,4,4)$	∞	269	245	210
$(5,2,5)$	25	4	21	21	$(6,4,5)$	∞	30660	973	728
$(5,2,6)$	31176	21	91	70	$(6,4,6)$	∞	∞	-	-
$(5,2,7)$	∞	135	199	108	$(6,5,2)$	1	9	36	30
$(5,2,8)$	∞	642	283	108	$(6,5,3)$	915	356	186	150
$(5,2,9)$	∞	950	311	108	$(6,5,4)$	∞	20310	726	540
$(5,2,10)$	∞	1106	311	108	$(6,5,5)$	∞	∞	-	-

Table 5.8 – Timings and degrees for dense Hankel linear matrices

The even moment curve

The even moment curve (*cf.* [141, Sec. 5.2]) is the intersection of the rank one locus and of the associated spectrahedron of the Hankel pencil

$$A(x) = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_n \\ x_1 & x_2 & x_3 & \cdots & x_{n+1} \\ x_2 & x_3 & x_4 & \cdots & x_{n+2} \\ \vdots & \vdots & \vdots & & \vdots \\ x_n & x_{n+1} & x_{n+2} & \cdots & x_{2n} \end{pmatrix},$$

subject to the affine constraint

$$1 - \sum_{j=0}^{n} \binom{n}{j} x_{2j} = 0.$$

The convex hull of the even moment curve is exactly the spectrahedron $\mathscr{S} = \{x \in \mathbb{R}^{2n} : A(x) \succeq 0\}$. The set $\mathscr{S}$ is an orbitope, called the universal Carathéodory orbitope. We relax the rank one constraint, by looking for solutions of $A(x) \succeq 0$ with different rank deficiencies.

We obtain Table 5.9 applying the variant for Hankel matrices of SolveLMI. In this table the symbol ∞ means that the computations did not end after 1 day of computation.

n	r	`reg`	`time (s)`	`deg`	`maxdeg`	$A \succeq 0$
1	1	0.1	7	4	2	≥ 2
2	2	0.1	17	18	7	≥ 2
2	1	0.1	17	10	4	≥ 2
3	3	0.4	63	93	33	≥ 2
3	2	0.25	32	66	27	≥ 4
3	1	0.14	29	16	6	≥ 4
4	2	2	244	146	61	≥ 4
4	1	0.2	40	32	8	≥ 2
5	2	9.4	7790	258	113	≥ 6
5	1	0.3	65	28	10	≥ 4
6	1	0.7	85	34	12	≥ 4

Table 5.9 – Timings and degrees for the even moment curve spectrahedron

In the column $A \succeq 0$ we report on the number of real solutions lying on the spectrahedron (we observe a typical lower bound). We remark that the algorithm has a better behavior in terms of timings when the rank defect is high. For example, for $n = 5$, it is able to handle a 6×6 linear Hankel matrix with 10 variables, and 4 and 5 rank defects, respectively, or for $n = 6$, a 7×7 matrix with 11 variables and 6 rank defects. The same examples cannot be handled by general algorithms and also by algorithm LowRankSym. We add that for cases $(n, r) = (4, 4), (4, 3), (5, 5), (5, 4), (5, 3)$ we succeded in checking the regularity assumptions but the algorithm did not stop after 24 hours of computations.

5.3.4 A quartic spectrahedron

Consider the 4×4 symmetric homogeneous pencil

$$A(x) = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{1,2} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{1,3} & a_{2,3} & a_{3,3} & a_{3,4} \\ a_{1,4} & a_{2,4} & a_{3,4} & a_{4,4} \end{pmatrix},$$

with linear entries

$$\begin{aligned}
a_{1,1} &= 45x_0 + 40x_1 + 65x_2 + 72x_3 & a_{1,2} &= 33x_0 + 50x_1 + 7x_2 - 60x_3 \\
a_{1,3} &= -84x_0 + 12x_1 + 3x_2 - 54x_3 & a_{1,4} &= -21x_0 + 16x_1 + 54x_2 - 18x_3 \\
a_{2,2} &= 82x_0 + 85x_1 + 2x_2 + 82x_3 & a_{2,3} &= -99x_0 + 12x_1 + 6x_2 - 25x_3 \\
a_{2,4} &= -46x_0 + 41x_1 + 18x_2 + 51x_3 & a_{3,3} &= 181x_0 + 4x_1 + 26x_2 + 53x_3 \\
a_{3,4} &= 59x_0 + 2x_1 + 58x_2 - 9x_3 & a_{4,4} &= 26x_0 + 26x_1 + 164x_2 + 45x_3.
\end{aligned}$$

The spectrahedron associated to $A(x)$ (actually, one of its 3−dimensional affine sections) is pictured on page 4 of [130]. The loci of rank defects of homogeneous pencils define real algebraic sets that are affine cones containing the origin, hence connected sets. Consequently the related real root finding problems or the emptiness problem for the associated spectrahedron are easily solved by returning the null vector. In order to break the homogeneity of the problem we cut the linear pencil $A(x)$ with an affine constraint

$$a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + b = 0.$$

When coefficients a_i and b are chosen randomly with respect to a predetermined distribution, this gives rise to a generic affine section of the aforementioned cone. Suppose from now on that such coefficients are randomly generated.

We tested SolveLMI with input the obtained affine matrix. First, we note that all regularity assumptions on the incidence varieties of such example are satisfied. Let us denote with $\mathcal{D}_r$, $r = 0, 1, 2, 3$, the low rank loci associated to the affine matrix. These algebraic sets have the expected dimension, that is $\mathcal{D}_0 = \mathcal{D}_1$ are empty and $\dim \mathcal{D}_2 = 0$ and $\mathcal{D}_3$ is the hypersurface defined by $\det A = 0$ and has dimension 2. Moreover, the set $\mathcal{D}_2$ has degree 10, and it corresponds to the 10 real singular points of the determinantal hypersurface. Algorithm SolveLMI provides the following information:

- SolveLinear with input $A = 0$ returns the empty list after a few seconds.
- LowRankSym, with input $(A, 1)$, outputs the empty list after a few seconds.
- With input $(A, 2)$, the output of the first recursion is a rational parametrization of degree 10, encoding 10 real distinct points. Depending on the affine section that we apply in order to de-homogenize the pencil $A(x)$, the hyperplane $a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + b = 0$ either cuts the locus $\mathscr{S} = \{(x_0, x_1, x_2, x_3) \in \mathbb{R}^4 : A(x) \succeq 0\}$ or its opposite $-\mathscr{S} = \{(x_0, x_1, x_2, x_3) \in \mathbb{R}^4 : A(x) \preceq 0\}$. In any case we verify that at all solutions of the rational parametrizations, the dehomogenized pencil is semidefinite with constant sign.
- With input $(A, 3)$, the degree of the output parametrization is 26 at the first recursion step, 11 at the second and 3 at the third. The number of real solutions varies in a set of typical values.

5.3.5 Degenerate example: the Scheiderer spectrahedron

We consider the Scheiderer spectrahedron, *cf.* Example 4.7. As explained, this is the convex set defined by a 6−variate 6×6 symmetric pencil representing the Gram matrix of a homogeneous ternary form $f \in \mathbb{Q}[u_1, u_2, u_3]$ of degree 4 (this matrix is displayed on page 118). This polynomial is nonnegative over $\mathbb{R}^3$ and it is, by Hilbert's theorem, a sum of at most three squares. The spectrahedron $\mathscr{S}$ parametrizes all the possible sum-of-squares decompositions of f. Indeed, if

$$v = (u_1^2, u_1u_2, u_2^2, u_1u_3, u_2u_3, u_3^2)$$

is the row vector whose entries span the space of homogeneous polynomials of degree 2 in 3 variables, then f is a sum of squares if and only if there exists x^* such that $f = vA(x^*)v'$, with $A(x^*) \succeq 0$.

Scheiderer proved that f does not admit a decomposition as a sum of squares over $\mathbb{Q}[u_1, u_2, u_3]$. This result implies that the spectrahedron $\mathscr{S} = \{x \in \mathbb{R}^n : A(x) \succeq 0\}$ is not empty but does not contain rational points. Guo *et al.* designed in [60] an exact algorithm for deciding the emptiness of $\mathscr{S} \cap \mathbb{Q}^6$, providing a computer algebra (probabilistic) verification of Scheiderer's result: when running the algorithm in [60] with input the Gram matrix of Scheiderer's spectrahedron $\mathscr{S}$, the output states that $\mathscr{S} \cap \mathbb{Q}^6$ is empty.

We have tested SolveLMI on Scheiderer's spectrahedron. Once checked that the associated incidence varieties are smooth and equidimensional, the recursive subroutines first easily check that $\mathscr{S}$ does not contain any point x with $\operatorname{rank} A(x) = 0$ and 1 (and precisely, deciding that $\mathcal{D}_0 \cap \mathbb{R}^6 = \mathcal{D}_1 \cap \mathbb{R}^6 = \emptyset$). Further, for $r = 2$, the algorithm returns the following rational parametrization of $\mathcal{D}_2 \cap \mathbb{R}^6$:

$$
\begin{array}{ll}
x_1 = \frac{3+16t}{-8+24t^2} & x_2 = \frac{8-24t^2}{-8+24t^2} \\
x_3 = \frac{8+6t+8t^2}{-8+24t^2} & x_4 = \frac{16+6t-16t^2}{-8+24t^2} \\
x_5 = \frac{-3-16t}{-8+24t^2} & x_1 = \frac{3+16t}{-8+24t^2}
\end{array}
$$

where t has to be chosen among the solutions of the univariate equation

$$8t^3 - 8t - 1 = 0.$$

The set $\mathcal{D}_2$ is, indeed, of dimension 0 and degree 3. We also remark that the rational parametrization makes explicit and certifies the following relations between the coordinates of the three points of $\mathcal{D}_2$:

$$
\begin{aligned}
x_1 - x_6 &= 0 \\
x_5 + x_6 &= 0 \\
x_3 - x_4 - 1 &= 0 \\
x_2 + 1 &= 0.
\end{aligned}
$$

The previous equations define the 2-dimensional affine plane spanned by the three points.

By applying CheckLMI one gets that two of the three points lie on $\mathscr{S}$, that is those with the following floating point approximation up to 9 certified digits:

$$
\begin{pmatrix} -0.930402926 \\ -1.000000000 \\ 0.731299211 \\ -0.268700788 \\ 0.930402926 \\ -0.930402926 \end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix} -0.127050844 \\ -1.000000000 \\ -0.967716166 \\ -1.967716166 \\ 0.127050844 \\ -0.127050844 \end{pmatrix}
$$

and a third point:

$$\begin{pmatrix} 1.057453771 \\ -1.000000000 \\ 1.236416954 \\ 0.2364169545 \\ -1.057453771 \\ 1.057453771 \end{pmatrix}$$

at which the matrix is indefinite, of rank 2.

We conclude this example with the following considerations:

- SolveLMI allows to certify that $\mathscr{S}$ is not empty, that the minimal rank attained by $A(x)$ on $\mathscr{S}$ is 2. This rank is attained at the two points whose coordinates have been given above.

- Since $\mathscr{S}$ does not contains rational points, we can easily deduce that the interior of $\mathscr{S}$ is empty and that $\mathscr{S}$ is not full-dimensional, that is, its affine span is contained in a proper hyperplane of $\mathbb{R}^6$ (*cf.* [85, Th. 1.2]). This is a degenerate behavior since typically (for generic matrices $A_0, A_1, \ldots, A_n$) the spectrahedron $\mathscr{S}$ is either empty or full-dimensional.

- One can easily verify that $\mathcal{D}_3 \cap \mathbb{R}^6$ is also zero-dimensional, of degree 57. It contains 4 real points, 3 of which are the elements of $\mathcal{D}_2 \cap \mathbb{R}^6$, and, at the fourth one, A has rank 3 and is not positive semidefinite. This shows that this example is particularly degenerate also with respect to the dimension of the rank stratifications of the matrix A. Indeed, the expected dimension of $\mathcal{D}_2$ is -1, while $\mathcal{D}_3$ has the expected dimension, that is 0. The fact that $\mathcal{D}_3$ does not intersect $\mathscr{S}$ is consistent with [142, Ex. 2.8].

- As previously observed, $\mathscr{S}$ parametrizes the decompositions of the polynomial f as a sum of squares. Further, SolveLMI is able to parametrize all the connected components of $\mathcal{D}_2 \cap \mathbb{R}^6$ that intersect (and hence that are contained in) $\mathscr{S}$. Hence, SolveLMI is able to establish, with a computer algebra proof:

  - that f is not a pure square, since $\mathcal{D}_1 \cap \mathscr{S}$ is empty;
  - that f is a sum of 2 squares, and that there are two such decompositions (already established via real algebra by Scheiderer, *cf.* [142, Ex. 2.8]), since $\mathcal{D}_2 \cap \mathscr{S}$ consists of 2 points;

  Moreover, one can provide a parametrization of the 2 minimal-rank decompositions. Writing down such expressions can be done by factorizing the matrix $A(x(t)) = VV'$, where t is chosen among the two solutions of $8t^3 - 8t - 1 = 0$ at which $A(x(t)) \succeq 0$.

Bibliography

[1] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. SIAM J. on Opt. 5(1):13–51, 1995.

[2] R.G. Bartle, D.R. Sherbert. Introduction to real analysis. John Wiley & Sons, Third Edition. New York 1992.

[3] J. Bokowski, B. Sturmfels. Computational synthetic geometry. Lecture Notes in Mathematics, N.1355, Springer, Heidelberg, 1989.

[4] M. Abril Bucero, B. Mourrain. Exact relaxation for polynomial optimization on semi-algebraic sets. https://hal.inria.fr/hal-00846977, 2014.

[5] A. Alzati, E. Ballico, G. Ottaviani. The theorem of Mather on generic projections for singular varieties. Geom. Dedicata 85, 113–117 (2001).

[6] A. Alzati, G. Ottaviani. The theorem of Mather on generic projections in the setting of algebraic geometry. Manuscr. Math. 74 (1992), 391–412.

[7] C.L. Alonso, J. Puente, J.L. Montaña. Straight line programs: a new linear genetic programming approach. IEEE International Conference on Tools with Artificial Intelligence Vol.2:517–524, 2008.

[8] G. Blekherman. Dimensional differences between nonnegative polynomials and sums of squares. Preprint arXiv:0907.1339, 2009.

[9] M.F. Anjos and J.B. Lasserre editors. Handbook of semidefinite, conic and polynomial optimization. International Series in Operational Research and Management Science. Volume 166, 2012.

[10] E. Arbarello, J. Harris, M. Cornalba, P. Griffiths. Geometry of algebraic curves. Volume I. Grundlehren der Mathematischen Wissenschaften, vol. 267, Springer-Verlag, New York, 1985.

[11] B. Bank, M. Giusti, J. Heintz, M. Safey El Din, É. Schost. On the geometry of polar varieties. Applicable Algebra in Engineering, Communication and Computing. 21(1): 33–83, 2010.

[12] B. Bank, M. Giusti, J. Heintz, G.-M. Mbakop. Polar varieties and efficient real equation solving: the hypersurface case. Journal of Complexity, 13(1):5–27, 1997.

[13] B. Bank, M. Giusti, J. Heintz, G.-M. Mbakop. Polar varieties and efficient real elimination. Mathematische Zeitschrift, 238(1):115–144, 2001.

[14] B. Bank, M. Giusti, J. Heintz, L.-M. Pardo. Generalized polar varieties and efficient real elimination procedure. Kybernetika, 40(5):519–550, 2004.

[15] B. Bank, M. Giusti, J. Heintz, L.-M. Pardo. Generalized polar varieties: geometry and algorithms. Journal of Complexity, 21(4):377–412, 2005.

[16] B. Bank, M. Giusti, J. Heintz, L. Pardo. Bipolar varieties and real solving of a singular polynomial equation. Jaen Journal of Approximation, 2(1):65–77, 2010.

[17] A. Bhardwaj, P. Rostalski, R. Sanyal. Deciding polyhedrality of spectrahedra. SIAM J. on Opt., 2015.

[18] S. Basu, R. Pollack, and M.-F. Roy. On the number of cells defined by a family of polynomials on a variety Mathematika, 43:120–126, 1996.

[19] S. Basu, R. Pollack, and M.-F. Roy. On the Betti numbers of sign conditions Proceedings of the American Mathematical Society, 133(4):965–974, 2004.

[20] S. Basu, R. Pollack, and M.-F. Roy. Algorithms in real algebraic geometry, volume 10 of Algorithms and Computation in Mathematics. 2nd edition. Springer-Verlag, Berlin, 2006.

[21] A. Beauville. Determinantal hypersurfaces. Mich. Math. J. 48:39–64 (2000).

[22] A. Ben-Tal, A. Nemirovski. Lectures on modern convex optimization: analysis, algorithms, engineering applications. MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2001.

[23] G. Blekherman, P. A. Parrilo, R. R. Thomas (Editors). Semidefinite optimization and convex algebraic geometry. SIAM, Philadelphia, 2013.

[24] J. Bochnak, M. Coste, and M.-F. Roy. Real algebraic geometry. Springer-Verlag, Berlin Heidelberg, 1998.

[25] S.P. Boyd, L. El Ghaoui, E. Feron, V. Balakrishnan. Linear matrix inequalities in system and control theory. Studies in Applied Mathematics, Vol 15. SIAM, Philadelphia, 1994.

[26] A. Bostan, F. Chyzak, M. Giusti, R. Lebreton, G. Lecerf, B. Salvy, É. Schost. Algorithmes efficaces en calcul formel. Lecture notes of MPRI, 2014.

[27] L. Vandenberghe, S. Boyd. Semidefinite programming. SIAM Review, 38(1):49–95, 1996.

[28] W. Bruns, U. Vetter. Determinantal rings, Springer Verlag, Berlin Heidelberg, 1988.

[29] B. Buchberger. A theoretical basis for the reduction of polynomials to canonical forms. SIGSAM Bull., 10(3):19–29, 1976.

[30] P. Bürgisser, M. Clausen, M.A. Shokrollahi. Algebraic complexity theory. Grundlehren der mathematischen Wissenschaften 315, Springer, 1997.

[31] M.D. Choi, T.Y. Lam, B. Reznick. Sums of squares of real polynomials. Proceedings of Symposia in Pure mathematics 58:103–126, 1995.

[32] M. Claeys. Mesures d'occupation et relaxations semi-définies pour la commande optimale. PhD thesis, LAAS CNRS, Toulouse, October 2013.

[33] G. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decompostion. Automata Theory and Formal Languages, pages 134–183. Springer, 1975.

[34] A. Conca. Straightening law and powers of determinantal ideals of Hankel matrices. Advances in Mathematics, 138(2):263–292, 1998.

[35] D. A. Cox, J. Little, D. O'Shea. Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra. Springer, 3rd edition, New York, 2007.

[36] M. Demazure. Bifurcations et catastrophes. Ellipses, Paris, 1989.

[37] J. Draisma, E. Horobet, G. Ottaviani, B. Sturmfels, R.R. Thomas. The Euclidean distance degree of an algebraic variety. Foundations of Computational Mathematics, DOI:10.1007/s10208-014-9240-x, 2015.

[38] C. Durvye. Algorithmes pour la décomposition primaire des idéaus polynomiaux de dimension nulle donné en évaluation. PhD Thesis, Université de Versailles, 2008.

[39] D. Eisenbud. Commutative algebra with a view toward algebraic geometry. Springer-Verlag, New York, 1995.

[40] D. Eisenbud. Linear sections of determinantal varieties. American Journal of Mathematics, 110(3):541–575, 1988.

[41] J.-C. Faugère. A new efficient algorithm for computing Gröbner bases (F4). Journal of Pure and Applied Algebra, 139(1–3):61–88, 1999.

[42] J.-C. Faugère. A new efficient algorithm for computing Gröbner bases without reductions to zero (F5). Proceedings of ISSAC 2002, Lille, France, 2002.

[43] J.-C. Faugère. FGb: a library for computing Gröbner bases. In Mathematical Software–ICMS 2010, pages 84–87, Springer, 2010.

[44] J.-C. Faugère, M. Safey El Din, P.-J. Spaenlehauer. On the complexity of the generalized MinRank problem. Journal of Symbolic Computation. 55:30–58, 2013.

[45] J.-C. Faugère, M. Safey El Din, P.-J. Spaenlehauer. Computing loci of rank defects of linear matrices using Gröbner bases and applications to cryptology. Proceedings of ISSAC 2010, Munich, Germany, 2010.

[46] J.-C. Faugère, F. Levy-dit Vehel, L. Perret. Cryptanalysis of minrank. In David Wagner, editor, Advances in Cryptology CRYPTO 2008, volume 5157 of Lecture Notes in Computer Science, pages 280–296, Berlin, Heidelberg, August 2008. Springer-Verlag.

[47] J.-C. Faugère, P. Gaudry, L. Huot, G. Renault. Polynomial systems solving by fast linear algebra. Preprint arXiv:1304.6039, 2013.

[48] J.-C. Faugère, P. Gianni, D. Lazard, T. Mora. Efficient computation of zero-dimensional Gröbner bases by change of ordering. Journal of Symbolic Computation, 16(4):329–344, 1993.

[49] J.-C. Faugère, C. Mou. Sparse FGLM algorithms. Preprint arXiv:1304.1238, 2013.

[50] J.-C. Faugère, C. Mou. Fast algorithm for change of ordering of zero-dimensional Gröbner bases with sparse multiplication matrices. In Proceedings of ISSAC 2011, San Jose, USA, 2011.

[51] I.M. Gel'fand, M.M. Kapranov, A.V. Zelevinsky. Discriminants, resultants and multidimensional determinants. Mathematics: Theory & Applications, Birkhäuser Boston Inc., Boston, MA, 1994.

[52] M. Giusti, K. Hägele, G. Lecerf, J. Marchand, B. Salvy. The projective Noether Maple package: computing the dimension of a projective variety. Journal of Symbolic Computation. 30(3):291–307, 2000.

[53] M. Giusti, G. Lecerf, B. Salvy. A Gröbner-free alternative for polynomial system solving. Journal of Complexity, 17(1):154–211, 2001.

[54] H.-C. G. v. Bothmer, K. Ranestad. A general formula for the algebraic degree in semidefinite programming Bulletin of LMS, 41:193–197, 2009.

[55] D. Grigoriev, N. Vorobjov. Solving systems of polynomial inequalities in subexponential time. Journal of Symbolic Computation, 5:37–64, 1988.

[56] D. Grigoriev, D. Pasechnik. Polynomial time computing over quadratic maps I. Sampling in real algebraic sets. Computational complexity, 14:20–52, 2005.

[57] A. Greuet. Polynomial optimization and polar varieties: theory, algorithms and implementations. PhD Thesis. University of Versailles - University Pierre et Marie Curie, Paris. 2013.

[58] A. Greuet, M. Safey El Din. Probabilistic Algorithm for the Global Optimization of a Polynomial over a Real Algebraic Set. SIAM J. on Opt., 24(3):1313–1343, 2014.

[59] M. Grötschel, L. Lovász, A. Schrijver. Geometric algorithms and combinatorial optimization. Springer-Verlag, Berlin, 1988.

[60] Q. Guo, M. Safey El Din, L. Zhi. Computing rational solutions of linear matrix inequalities. Proceedings of ISSAC 2013, Boston, USA, 2013.

[61] F. Guo, E. Kaltofen, L. Zhi. Certificates of impossibility of Hilbert-Artin representations of a given degree for definite polynomials and functions. Proceedings of ISSAC 2012, Grenoble, France, 195–202, 2012.

[62] J. Harris. Algebraic geometry. A first course. Springer Verlag, New York, 1992.

[63] R. Hartshorne. Algebraic geometry. Springer Science and Business Media, 1977.

[64] G. Heinig and K. Rost. Algebraic methods for Toeplitz-like matrices and operators. Springer, 1984.

[65] J. Heintz, M.-F. Roy, P. Solerno. On the theoretical and practical complexity of the existential theory of the reals. The Computer Journal, vol 36, 5 427–431 (1993).

[66] J. Heintz, M.-F. Roy, P. Solerno. Description of the connected components of a semi-algebraic set in single exponential time. Discrete and Computational Geometry 11:121–140 (1994).

[67] J. Heintz, M.-F. Roy, P. Solerno. On the complexity of semi-algebraic sets. Proc. IFIP 89, San Francisco. North-Holland 293–298 (1989).

[68] D. Henrion, S. Naldi, M. Safey El Din. Real root finding for determinants of linear matrices. Journal of Symbolic Computation, DOI:10.1016/j.jsc.2015.06.010, 2015.

[69] D. Henrion, S. Naldi, M. Safey El Din. Real root finding for rank defects in linear Hankel matrices. Proceedings of ISSAC 2015, Bath UK, 2015.

[70] D. Henrion, S. Naldi, M. Safey El Din. Real root finding for low rank linear matrices. To be registered as a LAAS-CNRS Research Report. Submitted to Journal of Complexity, 2015.

[71] D. Henrion, S. Naldi, M. Safey El Din. Exact algorithms for linear matrix inequalities. Submitted to SIAM J. on Opt., August 2015, https://hal.archives-ouvertes.fr/hal-01184320.

[72] B. Helton, J. Nie. Sufficient and necessary conditions for semidefinite representability of convex hulls and sets. SIAM J. on Opt. 20, 759–791 (2009).

[73] D. Henrion, J.B. Lasserre, J. Löfberg. GloptiPoly 3: moments, optimization and semidefinite programming. Optimization Methods and Software 24(4-5):761–779, 2009.

[74] B. Huber, B. Sturmfels. A polyhedral method for solving sparse polynomial systems. Math. Comput. 64(212):1541–1555 (1995).

[75] J.B. Lasserre, D. Henrion, C. Prieur, E. Trélat. Nonlinear optimal control via occupation measures and LMI relaxations. SIAM J. Control Opt. 47(4):1643–1666, 2008.

[76] D. Henrion, J.B. Lasserre, C. Savorgnan. Approximate volume and integration for basic semialgebraic sets. SIAM Review 51(4):722–743, 2009.

[77] D. Henrion. Semidefinite geometry of the numerical range. Electronic Journal of Linear Algebra, 20:322–332, 2010.

[78] D. Henrion. Optimization on linear matrix inequalities for polynomial systems control. Lecture notes of the International Summer School of Automatic Control, Grenoble, France, September 2014. Journal du CIRM 2013.

[79] D. Henrion, J. Malick. Projection methods for conic feasibility problems, applications to polynomial sum-of-squares decompositions. Optimization Methods and Software 26(1):23–46, 2011.

[80] V. Magron, D. Henrion, J.B. Lasserre. Semidefinite approximations of projections and polynomial images of semialgebraic sets To be registered as a LAAS-CNRS Research Report, October 2014. Submitted to SIAM F. on Opt. hal-01075343.

[81] D. Hilbert. Über die Theorie der algebraischen Formen. Math. Annalen 36 (4): 473–534 (1890).

[82] D. Hilbert. Über die vollen Invariantensysteme. Math. Ann. 42:13–373. (1893).

[83] D. Hilbert. Über die Dartstellung definiter Formen als Summe von Formenquadraten. Math. Ann. 32, 342–350. (1888).

[84] R. Hildebrand. Spectrahedral cones generated by rank 1 matrices. Journal of Global Optimization, DOI:10.1007/s10898-015-0313-4.

[85] C. Hillar. Sums of polynomial squares over totally real fields are rational sums of squares. Proc. American Math. Society, 137: 921–930, 2009.

[86] J. Huh, B. Sturmfels. Likelihood geometry. In Aldo Conca et al. editors. Combinatorial Algebraic Geometry, Lecture Notes in Mathematics 2108, Springer, 63–117, 2014.

[87] Z. Jelonek. Testing sets for properness of polynomial mappings. Mathematische Annalen, 315(1):1–35, 1999.

[88] G. Jeronimo, G. Matera, P. Solernó, and A. Waissbein. Deformation techniques for sparse systems. Foundations of Computational Mathematics, 9(1):1–50, 2009.

[89] G. Jeronimo, D. Perrucci, J. Sabia. On sign conditions over real multivariate polynomials. Discrete and Computational Geometry, 44:195–222, 2010.

[90] D. Jibetean, E. de Klerk. Global optimization of rational functions: a semidefinite programming approach. Mathematical Programming, 106:93–109, 2006.

[91] L. Khachiyan and L. Porkolab. On the complexity of semidefinite programs. J. Global Optim., 10:351–365, 1997.

[92] H. Khalil. Nonlinear systems. 3rd Edition. Prentice Hall, Upper Saddle River, NJ USA, 2002.

[93] J. Kollar. Sharp effective Nullstellensatz. Journal of the American Mathematical Society, 1, No.(4):963–975, 1988.

[94] I. Klep, M. Schweighofer. An exact duality theory for semidefinite programming based on sums of squares. Mathematics of Operations Research 38:569–590, 2013.

[95] I. Klep, M. Schweighofer. Pure states, positive matrix polynomials and sums of hermitian squares. Indiana University Mathematics Journal 59(3):857–874, 2010.

[96] I. Klep, M. Schweighofer. Infeasibility certificates for linear matrix inequalities. Oberwolfach Preprints (OWP), 2011, vol. 28.

[97] F. Le Gall. Powers of tensors and fast matrix multiplication. Proceedings of ISSAC 2014, Kobe JP, 296–303, 2014.

[98] S. Lang. Algebra. Springer-Verlag New York Inc., 3rd edition, 2002.

[99] J.B. Lasserre. Moments, positive polynomials and their applications. Imperial College Press, London, UK, 2010.

[100] J.B. Lasserre. Global optimization with polynomials and the problem of moments. SIAM J. on Opt., 11(3):796–817, 2001.

[101] J.B. Lasserre. A semidefinite programming approach to the generalized problem of moments. Mathematical Programming, 112:65–92, 2008.

[102] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In M. Putinar and S. Sullivant (eds.). Emerging Applications of Algebraic Geometry, Vol. 149 of IMA Volumes in Mathematics and its Applications, Springer, pages 157–270, 2009

[103] C. Le Guernic, F. Rouillier, M. Safey El Din. On the practical computation of one point in each connected component of a semi-algebraic set defined by a polynomial system of equations and non-strict inequalities. In Proceedings of EACA 04 conference, L.Gonzalez-Vega and T. Recio Eds, 2004.

[104] A. Logar. A computational proof of the Noether normalization lemma. Applied Algebra, Algebraic Algorithms and Error-Correcting Codes, pages 259–273, Lecture Notes in Computer Science, 357, Springer, Berlin, 1989.

[105] P. Gahinet, A. Nemirovsky. LMI Lab: A package for manipulating and solving LMIs. INRIA, Le Chesnay, 1993.

[106] H. Lombardi, D. Perrucci, M.-F. Roy. An elementary recursive bound for the effective Positivestellensatz and Hilbert 17th problem. Preprint arXiv 1404.2338, 2013.

[107] D.R. Grayson, M.E. Stillman. MACAULAY2, a software system for research in algebraic geometry: http://www.math.uiuc.edu/Macaulay2/.

[108] MAGMA Computational Algebra System. University of Sydney. http://magma.maths.usyd.edu.au/magma/

[109] MAPLE. Computer Algebra System: www.maplesoft.com/products/maple/. Waterloo Maple (Maplesoft).

[110] J.N. Mather. Generic projections. Ann. of Math 98:226–245, 1973.

[111] MATLAB. Version 2015a. MathWorks, March 5, 2015.

[112] Y. Ma, L. Zhi. Computing real solutions of polynomial systems via low-rank moment matrix completion. Proceedings of ISSAC 2012, Grenoble, France, 249–256, 2012.

[113] T.M. Developement Team. The mosek optimization software. 2013.

[114] S. Naldi. Nonnegative polynomials and their Carathéodory number. Discrete and Computational Geometry 51(3):559–568 (2014).

[115] Y. Nesterov and A. Nemirovsky. Interior-point polynomial algorithms in convex programming. Studies in Applied Mathematics 13. SIAM, Philadelphia, 1994.

[116] Y. Nesterov. Squared functional systems and optimization problems. In H. Frenk, K. Roos, T. Terlaky (Editors). High performance optimization. Kluwer Academic Publishers, Dordrecht, Pages 405–440, 2000.

[117] J. Nie, K. Ranestad, B. Sturmfels. The algebraic degree of semidefinite programming. Mathematical Programming 122:379–405, 2010.

[118] J. Nie. Optimality conditions and finite convergence of Lasserre's hierarchy. Mathematical Programming, Ser. A, Vol 146, No. 1-2, pp. 97–121, 2014.

[119] J. Nie, M. Schweighofer. On the complexity of Putinar Positivstellensatz. Journal of Complexity 23, No. 1, 135–150 (2007).

[120] G. Ottaviani, P.-J. Spaenlehauer, B. Sturmfels. Exact Solutions in Structured Low-Rank Approximation. SIAM J. on Mat. Anal. and App.. 35(4):1521–1542, 2014.

[121] P. Parrilo. Semidefinite programming relaxations for semialgebraic problems. Mathematical Programming Ser.B 96(2):293–320, 2003.

[122] P. Parrilo, B. Sturmfels. Minimizing polynomial functions. Proceedings of the DIMACS Workshop on Algorithmic and Quantitative Aspects of Real Algebraic Geometry in Mathematics and Computer Science (March 2001), (eds. S. Basu and L. Gonzalez-Vega), American Mathematical Society, 2003, pp. 83–100.

[123] V.Y. Pan, E. Tsigaridas, Z. Liang. Simple and efficient real root finding for a univariate polynomial. hal-01105309. 2015.

[124] G. Pataki. The geometry of cone-LP's. In: Wolkowicz, Saigal, Vandenberghe Eds. Handbook of Semidefinite Programming, 29–66, 2000.

[125] M. Kocvara, M. Stingl. PENSDP

[126] D. Plaumann, B. Sturmfels, C. Vinzant. Quartic curves and their bitangents. Journal of Symbolic Computation, 46:712–733, 2011.

[127] POVRAY Persistence of Vision Raytracer (Version 3.6): http://www.povray.org/download/.

[128] V. Powers, T. Woermann. An algorithm for sums of squares of real polynomials. J. Pure and Appl. Alg. 127 (1998), 99–104.

[129] M. Putinar. Positive polynomials on compact sets. Indiana University Mathematics Journal. 42(3):969–984, 1993.

[130] J.C. Ottem, K. Ranestad, B. Sturmfels, C. Vinzant. Quartic spectrahedra. To appear in Mathematical Programming, Series B, Special issue on Polynomial Optimization. Preprint arXiv:1311.3675, 2013.

[131] M. Ramana, A.J. Goldman. Some geometric results in semidefinite programming. J. Global Optim., 7:33–50, 1995.

[132] M. Ramana, L. Tunçel, H. Wolkowicz. Strong duality for semidefinite programming. SIAM J. on Opt. 7(3):641–662, 1997.

[133] J. Renegar. On the computational complexity and geometry of the first order theory of the reals. Journal of Symbolic Computation 13(3):255–352, (1992).

[134] B. Reznick. Sums of even powers of real linear forms. Mem. Amer. Math. Soc., 96(463), 1992.

[135] M. Safey El Din. Raglib (real algebraic geometry library), Maple package. `www-polsys.lip6.fr/~safey`

[136] F. Rouillier. Solving zero-dimensional systems through the rational univariate representation. Journal of Applicable Algebra in Engineering, Communication and Computing. 9(5):433–461, 1999.

[137] M. Safey El Din, É. Schost. Polar varieties and computation of one point in each connected component of a smooth real algebraic set. Proceedings of ISSAC 2003, Philadelphia, 224–231, 2003.

[138] M. Safey El Din, É. Schost. Properness defects of projections and computation of one point in each connected component of a real algebraic set. Discrete and Computational Geometry, 32(3):417–430, 2004.

[139] M. Safey El Din, É. Schost. A nearly optimal algorithm for deciding connectivity queries in smooth and bounded real algebraic sets. Preprint arXiv:1307.7836, 2013.

[140] M. Safey El Din. Finding sampling points on real hypersurfaces is easier in singular situations. In Electronic proceedings of MEGA (Effective Methods in Algebraic Geometry), 2005.

[141] R. Sanyal, F. Sottile, B. Sturmfels. Orbitopes. Mathematika 57:275–314, 2011.

[142] C. Scheiderer. Sums of squares of polynomials with rational coefficients. To appear in J. European Math. Soc., 15 pp., 2013.

[143] C. Scheiderer. Semidefinite representation for convex hulls of real algebraic curves. Preprint arXiv 1208.3865v3, August 2012.

[144] C. Scheiderer. Positivity and sums of squares: A guide to recent results. In M. Putinar, S. Sullivant, eds. Emerging Applications of Algebraic Geometry, IMA Volumes Math. Appl. 149, Springer, pp. 271–324, 2009.

[145] I. Shafarevich. Basic algebraic geometry 1. Springer, Berlin, 1977.

[146] K. Schmüdgen. The K-moment problem for compact semi-algebraic sets. Mathematische Annalen, 289:203–206, 1991.

[147] M. Schweighofer. On the complexity of Schmüdgen Positivstellensatz. Journal of Complexity 20, 529–543 (2004).

[148] R. Sinn, B. Sturmfels. Generic spectrahedral shadows. SIAM J. on Opt., 25(2):1209–1220, 2015.

[149] P.-J. Spaenlehauer. Résolution de systèmes multi-homogènes et déterminantiels. Algorithmes, Complexité, Applications. PhD thesis, Université Pierre et Marie Curie, Paris, October 2012.

[150] M. Spivak. Calculus on manifolds. WA Benjamin New York. Vol 1 (1965).

[151] V. Strassen. Gaussian elimination is not optimal. Numerische Mathematik, 13:354–356, 1969.

[152] J.F. Sturm. SeDuMi version 1.1R3. Advanced Optimization Laboratory at McMaster University, Canada, 2006.

[153] B. Sturmfels. What is a Gröbner basis. Notices of the American Mathematical Society, 52:1199–1200, 2005.

[154] F. Tanturri. Degeneracy loci of morphisms between vector bundles. PhD Thesis, SISSA, International School for Advanced Studies, Trieste, 2013.

[155] S. Tarbouriech, G. Garcia, J.M. Gomes da Silva, I. Queinnec. Stability and stabilization of linear systems with saturating actuators. Springer Verlag London limited, 2011.

[156] A. Tarski. A decision method for elementary algebra and geometry. University of California Press, 1951.

[157] M.J. Todd. Semidefinite optimization. Acta Numerica, 10:515–560, 2001.

[158] J. Löfberg. YALMIP: a toolbox for modeling and optimization in Matlab. Proceedings of the IEEE Symposium on Computer-Aided Control System Design (CACSD), Taipei, Taiwan, 2004.

[159] A. Varvitsiotis. Combinatorial conditions for low rank solutions in semidefinite programming. PhD Thesis. Tilburg University, Tilburg, Netherlands, 2013.

[160] C. Vinzant. Real algebraic geometry in convex optimization. PhD Thesis. University of California at Berkeley, 2011.

Index