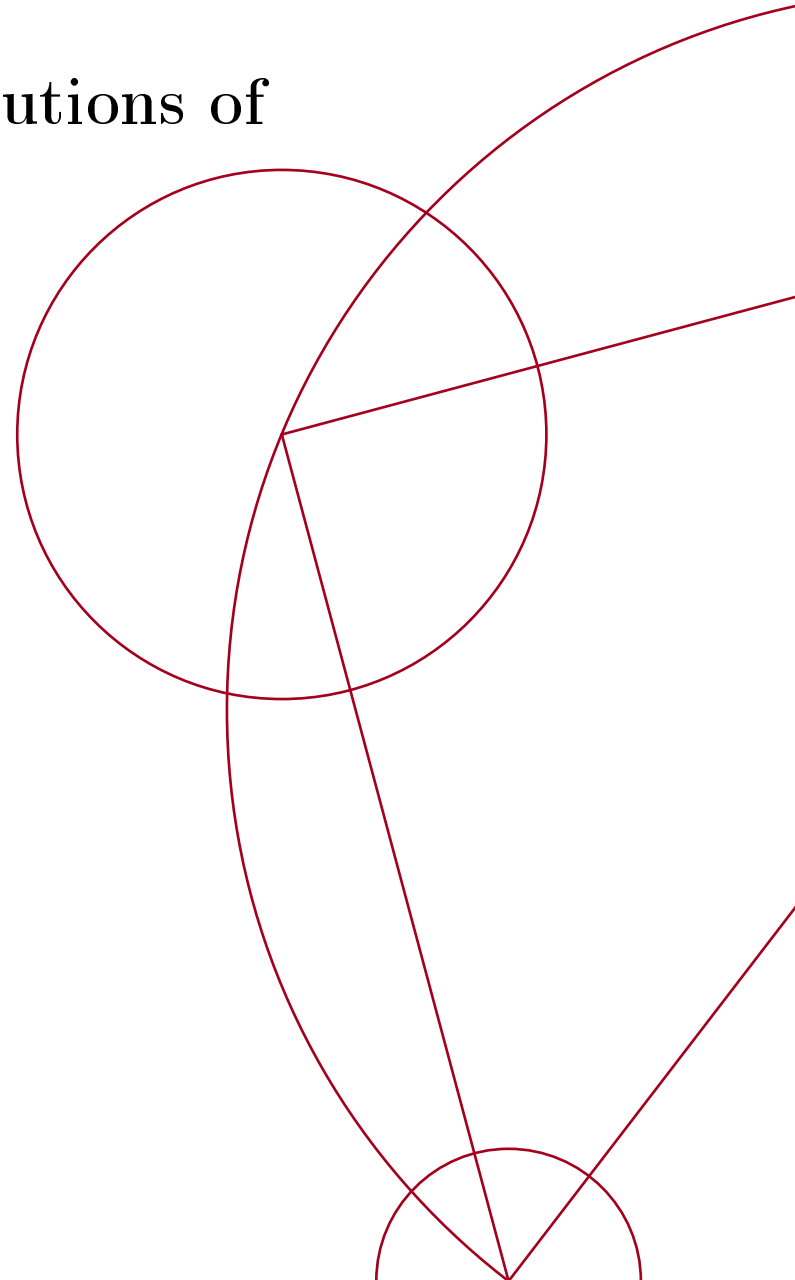
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF COPENHAGEN

June 2025

Bounds on positive solutions of polynomial systems

Marie Stuhr Kaltoft



Master's Thesis

Advisors: Elisenda Feliu, Carles Checa

Abstract

In this thesis, we consider the problem of bounding the number of positive solutions to a multivariate system of polynomials. We first consider the case of a single univariate polynomial. This is quite well understood, and we will consider bounds on both the number of positive solutions, as well as the number of solutions within an interval. We then give a characterization of when all the roots of a univariate polynomial are real. After the characterization, we consider an easier to verify condition for all roots to be real, which we further explore through computer implementations. For the rest of the thesis, we work with systems of n polynomials in n variables. First, we consider the BKK-bound, which bounds the number of complex solutions based on the Newton polytope(s) of the polynomials. Next, we introduce fewnomial theory and cover Gale duality, which we finally use to prove a generalization of Descartes' rule of signs for polynomial systems supported on circuits.

Declaration of use of generative AI tools

No generative AI has been used for this thesis.

Acknowledgments

I am very grateful to my advisors, Elisenda Feliu and Carles Checa, for their guidance throughout the process. Your enthusiasm and encouragement meant a lot to me.

I would like to thank my friend Anton Fehnker, who has been sitting across the table from me for these last four months. Thank you for the exceptionally thorough proofreading, and for always being ready to discuss various mathematical problems. Going through the whole master's thesis process together has made it quite fun and certainly less daunting. On that note, I would also like to thank the members of the Thesis Breakfast Club. Our biweekly meetings have been both fun and very motivational.

Lastly, I am thankful for my friends and family for their support during this process. In particular, I would like to thank my husband, Stig, for his patience and for always listening to my ramblings about math — both during the writing of my master's thesis and for the duration of my degree.

Contents

List of Symbols

General symbols

$\mathbb{N}$	Natural numbers with zero, page 3
$[m]$	$\{0, \ldots, m-1\}$, page 3
S_n	Symmetric group of degree n, page 3
M^T	Transpose of a matrix M, page 4
vol	Euclidean volume, page 4

Algebraic geometry

$\mathbb{T}$	Complex torus $\mathbb{C}^\times$, page 4
LC	Leading coefficient, page 4
supp	Support, page 3
NP	Newton polytope, page 3
conv	Convex hull, page 3
pos	Positive cone, page 4
Aff	Affine hull, page 4
MV	Mixed volume, page 4

Thesis specific notation

sgnvar	Variation in sign, page 5
f	Polynomial, page 3
δf	Sequence of derivatives of f, page 7
$\mathcal{A} \subset \mathbb{Z}^n$	Set of exponent vectors, page 3
A	Matrix corresponding to $\mathcal{A}$, page 3
C	Coefficient matrix, page 3
$C \cdot x^A = 0$	Matrix representation of polynomial system, page 3
$\varphi_{\mathcal{A}}$	$\mathbb{T}^n \to \mathbb{P}^{\mathcal{A}}$, $x \mapsto [x^a \mid a \in \mathcal{A}]$, page 26
B	Gale dual of A, page 38
P	Gale dual of C, page 38
$n_{\mathcal{A}}(C)$	Number of positive solutions of $C \cdot x^A = 0$, page 43
Δ_p	Positive chamber, page 38

1 Introduction

Given a system of equations in multiple variables, how can we determine the common solutions? Or even just find out how many solutions the system has? In both mathematics itself and in its application, we often need to solve multivariate systems of equations. This root finding problem is, in general, extremely difficult computationally. Even just determining the *number* of solutions is surprisingly hard. If we consider general systems with no additional constraints, the problem is practically impossible. However, as luck will have it, most polynomial systems coming from problems appearing in nature have some sort of built-in structure, which we can exploit to better understand the solutions of said systems. If we want to consider the complex solutions of a system of polynomials, the situation is very well understood. The theory is nicely developed, and we have algorithms for both symbolic computation (using, e.g., Gröbner bases or resultants) and numerical computation (using, e.g., homotopy continuation). However, the primary goal of considering problems from nature as systems of equations is to determine potential real solutions and, in particular, positive solutions, both of which are not considered by the above mentioned theory. A big challenge of real algebraic geometry is that there are far fewer strong results available compared to complex algebraic geometry. And the results we do have are typically not analogous to the results for complex solutions. Thus, it is clear that many problems arise, when we want to restrict our search for solutions to $\mathbb{R}$ instead of all of $\mathbb{C}$.

Looking back, the problem of determining the number of complex solutions in the univariate case has long ago been solved by the fundamental theorem of algebra.

Theorem (Fundamental Theorem of Algebra)**.** *A polynomial $f \in \mathbb{R}[x]$ of degree n has precisely n roots in $\mathbb{C}$ counted with multiplicity.*

For bounds on the number of positive solutions, the first well-known milestone for univariate polynomials (Descartes' rule of signs) was formulated by Réne Descartes in 1637. However, it was in the 19th century that the now standard results about real solutions of univariate polynomials were formulated. These are for example the Budan–Fourier Theorem (Budan in 1807 and Fourier in 1820), which generalizes Descartes' rule of signs to bound the number of roots within an interval, and Sturm's Theorem (1829), which gives an algorithm to find real roots within an interval. These are all now considered standard results, and it is this accuracy and strength, which we wish for in the multivariate case.

For multivariate systems of polynomials, Bézout's Theorem (1779) gives an upper bound to the number of complex solutions to general systems of polynomials. It is the most immediate generalization of the Fundamental Theorem of Algebra, but the bound is naturally very large. When the system of polynomials has some more "structure", we can say more. In 1976, Bernstein, Kushnirenko, and Khovanskii published the (now well-known) BKK-bound, which gives a bound on the number of complex solutions away from zero. The number of complex solutions to a univariate polynomial is, in general, equal to the degree of the polynomial, while the number of real solutions (as we will see) is bounded by the number of terms. Khovanskii generalized this concept with the introduction of fewnomial theory and his famous Fewnomial Bound (1980). It was revolutionary exactly due to the idea that even in the multivariate case, the number of real roots should depend on the number of terms, not the degree. Many others have come up with partial generalizations in the same spirit, which brings us to today. In 2022, Frédéric Bihan, Alicia Dickenstein, and Jens Forsgård published a partial generalization of Descartes' rule of signs for multivariate polynomial systems.

Example. To better understand the significance of these improvements, we consider the following system of 3 polynomials in 3 variables:

$$\begin{aligned} f &= x^{53} + y^{42}z^{10} - 3z^{47} + x^{31}y^{11}z^{27} + y^{37}, \\ g &= -2x^{53} + 5y^{42}z^{10} + 42z^{47} - 11x^{31}y^{11}z^{27} + 4y^{37}, \\ h &= 2x^{53} + 3y^{42}z^{10} - z^{47} + 83x^{31}y^{11}z^{27} - 13y^{37}. \end{aligned}$$

When we discuss number of solutions, we will, in this example, assume that we do not have any infinite component in the solution set.

In $\mathbb{C}^3$, Bézout's Theorem would give a bound of 328,509 solutions. Using the BKK-bound, we find that there are at most 70,310 solutions in the complex numbers without zero, which is already a big improvement. Moving to the realm of positive real solutions, Khovanskii's Fewnomial Bound implies that there are at most 16,384 (nondegenerate) positive solutions. •

In this thesis, we explore the problem of bounding the number of positive solutions to a multivariate system of polynomials. Essential for both the univariate and multivariate results is the notion of variation in sign of a finite sequence of real numbers. The variation in sign counts the number of consecutive entries in the sequence c_i and c_{i+1}, where $c_i c_{i+1} < 0$. We begin in Chapter 2, where we explore the problem in the case of a single univariate polynomial. In Section 2.1, which is primarily based on the exposition in [Sot11, Chapter 2], we consider different bounds on the number of real roots, as well as their implications. We begin by studying the classical Descartes' rule of signs (Theorem 2.5), which states that the number of positive roots of a univariate polynomial is bounded from above by the variation in sign of the coefficients of said polynomial. We mention several corollaries to Descartes' rule of sign. One conceptually significant corollary of Descartes' rule of signs, which describes the idea that we want to generalize, is that a polynomial with m terms has at most $m-1$ positive solutions. To better understand the real solutions, we then consider the Budan–Fourier Theorem (Theorem 2.12), which builds on Descartes' rule of signs — it is essentially the same statement, except instead of bounding the number of roots in the positive orthant, it bounds the number of roots in an interval. We conclude the section with Sturm's Theorem, which (unlike the others) tells us *exactly* how many roots a polynomial has in a given interval. To round out our exploration of univariate polynomials, we briefly consider a somewhat different question; when are all roots of a polynomial real? In Section 2.2, we state the Hermite–Sylvester Criterion, which characterizes exactly this. As this characterization is computationally burdensome for higher degrees, we also consider, in Section 2.3, a simpler sufficient condition due to Kurtz (Theorem 2.26). As this is not a necessary condition, an obvious question is when does it hold? We explore this (and another question) through implementations in Julia with the use of the OSCAR package [OSC25]. The computer implementations we refer to can be found at [Kal25], where they are listed according to section.

Next, we are ready to consider the multivariate version of our problem. From Chapter 3 and onwards, we work with systems of n polynomials in n variables. In Chapter 3, we begin by considering bounds on the number of complex solutions to such polynomial systems. This understanding will be helpful, when we tackle the problem of positive solutions. This chapter is based on the exposition in [Sot11, Chapter 3], where we have fleshed out many of the missing details. The vast majority of the chapter is dedicated to proving Kushnirenko's Theorem (Theorem 3.2), which is the BKK-bound for unmixed system, i.e., systems of polynomials with the same support. In the last part of the chapter, we state Bernstein's Theorem (Theorem 3.21), which extends Kushnirenko's Theorem to the case of mixed systems. Together these theorems are known as the BKK-bound (the last K refers to Khovanskii).

Having explored the landscape of complex solutions, we now get ready to dive into the true purpose of this thesis — the study of bounds on positive real solutions! We open Chapter 4 by introducing the theory of fewnomials along with Khovanskii's Fewnomial Bound (Theorem 4.1). To bound the number of positive solutions, we will need a completely different tool from the ones we used in Chapter 3. The chapter is dedicated to introducing the reader to the concept of Gale duality, which will be the central tool needed for our problem. The chapter is based on the material from [Sot11, Chapter 6] with the exception of Subsection 4.1.1, where we formulate a special case of Gale duality.

Finally, we reach Chapter 5, where we consider the 2022 paper by Bihan, Dickenstein and Forsgård [BDF22], as well as the 2016 paper it is based on [BD16]. The paper concerns systems of n polynomials in n variables, which are supported on so-called circuits. A circuit is a set of exponent vectors of cardinality $n+2$, where the differences of the exponent vectors make up a minimally linearly dependent set spanning $\mathbb{R}^n$. Recall that Descartes' rule of signs in the univariate case gives the bound of $m-1$ positive solutions to a polynomial with m terms. The main result of the article (Theorem 5.10) implies the bound $n+1$ on the number of positive solutions to such a system. The number of monomials in the system is $m = n+2$. Hence, we again get the bound $m-1$ on the number of positive solutions. In this way, Theorem 5.10 turns out to nicely generalize Descartes' rule of signs to the case of circuits.

It turns out that the exponent vectors, in the above example, form a circuit. The bound from Theorem 5.10 reduces the bound on the number of positive solutions all the way down to 4, which is a remarkable improvement!

Notation and preliminaries

For general background on algebraic geometry, we refer to [CLO15], [CLO05], and [Gat14].

Before we truly get started, we introduce some conventions and notation, which will be used throughout the thesis. We assume that $\mathbb{N}$ includes zero. For a positive integer m, we let $[m] = \{0, \dots, m-1\}$. We let S_n denote the symmetric group of degree n. Let $\mathcal{A} \subset \mathbb{Z}^n$ be a set of integer vectors. We can consider $\mathcal{A}$ as the matrix $A = \begin{pmatrix} a_0 & \dots & a_{|\mathcal{A}|-1} \end{pmatrix} \in \mathbb{Z}^{n \times |\mathcal{A}|}$, where $a_i \in \mathbb{Z}^n$ are column vectors. For $x = (x_1, \dots, x_n) \in \mathbb{C}^n$, we write

$$x^A = \begin{pmatrix} x^{a_0} \\ \vdots \\ x^{a_{|\mathcal{A}|-1}} \end{pmatrix},$$

where $x^{a_i} = \prod_{j=1}^n x_j^{(a_i)_j}$. We say that

$$f(x) = \sum_{a \in \mathcal{A}} c_a x^a$$

is a *sparse polynomial* in n variables $x = (x_1, \dots, x_n)$ with *support* $\operatorname{supp}(f) = \mathcal{A}$. Then $\operatorname{NP}(f) = \operatorname{conv}(\operatorname{supp}(f))$ is called the *Newton polytope* of f, where conv denotes the convex hull. For a coefficient matrix $C \in \mathbb{C}^{n \times |\mathcal{A}|}$, the corresponding system of polynomials can be represented by matrices as $C \cdot x^A = 0$.

Example. Let $\mathcal{A} = \{(0,0), (1,0), (1,1)\}$ be a set of exponent vectors, and let $C = \begin{pmatrix} 3 & 1 & -4 \\ -1 & 23 & -51 \end{pmatrix}$

be a coefficient matrix. We get the matrix $A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. Hence,

$$x^A = \begin{pmatrix} 1 \\ x_1 \\ x_1 x_2 \end{pmatrix},$$

and our corresponding polynomial system is

$$C \cdot x^A = \begin{pmatrix} 3 & 1 & -4 \\ -1 & 23 & -51 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_1 x_2 \end{pmatrix} = \begin{pmatrix} 3 + x_1 - 4x_1x_2 \\ -1 + 23x_1 - 51x_1x_2 \end{pmatrix}.$$

•

For a univariate polynomial f, we define $\mathrm{LC}(f)$ to be the leading coefficient of f. We write M^T to denote the transpose of a matrix M.

By $\mathbb{R}_{>0}$ we denote the strictly positive real numbers. Analogously, we write $\mathbb{R}_{<0}, \mathbb{Q}_{>0}, \mathbb{Q}_{\geq 0}$, etc. We write $\mathbb{T} = \mathbb{C}^\times$ for the complex torus, and $\mathbb{T}^{\mathbb{R}} = \mathbb{R}^\times$ for the real torus. In general, we write $\mathbb{R}$ in the exponent, when a space is restricted to $\mathbb{R}$. Complex projective space in dimension n is denoted $\mathbb{P}^n$. When we write $\mathbb{P}^{\mathcal{A}}$, we mean projective space indexed over the elements of $\mathcal{A}$. Similarly, we also write, e.g., $\mathbb{T}^{\mathcal{A}}$ for the torus indexed over the elements of $\mathcal{A}$. Given a variety X, we use $\overline{Y}$ to denote the Zariski closure of a subset Y in X. To avoid ambiguity, we sometimes write $\overline{Y}^X$ to mean the same.

The results in this thesis are typically for isolated solutions, i.e., a solution x, such that there exists a neighborhood of x containing no other solutions to our system. We will often consider so-called generic systems of polynomials. The general definition of genericity is as follows.

Definition. A property holds *generically* on a variety, if it holds on a non-empty Zariski open subset of the variety.

Hence, a system of polynomials is *generic* given its support, if the coefficients lie in a non-empty Zariski open subset. This is equivalent to saying that, given a system of n polynomials in n variables, their derivatives at each solution span $\mathbb{C}^n$.

Given a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^n$, we define the *positive cone*

$$\mathrm{pos}(v_1, \ldots, v_m) = \left\{ \sum_{i=1}^m \eta_i v_i \;\middle|\; \eta_i \in \mathbb{R}_{>0} \right\},$$

and the *affine span*

$$\mathrm{Aff}(v_1, \ldots, v_m) = \left\{ \sum_{i=1}^m \eta_i v_i \;\middle|\; \eta_i \in \mathbb{R}, \sum_{i=1}^m \eta_i = 1 \right\}.$$

The $\mathbb{Z}$-affine span is the affine span with the additional condition that $\eta_i \in \mathbb{Z}$.

For a polytope Δ, we define $\mathrm{vol}(\Delta)$ to be the Euclidean volume of the polytope. We write vol_n (when relevant) to specify that it is the n-dimensional Euclidean volume. We use this in connection with mixed volume: Given a collection of polytopes $\Delta_1, \ldots, \Delta_n$, their n-dimensional *mixed volume*, $\mathrm{MV}(\Delta_1, \ldots, \Delta_n)$, is the coefficient of $\ell_1 \cdots \ell_n$ in $\mathrm{vol}_n(\ell_1\Delta_1 + \cdots + \ell_n\Delta_n)$, where $\ell_1\Delta_1 + \cdots + \ell_n\Delta_n$ denotes the Minkowski sum of the scaled polytopes.

2 The Univariate Case

In this chapter, we primarily consider the problem of bounding the number of (positive) real roots of univariate polynomials. We know a lot about this problem, and there are many sufficient conditions for real-rootedness, as well as several different characterizations. First, we will discuss different bounds on the number of (positive) real solutions to a univariate polynomial. We will prove Descartes' rule of signs, and a generalization by Budan and Fourier. Budan–Fourier gives us an upper bound on the number of real roots within an interval. Descartes' rule of signs is a special case, which gives an upper bound on the number of *positive* real roots. It states that the number of positive real roots of a polynomial is at most the number of sign changes between consecutive coefficients of the polynomial. Next, we consider Sturm's theorem, which allows us to determine the exact number of real roots in an interval. But none of these tells us, when a polynomial will have all roots real. In this regard, we will consider a way to characterize polynomials with all real roots. Characterizations, although very useful, can require a lot of computations — this leads us to consider a condition, due to Kurtz, for all roots of a polynomial to be real, which is easier to check. As a little extra treat, at the end of the chapter, we will experimentally explore the theorem by Kurtz to find out how "often" the main condition of the theorem is satisfied.

2.1 Bounds on the number of real roots

The proofs in this section rely heavily on the following classical result from analysis.

Lemma 2.1 (Lagrange's Mean Value Theorem)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function, and assume that f is differentiable on (a,b). Then there exists some $c \in (a,b)$, such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

For us, the following special case will be most important.

Lemma 2.2 (Rolle's Theorem)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function, and assume that f is differentiable on (a,b). If $f(a) = f(b)$, then there exists $c \in (a,b)$, such that $f'(c) = 0$.*

Throughout this chapter, we consider polynomials in $\mathbb{R}[x]$. We also assume throughout that the exponents of our polynomials are nonnegative integers, however, most results also hold for real exponents. All bounds in this section will be stated in terms of the following definition.

Definition 2.3. The *variation* of a finite sequence of real numbers is

$$\operatorname{sgnvar}(k_1, \ldots, k_m) := \left|\{i \mid 1 \leq i \leq m\colon k_{i-1}k_i < 0\}\right|.$$

Let $F = (f_1, \ldots, f_m)$ be a sequence of polynomials in $\mathbb{R}[x]$, and let $a \in \mathbb{R}$. Then the variation of $f_1(a), \ldots, f_m(a)$ is denoted $\operatorname{sgnvar}(F, a)$. We extend this definition to all of $\mathbb{R} \cup \{\pm\infty\}$ by letting

$$\begin{aligned}\operatorname{sgnvar}(F, \infty) &:= \operatorname{sgnvar}\left(\operatorname{LC}(f_1(x)), \ldots, \operatorname{LC}(f_m(x))\right), \quad \text{and}\\ \operatorname{sgnvar}(F, -\infty) &:= \operatorname{sgnvar}\left(\operatorname{LC}(f_1(-x)), \ldots, \operatorname{LC}(f_m(-x))\right).\end{aligned}$$

•

Remark 2.4. Note that multiplying a sequence by a non-zero real number does not change the variation of the sequence. Therefore, we can (and will sometimes) make assumptions without loss of generality about the sign of $f_j(c)$ for some $f_j \in F$. •

Let $r(f, I)$ denote the number of roots of a polynomial f in the subset $I \subseteq \mathbb{R}$ counted with multiplicity. We are now able to formulate our first bound, which Descartes described in *La Géométrie* from 1637. An important feature of this bound is that it is independent of the degree of the polynomial.

Theorem 2.5 (Descartes' rule of signs, [Des37])**.** *Let $f = \sum_{i=1}^{n} c_i x^{a_i}$ be a real polynomial, where $c_i \neq 0$ for all i and $a_1 < \cdots < a_n$. Then*

$$r(f, \mathbb{R}_{>0}) \leq \operatorname{sgnvar}(c_1, \ldots, c_n),$$

and the difference is even.

We will prove Descartes' rule of signs as a special case of Budan–Fourier. The following example will be used throughout the section to illustrate the capabilities of the different bounds.

Example 2.6. Consider the polynomial $f = x^5+4x^4+2x^3-2x^2+x-6$. As $\operatorname{sgnvar}(-6, 1, -2, 2, 4, 1) = 3$, it follows, from Descartes' rule of signs, that f has either one or three positive real roots. •

We can also use Descartes' rule of signs to bound the number of *negative* roots of a polynomial.

Corollary 2.7. *Let $f = \sum_{i=1}^{n} c_i x^{a_i}$ be a real polynomial, where $c_i \neq 0$ for all i and $a_1 < \cdots < a_n$. The number of negative roots of f counted with multiplicity is at most the variation of the sequence $((-1)^{a_i} c_i)_{i=1,\ldots,n}$.*

Proof. We have that $f(-x) = \sum_{i=1}^{n} c_i(-x)^{a_i} = \sum_{i=0}^{n}(-1)^{a_i} c_i x^{a_i}$. Hence,

$$r(f(-x), \mathbb{R}_{>0}) \leq \operatorname{sgnvar}\left((-1)^{a_i} c_i\right)_{i=1,\ldots,n},$$

by Descartes' rule of signs. It follows that $r(f, \mathbb{R}_{<0}) \leq \operatorname{sgnvar}\left((-1)^{a_i} c_i\right)_{i=1,\ldots,n}$. ■

Using this corollary, we try to understand our previous example better.

Example 2.8 (Continuation of Example 2.6)**.** We again consider $f = x^5 + 4x^4 + 2x^3 - 2x^2 + x - 6$. For f, we have that

$$\operatorname{sgnvar}\left((-1)^{a_i} c_i\right)_{i=1,\ldots,n} = \operatorname{sgnvar}(-6, -1, -2, -2, 4, -5) = 2.$$

Thus, f has either zero or two negative real roots, but we do not yet know which. •

The following corollary is immediate from Descartes' rule of signs, and it gives us a nice upper bound on the number of *positive* solutions, which is independent of the choice of coefficients. This result will be important to keep in mind, when we, in Chapter 5, consider a generalization of Descartes' rule of signs to the multivariate case.

Corollary 2.9. *A polynomial with n terms has at most $n-1$ positive real roots.*

This corollary highlights that the number of real roots (both positive and negative, by Corollary 2.7) does not so much depend on the degree, but instead depends on the number of monomials.

Remark 2.10. Let $f = \sum_{i=1}^{n} c_i x^{a_i}$ be a real polynomial, where $c_i \neq 0$ for all i and $a_1 < \cdots < a_n$. If $c_i > 0$ for all i, then $\operatorname{sgnvar}(c_1, \ldots, c_n) = 0$. So if a polynomial has only positive coefficients (or only negative coefficients), then, by Descartes' rule of signs, the polynomial cannot have any positive roots. Therefore, such a polynomial must have only negative (or zero) roots. •

One of the most interesting things about Descartes' rule of signs is the fact that it is a sharp bound. If all solutions are real, it is exact.

Corollary 2.11. *If all roots of f are real, then Descartes' rule of signs is exact.*

Proof. Let $f(x) = c_0 + c_1 x + \cdots + c_n x^n \in \mathbb{R}[x]$ be a real-rooted polynomial. By Descartes' rule of signs, $r(f, \mathbb{R}_{>0}) \leq \operatorname{sgnvar}(c_0, \ldots, c_n)$. Additionally, by Corollary 2.7,

$$r(f, \mathbb{R}_{<0}) \leq \operatorname{sgnvar}(c_0, -c_1, \ldots, (-1)^n c_n).$$

If all roots are positive, Descartes' rule of signs requires $\operatorname{sgnvar}(c_0, \ldots, c_n) = n$. Note that, in general, $\operatorname{sgnvar}(c_0, -c_1, \ldots, (-1)^n c_n) \leq n - \operatorname{sgnvar}(c_0, \ldots, c_n)$. Assume $x = 0$ is not a root. As all roots are real, we must have $\operatorname{sgnvar}(c_0, -c_1, \ldots, (-1)^n c_n) + \operatorname{sgnvar}(c_0, \ldots, c_n) = n$. Hence,

$$\begin{aligned} r(f, \mathbb{R}_{>0}) &= \operatorname{sgnvar}(c_0, \ldots, c_n), \\ r(f, \mathbb{R}_{<0}) &= \operatorname{sgnvar}(c_0, -c_1, \ldots, (-1)^n c_n). \end{aligned}$$

Now, assume $x = 0$ is a simple root. Then $c_0 = 0$, so $\operatorname{sgnvar}(c_0, \ldots, c_n) = \operatorname{sgnvar}(c_1, \ldots, c_n)$. As all roots are real, there are $n - 1$ nonzero roots, and the same argument applies. If $x = 0$ is a multiple root, then more of the coefficients will be zero, and the argument is the same. ■

None of these corollaries, however, help us understand our running example better. Thus, we now consider a generalization of Descartes' rule of signs. Let f be a polynomial of degree k. We denote the *sequence of derivatives* of f by

$$\delta f := \left(f, f', \ldots, f^{(k)}\right).$$

In 1807, Budan proved a generalization of Descartes' rule of signs. The same result was proven independently in 1820 by Fourier.

Theorem 2.12 (Budan–Fourier, [Bud07, Fou20])**.** *Let $f \in \mathbb{R}[x]$ and $a, b \in \mathbb{R} \cup \{\pm\infty\}$, such that $a < b$. Then*

$$\operatorname{sgnvar}(\delta f, a) - \operatorname{sgnvar}(\delta f, b) \geq r(f, (a, b]),$$

and the difference (between the left hand side and the right hand side of the inequality) is even.

We can now prove Descartes' rule of signs as a special case of Budan–Fourier.

Proof of Theorem 2.5. Let $f = \sum_{i=1}^{n} c_i x^{a_i}$ be a real polynomial, where $c_i \neq 0$ and $a_1 < \cdots < a_n$. We set $a = 0$, and $b = \infty$. Then

$$\begin{aligned} \operatorname{sgnvar}(\delta f, a) &= \operatorname{sgnvar}(\delta f, 0) = \operatorname{sgnvar}(c_1, \ldots, c_n), \\ \operatorname{sgnvar}(\delta f, b) &= \operatorname{sgnvar}(\delta f, \infty) = \operatorname{sgnvar}(c_n, \ldots, c_n) = 0. \end{aligned}$$

It now follows, from Budan–Fourier, that

$$\operatorname{sgnvar}(c_1, \ldots, c_n) = \operatorname{sgnvar}(\delta f, 0) - \operatorname{sgnvar}(\delta f, \infty) \geq r(f, \mathbb{R}_{>0}).$$

■

The proof of Budan–Fourier is quite long and technical, so we will first give a small sketch of the idea behind the proof so that the actual proof will be easier for the reader to follow.

Proof idea for Theorem 2.12. In the degree one case, the variation in sign is equal to the number of positive roots.

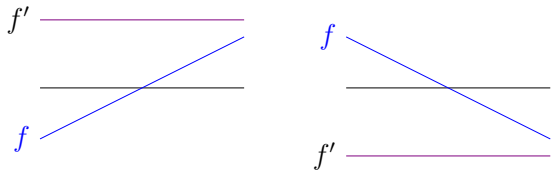


Figure 2.1: The two possibilities in degree one.

We now consider an example of a polynomial f of degree three as in Figure 2.2. We apply Rolle's Theorem — between two sign changes (roots) of f there will be at least one sign change (root) of f'.

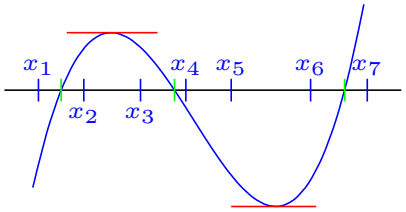


Figure 2.2

In Table 2.1, we list the signs of f and its derivatives at the marked positions. Set $a = x_1$ and $b = x_7$ from Figure 2.2. From Table 2.1, we see that $\operatorname{sgnvar}(\delta f, a) = 3$. Then, from the table and the figure, we see that $\operatorname{sgnvar}(\delta f, a) - \operatorname{sgnvar}(\delta f, t) - r(f, (a, t]) \geq 0$ for all $t \in (a, b]$.

t	$f(t)$	$f'(t)$	$f''(t)$	$f'''(t)$	$\operatorname{sgnvar}(\delta f, t)$	$r(f, (a, t])$
x_1	-	+	-	+	3	0
x_2	+	+	-	+	2	1
x_3	+	-	-	+	2	1
x_4	-	-	-	+	1	2
x_5	-	-	+	+	1	2
x_6	-	+	+	+	1	2
x_7	+	+	+	+	0	3

Table 2.1: Signs at the positions marked in Figure 2.2.

■

While the proof idea is not complete, it helps us gain a better intuition of the proof.

Proof of Theorem 2.12. To prove the theorem, we consider $\operatorname{sgnvar}(\delta f, t)$ as t increases from a to b. The variation can only change, when t passes a number, c, which is a root of some polynomial in δf, as then the sign of said polynomial would change. So suppose throughout that c is a root of some polynomial in δf. Whenever we consider some c, let $\varepsilon > 0$, such that $[c-\varepsilon, c+\varepsilon]$ contains no roots (other than c) of any polynomial in δf. Let m be the order of vanishing of f at c, i.e., $f^{(m)}(c) \neq 0$, but $f^{(i)}(c) = 0$ for $0 \leq i < m$. We want to show that the following holds:

$$\begin{aligned} \operatorname{sgnvar}(\delta f, c) &= \operatorname{sgnvar}(\delta f, c+\varepsilon) \\ \operatorname{sgnvar}(\delta f, c-\varepsilon) &\geq \operatorname{sgnvar}(\delta f, c) + m, \text{ and the difference is even.} \end{aligned} \tag{2.1}$$

This essentially says that $\operatorname{sgnvar}(\delta f, t)$ decreases as t increases, and that this decrease specifically occurs at roots of f (as we saw in Table 2.1).

We proceed by induction on the degree of f.

Suppose that $\deg(f) = 1$. Then we must have $m \leq 1$. The case $m = 0$ is trivial, so we assume $m = 1$. Then $f(c) = 0$, and $f'(c) \neq 0$. If $f'(t) < 0$, then $f(c-\varepsilon) > 0$, and $f(c+\varepsilon) < 0$. Hence, $\operatorname{sgnvar}(\delta f, c-\varepsilon) = 1$, $\operatorname{sgnvar}(\delta f, c) = 0$, and $\operatorname{sgnvar}(\delta f, c+\varepsilon) = 0$, so the conditions from Equation (2.1) hold. If $f'(t) > 0$, then $f(c-\varepsilon) < 0$, and $f(c+\varepsilon) > 0$. Analogously, the conditions from Equation (2.1) hold.

Suppose now that $\deg(f) > 1$. We consider two cases: $f(c) = 0$ and $f(c) \neq 0$. Suppose first that $f(c) = 0$. Then $m > 0$. Hence, f' has multiplicity $m-1$ at c. Applying the induction hypothesis to f' gives us that

$$\begin{aligned} \operatorname{sgnvar}(\delta f', c) &= \operatorname{sgnvar}(\delta f', c+\varepsilon), \text{ and} \\ \operatorname{sgnvar}(\delta f', c-\varepsilon) &\geq \operatorname{sgnvar}(\delta f', c) + (m-1), \text{ and the difference is even.} \end{aligned}$$

Note that even if $m - 1 = 0$, the above still holds. It follows directly, from Lemma 2.1, that f and f' have the same sign at $c+\varepsilon$, and that they have opposite signs at $c-\varepsilon$.

As $\operatorname{sgnvar}(\delta f, c) = \operatorname{sgnvar}(0, f'(c), \ldots, f^{(k)}(c)) = \operatorname{sgnvar}(f'(c), \ldots, f^{(k)}(c)) = \operatorname{sgnvar}(\delta f', c)$, it follows that

$$\begin{aligned} \operatorname{sgnvar}(\delta f, c) &= \operatorname{sgnvar}(\delta f', c+\varepsilon) = \operatorname{sgnvar}(\delta f, c+\varepsilon), \text{ and} \\ \operatorname{sgnvar}(\delta f, c-\varepsilon) &= \operatorname{sgnvar}(\delta f', c-\varepsilon) + 1 \geq (\operatorname{sgnvar}(\delta f', c) + (m-1)) + 1 = \operatorname{sgnvar}(\delta f, c) + m, \end{aligned}$$

and the difference is even.

Suppose next that $f(c) \neq 0$, i.e., $m = 0$. We assume that $f'(c) = 0$ (if not, the same argument applies to the first polynomial in δf, which vanishes at c.) Let n be the multiplicity of f' at c, i.e., $f'(c) = \cdots = f^{(n)}(c) = 0$, and $f^{(n+1)}(c) \neq 0$. We assume that $f^{(n+1)}(c) > 0$ (recall Remark 2.4). We apply the first case (when $f(c) = 0$) to f' and get that

$$\begin{aligned} \operatorname{sgnvar}(\delta f', c) &= \operatorname{sgnvar}(\delta f', c+\varepsilon), \text{ and} \\ \operatorname{sgnvar}(\delta f', c-\varepsilon) &\geq \operatorname{sgnvar}(\delta f', c) + n, \text{ and the difference is even.} \end{aligned} \tag{2.2}$$

From the assumption that $f^{(n+1)}(c) > 0$, and by considering the Taylor expansion for f' at c, we get the following cases:

- If n is even, then $f'(c-\varepsilon), f'(c+\varepsilon) > 0$.
- If n is odd, then $f'(c-\varepsilon) < 0$, and $f'(c+\varepsilon) > 0$.

Then, Table 2.2 describes the signs of the first nonzero term in the sequence $\delta f'(t)$ for $t \in \{c-\varepsilon, c, c+\varepsilon\}$.

t	$c-\varepsilon$	c	$c+\varepsilon$
n even	+	+	+
n odd	-	+	+

Table 2.2: Sign of the first nonzero term of $\delta f'(t)$

In Table 2.3, we list the possible options for the signs of f at the different values of t.

t	$c-\varepsilon$	c	$c+\varepsilon$
Option 1	+	+	+
Option 2	-	-	-

Table 2.3: Sign of $f(t)$ (independent of n).

Lastly, we can complete the proof by considering Table 2.4. The equality

$$\operatorname{sgnvar}(\delta f, c) = \operatorname{sgnvar}(\delta f, c+\varepsilon)$$

follows from the table together with Equation (2.2). We also want to show the relation

$$\operatorname{sgnvar}(\delta f, c) \geq \operatorname{sgnvar}(\delta f, c-\varepsilon) + m,$$

where the difference should be even. When n is even, the two columns in Table 2.4 are equal. When n is odd, the two columns have a difference of 1. Put together with Equation (2.2), the desired inequality holds with the difference being even in both cases.

Option	Parity of n	$c-\varepsilon$	c	$c+\varepsilon$
1	even	0	0	0
1	odd	+1	0	0
2	even	+1	+1	+1
2	odd	0	+1	+1

Table 2.4: Possible values a, where $\operatorname{sgnvar}(\delta f, t) = \operatorname{sgnvar}(\delta f', t) + a$.

This completes the proof of the conditions from Equation (2.1). We now apply this condition to prove the theorem. Let $c_1, \ldots, c_l$ be the roots of f in the interval $(a, b]$. We consider what happens as t passes through $(a, b]$. By definition, $r(f, (a, t])$ only changes, when t passes a root of f. By Equation (2.1), $\operatorname{sgnvar}(\delta f, t)$ either stays the same or decreases, when t increases. Let m_i be the

multiplicity of c_i as a root of f. Denote by ε_i the above defined ε corresponding to the root c_i. By repeated use of the conditions from Equation (2.1), we get that

$$\begin{aligned} r(f,(a,b]) + \operatorname{sgnvar}(\delta f, b) &\leq r(f,(a,c_l]) + \operatorname{sgnvar}(\delta f, c_l) \\ &\leq (r(f,(a,c_l-\varepsilon_l]) + m_l) + (\operatorname{sgnvar}(\delta f, c_l-\varepsilon_l) - m_l) \\ &= r(f,(a,c_l-\varepsilon_l]) + \operatorname{sgnvar}(\delta f, c_l-\varepsilon_l) \\ &\vdots \\ &\leq r(f,(a,c_1-\varepsilon_1]) + \operatorname{sgnvar}(\delta f, c_1-\varepsilon_1) \\ &\leq r(f,(a,a]) + \operatorname{sgnvar}(\delta f, a) = \operatorname{sgnvar}(\delta f, a), \end{aligned}$$

where the difference in each inequality is even. Finally, it follows that

$$\operatorname{sgnvar}(\delta f, a) - \operatorname{sgnvar}(\delta f, b) \geq r(f,(a,b]),$$

and the difference is even. ■

With this new knowledge, we attempt to further specify the number of negative roots in our running example.

Example 2.13 (Continuation of Example 2.8)**.** As before, we consider $f = x^5+4x^4+2x^3-2x^2+x-6$. We have that

$$\begin{aligned} \operatorname{sgnvar}(\delta f, -\infty) &= \operatorname{sgnvar}(-1,5,-20,60,-120,120) = 5, \\ \operatorname{sgnvar}(\delta f, 0) &= \operatorname{sgnvar}(-6,1,-4,12,96,120) = 3, \\ \operatorname{sgnvar}(\delta f, \infty) &= \operatorname{sgnvar}(1,5,20,60,120,120) = 0. \end{aligned}$$

This does not give us more information. However, we can also consider other intervals. We have that

$$\operatorname{sgnvar}\left(\delta f, \frac{1}{2}\right) = \operatorname{sgnvar}\left(-\frac{175}{32}, \frac{45}{16}, \frac{33}{2}, 75, 156, 120\right) = 1.$$

Hence, by Budan-Fourier, there must be exactly one root of f in $(1/2,\infty)$, zero or two roots in $(0,1/2]$ and zero or two roots in $(-\infty,0]$. •

The last result of this section will give us the exact number of real roots in an interval. It relies on the following definition.

Definition 2.14. The *Sylvester sequence* of two univariate polynomials f and g is $f_0 := f$, $f_1 := g$, $f_2, \ldots, f_k$, where $f_k = \gcd(f,g)$, and $-f_{i+1} = \operatorname{remainder}(f_{i-1}, f_i)$ is the remainder from the Euclidean algorithm, i.e., $f_{i-1} = q_i f_i - f_{i+1}$ for some polynomial q_i. The *Sturm sequence* of a univariate polynomial f is the Sylvester sequence of f and f'. •

In 1829, Sturm discovered the following result, which can be used to approximate real roots to arbitrary precision.

Theorem 2.15 (Sturm's Theorem, [Stu29])**.** *Let f be a univariate polynomial, and let $a, b \in \mathbb{R} \cup \{\pm\infty\}$ with $a < b$ and $f(a), f(b) \neq 0$. Then the number of roots of f in (a,b) counted* without *multiplicity is*

$$\operatorname{sgnvar}(F, a) - \operatorname{sgnvar}(F, b),$$

where F is the Sturm sequence of f.

Proof. The strategy of the proof is to consider $\operatorname{sgnvar}(F,t)$ as t passes roots of polynomials in F. We will show that it is only necessary to consider roots of f. For simple roots of f, the situation is similar to the degree one case in the proof of Budan–Fourier. The case of multiple roots can then be reduced to the simple root case.

Let $f \in \mathbb{R}[t]$ with Sturm sequence F. To prove the theorem, we consider $\operatorname{sgnvar}(F,t)$ as t increases from a to b (similarly to the proof of Theorem 2.12). The variation can only change, when t passes a real number, c, which is a root of some polynomial in the sequence F, as then the sign of said polynomial would change. Suppose throughout that c is a root of some polynomial in the sequence F. For any such c, we associate an $\varepsilon > 0$, such that $[c-\varepsilon, c+\varepsilon]$ contains no roots (other than c) of any polynomial in F.

We want to show the following: If $i > 0$, then the roots of f_i does not affect $\operatorname{sgnvar}(F,t)$, when t passes over them. However, when c is a root of $f_0 = f$, the variation decreases by exactly one, when t passes c.

We split the proof into three cases:

- Case 1: $f(c) \neq 0$
- Case 2: c is a simple root of f
- Case 3: c is a multiple root of f

Suppose $f_i(c) = f_{i+1}(c) = 0$ for some i. Then $f_{i-1}(c) = 0$, as $f_{i-1} = q_i f_i - f_{i+1}$ for some polynomial q_i (as in the Euclidean algorithm). Then the other polynomials in F will also vanish at c, by the same reasoning. In particular, $f(c) = f'(c) = 0$, so c is a multiple root of f. It follows that two subsequent polynomials in F cannot both vanish at c, unless we are in the third case.

Case 1: Suppose first that $f(c) \neq 0$. Then c must be a root of another polynomial in F. Suppose $f_i(c) = 0$ for some $i > 0$. As c is not a root of f, we must have $f_{i-1}(c), f_{i+1}(c) \neq 0$. We have that $f_{i-1}(c) = q_i(c) \cdot 0 - f_{i+1}(c) = -f_{i+1}(c)$, so $f_{i-1}(c)$ and $f_{i+1}(c)$ must have opposite signs. Therefore, $f_{i-1}(t)$ and $f_{i+1}(t)$ have opposite signs for any $t \in [c-\varepsilon, c+\varepsilon]$, as their signs cannot change in this interval. Hence, there is exactly one variation in sign coming from the subsequence $f_{i-1}(t), f_i(t), f_{i+1}(t)$ for all $t \in [c-\varepsilon, c+\varepsilon]$. So f_i vanishing at c has no effect on $\operatorname{sgnvar}(F,t)$, when t passes c. Note that this argument works for any Sylvester sequence, as we did not use that $f_1 = f'$.

Case 2: Next, suppose that c is a simple root of f. Then $f'(c) \neq 0$, so assume $f'(c) > 0$ (recall Remark 2.4). Hence, $f(t) < 0$ for $t \in [c-\varepsilon, c)$ and $f(t) > 0$ for $t \in (c, c+\varepsilon]$. Therefore, $\operatorname{sgnvar}(F,t)$ decreases by exactly one, when t passes c.

Case 3: Finally, we assume that c is a multiple root of f. Let $m+1$ be the multiplicity of c as a root of f. Then m is the multiplicity of c as a root of f'. It follows from the recursive definition of F that $(t-c)^m$ divides all polynomials in F. So $G = (g_0, \ldots, g_k) := \left(\frac{f}{(t-c)^m}, \frac{f'}{(t-c)^m}, \frac{f_2}{(t-c)^m}, \ldots, \frac{f_k}{(t-c)^m}\right)$ is a sequence of polynomials. By Remark 2.4, we have that $\operatorname{sgnvar}(G,t) = \operatorname{sgnvar}(F,t)$ for $t \neq c$. Note that G is, in particular, a Sylvester sequence. As the multiplicity of c as a root of f' was m, $g_1(c) \neq 0$. So we are in the first case with the sequence $(g_1, \ldots, g_k)$. As the argument also worked for any Sylvester sequence, there is no contribution to a change in variation from any g_i with $i > 0$. We now consider the contribution of g_0 to the change in variation as t passes c. Let h be a polynomial, such that $f(t) = (t-c)^{m+1}h(t)$. Clearly, $h(c) \neq 0$. Then

$$f'(t) = (m+1)(t-c)^m h(t) + (t-c)^{m+1}h'(t).$$

It follows that $g_0(t) = (t-c)h(t)$ and $g_1(t) = (m+1)h(t) + (t-c)h'(t)$. Assume that $h(c) > 0$ (recall again Remark 2.4). Note that we then also have $h(c-\varepsilon), h(c+\varepsilon) > 0$. After possibly adjusting ε to be smaller, we get that $g_1(c-\varepsilon), g_1(c), g_1(c+\varepsilon) > 0$, and $g_0(c-\varepsilon) < 0$, while $g_0(c+\varepsilon) > 0$. Hence, $\operatorname{sgnvar}(G,t)$ decreases by exactly 1, as t passes c. As $\operatorname{sgnvar}(G,t) = \operatorname{sgnvar}(F,t)$ for $t \in [c-\varepsilon, c+\varepsilon]\backslash\{c\}$, the same happens to $\operatorname{sgnvar}(F,t)$ as t passes c, i.e., when t passes a root of f.

So $\operatorname{sgnvar}(F,t)$ decreases by exactly one, as t passes a root of f, and otherwise it stays the same. Thus, the number of roots of f (counted without multiplicity) in the interval (a,b) is the difference $\operatorname{sgnvar}(F,a) - \operatorname{sgnvar}(F,b)$, which is exactly the statement of the theorem. ■

We are now finally able to properly describe the number of positive and negative roots of our running example.

Example 2.16 (Continuation of Example 2.13)**.** We again consider the polynomial

$$f = x^5 + 4x^4 + 2x^3 - 2x^2 + x - 6.$$

The Sturm sequence F of f is

$$\begin{aligned}
f_0 &= f,\\
f_1 &= f' = 5x^4 + 16x^3 + 6x^2 - 4x + 1,\\
f_2 &= \frac{44}{25}x^3 + \frac{54}{25}x^2 - \frac{36}{25}x + \frac{154}{25},\\
f_3 &= \frac{975}{484}x^2 + \frac{1625}{121}x + \frac{1475}{44},\\
f_4 &= -\frac{3872}{117}x - \frac{19360}{117},\\
f_5 &= -\frac{2025}{121}.
\end{aligned}$$

Hence, we get that

$$\begin{aligned}
\operatorname{sgnvar}(F,-\infty) &= \operatorname{sgnvar}\left(-1, 5, -\frac{44}{25}, \frac{975}{484}, \frac{3872}{117}, -\frac{2025}{121}\right) = 4,\\
\operatorname{sgnvar}(F,0) &= \operatorname{sgnvar}\left(-6, 1, \frac{154}{25}, \frac{1475}{44}, -\frac{19360}{117}, -\frac{2025}{121}\right) = 2,\\
\operatorname{sgnvar}(F,\infty) &= \operatorname{sgnvar}\left(1, 5, \frac{44}{25}, \frac{975}{484}, -\frac{3872}{117}, -\frac{2025}{121}\right) = 1.
\end{aligned}$$

By Sturm's Theorem, it follows that there is exactly three real roots — one positive and two negative. •

2.2 A characterization of real-rootedness: Hermite–Sylvester Criterion

We have explored bounds on the number of (positive) real solutions. But what conditions must be met for a polynomial to have all roots real? In this section, we will give a characterization of real-rooted polynomials (i.e., polynomials with all roots real). This characterization is known as the *Hermite–Sylvester Criterion*, and we state it in Theorem 2.21.

The concept of the discriminant is central to the characterization.

Definition 2.17. Let $f = \sum_{i=0}^{n} c_i x^i$ be a polynomial in $\mathbb{R}[x]$, where $c_n \neq 0$ and $n > 0$. The *Sylvester matrix* of f and its derivative is the $(n + (n-1)) \times (n + (n-1))$ matrix

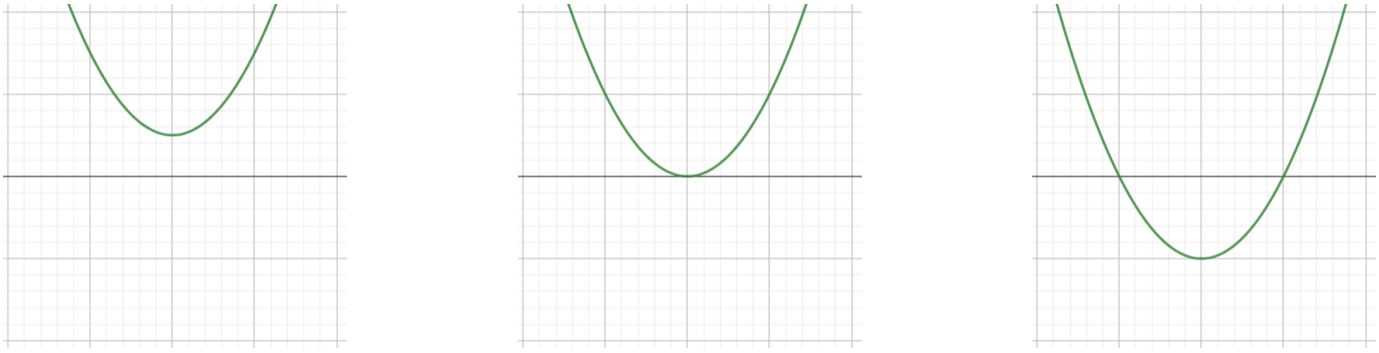
$$\mathrm{Syl}(f, f') = \begin{pmatrix} c_n & & & nc_n & & \\ c_{n-1} & \ddots & & (n-1)c_{n-1} & \ddots & \\ \vdots & & c_n & \vdots & & nc_n \\ c_1 & & c_{n-1} & 2c_2 & & (n-1)c_{n-1} \\ c_0 & & \vdots & c_1 & & \vdots \\ & \ddots & c_1 & & \ddots & 2c_2 \\ & & c_0 & & & c_1 \end{pmatrix},$$

where all empty spaces are filled with zeros. So the first $n-1$ columns are the coefficients of f, and the last n columns are the coefficients of f'. The *resultant* of f and its derivative is the determinant of the Sylvester matrix $\mathrm{Res}(f, f') = \det(\mathrm{Syl}(f, f'))$. The *discriminant* of f is

$$\mathrm{Disc}(f) = \frac{(-1)^{\frac{n(n-1)}{2}}}{c_n} \mathrm{Res}(f, f').$$

•

In Figure 2.3, we recall the geometric meaning of the discriminant for quadratic polynomials.



(a) $\mathrm{Disc}(f) < 0$ (b) $\mathrm{Disc}(f) = 0$ (c) $\mathrm{Disc}(f) > 0$

Figure 2.3: The discriminant of a quadratic polynomial f.

In general, the discriminant of a polynomial is zero, if the polynomial has any multiple roots. In the following example, we explore the connection between the discriminant of a cubic polynomial and the realness of its roots.

Example 2.18 (A geometric argument for real-rootedness of cubics)**.** Let $f(x) = x^3 + ax^2 + bx + c$ be a polynomial in $\mathbb{R}[x]$. Then $f'(x) = 3x^2 + 2ax + b$, and it has the roots

$$\alpha_i = \frac{-a + (-1)^i \sqrt{a^2 - 3b}}{3}, \ i = 1, 2.$$

Hence, f' has two distinct real roots, if and only if, $\mathrm{Disc}(f') = a^2 - 3b > 0$. We have that

$$f(\alpha_1) = \frac{1}{27}\left(2a^3 + 3a^2\sqrt{a^2 - 3b} - \sqrt{a^2 - 3b}^3 - 9b\sqrt{a^2 - 3b} - 9ab + 27c\right),$$
$$f(\alpha_2) = \frac{1}{27}\left(2a^3 - 3a^2\sqrt{a^2 - 3b} + \sqrt{a^2 - 3b}^3 + 9b\sqrt{a^2 - 3b} - 9ab + 27c\right).$$

Thus, we get that

$$f(\alpha_1)f(\alpha_2) = -\frac{1}{27}b^2(a^2 - 4b) + \frac{2}{27}ac(2a^2 - 9b) + c^2.$$

Note that

$$\begin{aligned} -27f(\alpha_1)f(\alpha_2) &= b^2(a^2 - 4b) - 2ac(2a^2 - 9b) - 27c^2 \\ &= a^2b^2 - 4b^3 - 4a^3c + 18abc - 27c^2. \end{aligned}$$

The discriminant of a cubic polynomial $f = a_3x^3 + a_2x^2 + a_1x + a_0$ is

$$\text{Disc}(f) = a_2^2a_1^2 - 4a_3a_1^3 - 4a_2^3a_0 - 27a_3^2a_0^2 + 18a_3a_2a_1a_0.$$

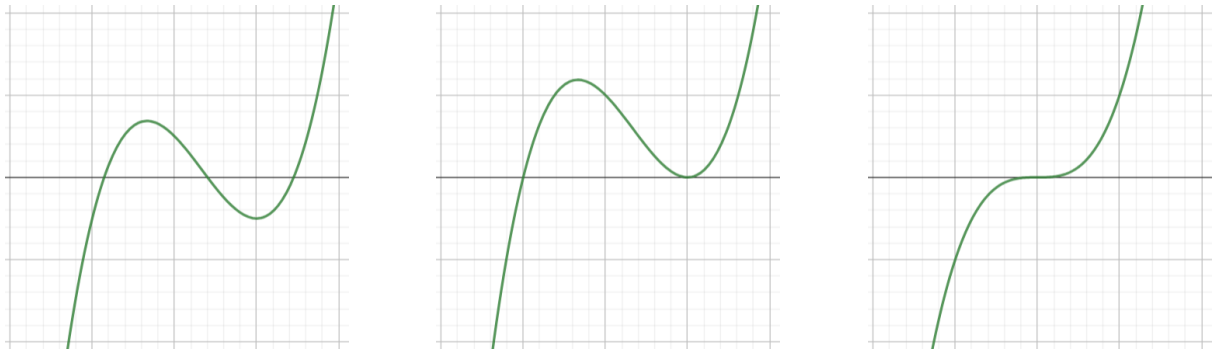
Hence, the discriminant of f is

$$\text{Disc}(f) = a^2b^2 - 4b^3 - 4a^3c - 27c^2 + 18abc.$$

This gives us the relation

$$f(\alpha_1)f(\alpha_2) = -\frac{\text{Disc}(f)}{27}.$$

Hence, $\text{Disc}(f) > 0$, if and only if, $f(\alpha_1)f(\alpha_2) < 0$.



(a) $f(\alpha_1) > 0, f(\alpha_2) < 0$ (b) $f(\alpha_1) > 0, f(\alpha_2) = 0$ (c) $f(\alpha_1) = f(\alpha_2) = 0$

Figure 2.4: Possibile configurations of the real roots f, where α_1, α_2 are the roots of the derivative.

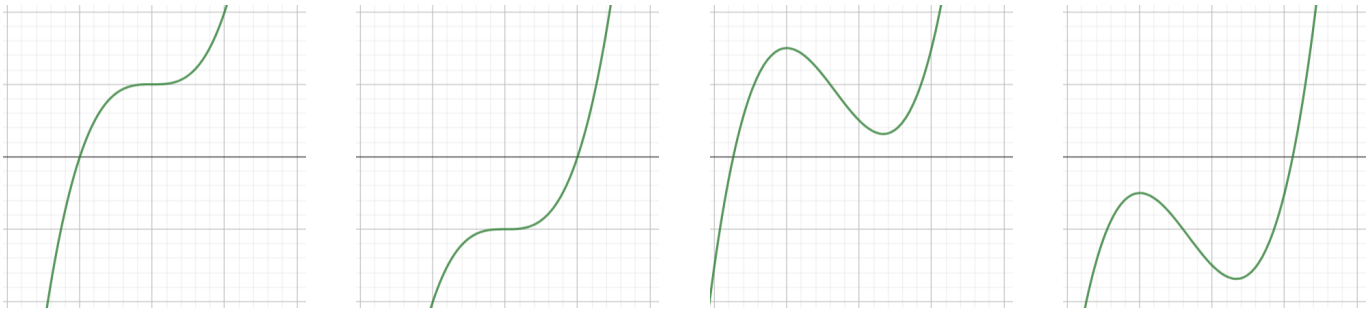
Our goal is to find a characterization of real-rootedness, which is based on the discriminants of f and f'. It will turn out that f has three distinct real roots, if and only if, $\text{Disc}(f) > 0$ and $\text{Disc}(f') > 0$.

$\underline{\text{Disc}(f')}$:

- If all roots of f are real, then there are three possible configurations of the roots as seen in Figure 2.4. Note that $\text{Disc}(f') > 0$ is equivalent to f' having two distinct real roots, so this rules out the configuration in Figure 2.4c. In Figure 2.4b, we get $f(\alpha_1)f(\alpha_2) = 0$, so $\text{Disc}(f) > 0$ rules out that configuration. The configuration in Figure 2.4a satisfies both the conditions listed.
- In Figure 2.5 we see the possible configurations of roots, if f has a non-real root. The first two cases (Figures 2.5a and 2.5b) are somewhat deceptive in that the inflection point has vanishing derivative, which is not always the case for the conditions stated. There

only being one inflection point means that either f' has only one real root (of multiplicity two) or it has a pair of non-real roots. If it has a single real root of multiplicity two, then $\mathrm{Disc}(f') = 0$. If it has a pair of non-real roots, then $\mathrm{Disc}(f') < 0$. Hence, $\mathrm{Disc}(f') > 0$ rules out the first two cases in Figure 2.5.

$\mathrm{Disc}(f)$: In both Figures 2.5c and 2.5d, we have that $f(\alpha_1)f(\alpha_2) > 0$. Hence, $\mathrm{Disc}(f) > 0$ rules out these two cases.



(a) $f(\alpha_1) = f(\alpha_2) > 0$ (b) $f(\alpha_1) = f(\alpha_2) < 0$ (c) $f(\alpha_1), f(\alpha_2) > 0$ (d) $f(\alpha_1), f(\alpha_2) < 0$

Figure 2.5: The possible configurations of non-real roots of f, where α_1, α_2 are the roots of the derivative.

It follows that f has three distinct real roots, if and only if, $\mathrm{Disc}(f) > 0$ and $\mathrm{Disc}(f') > 0$. •

We now move to the general setting. Let $f(x) = x^n + \cdots + a_1 x + a_0$ be a polynomial over $\mathbb{C}$. Assume that $\{1, \ldots, x^{n-1}\}$ is a basis of $\mathbb{C}[x]/(f)$. We consider the multiplication map

$$\begin{aligned} M_k \colon \mathbb{C}[x]/(f) &\xrightarrow{\cdot x^k} \mathbb{C}[x]/(f) \\ g &\mapsto g \cdot x^k. \end{aligned}$$

Hence, multiplication by x^k is equivalent to simply multiplying our coefficient matrix by M_k. It is easy to see that M_1 can be represented by the companion matrix of f, i.e.,

$$M_1 = \begin{pmatrix} 0 & & 0 & -a_0 \\ 1 & \ddots & & -a_1 \\ & \ddots & 0 & \vdots \\ 0 & & 1 & -a_{n-1} \end{pmatrix}.$$

Note that $M_k = M_1^k$. Let us first explore this definitions for quadratic polynomials.

Example 2.19. Let $f(x) = x^2 + ax + b$ be a real polynomial. Then we get that $M_0 = I_2$,

$$\begin{aligned} M_1 &= \begin{pmatrix} 0 & -b \\ 1 & -a \end{pmatrix}, \\ M_2 &= \begin{pmatrix} 0 & -b \\ 1 & -a \end{pmatrix}^2 = \begin{pmatrix} -b & ab \\ -a & a^2 - b \end{pmatrix}. \end{aligned}$$

Hence, we get that $\mathrm{Tr}(M_0) = 2$, $\mathrm{Tr}(M_1) = -a$, and $\mathrm{Tr}(M_2) = a^2 - 2b$. Let

$$H = \begin{pmatrix} \mathrm{Tr}(M_0) & \mathrm{Tr}(M_1) \\ \mathrm{Tr}(M_1) & \mathrm{Tr}(M_2) \end{pmatrix} = \begin{pmatrix} 2 & -a \\ -a & a^2 - 2b \end{pmatrix}.$$

Note that the determinant of H,

$$\det(H) = 2a^2 - 4b - a^2 = a^2 - 4b,$$

is exactly the discriminant of f. We know that f has two distinct real roots, if and only if, the discriminant is strictly positive. Hence, $\det(H) > 0$ is equivalent to f having two distinct real roots. •

The matrix H from Example 2.19 is called the *Hermite matrix* of f.

Definition 2.20. The *Hermite matrix* of $f \in \mathbb{C}[x]$ of degree n is $H = (\mathrm{Tr}(M_{i+j-2}))_{i,j=1,\dots,n}$. •

Note that the Hermite matrix is symmetric. This definition allows us to characterize real-rooted polynomials (see [Nat19] for a proof).

Theorem 2.21 (Hermite–Sylvester Criterion)**.** *The Hermite matrix associated to $f \in \mathbb{C}[x]$ of degree n is positive definite, if and only if, all roots of f are real and distinct.*

The following example characterizes the situation in degree three.

Example 2.22. Let $f(x) = x^3 + ax^2 + bx + c$ be a real polynomial. We find that $M_0 = I_3$,

$$M_1 = \begin{pmatrix} 0 & 0 & -c \\ 1 & 0 & -b \\ 0 & 1 & -a \end{pmatrix},$$

$$M_2 = M_1^2 = \begin{pmatrix} 0 & -c & ac \\ 0 & -b & ab - c \\ 1 & -a & a^2 - b \end{pmatrix},$$

$$M_3 = M_1^3 = \begin{pmatrix} -c & ac & bc - a^2c \\ -b & ab - c & b^2 + ac - a^2b \\ -a & a^2 - b & 2ab - c - a^3 \end{pmatrix},$$

$$M_4 = M_1^4 = \begin{pmatrix} ac & bc - a^2c & a^3c - 2abc + c^2 \\ ab - c & b^2 + ac - a^2b & a^3b - a^2c - 2ab^2 + 2bc \\ a^2 - b & 2ab - c - a^3 & a^4 + b^2 - 3a^2b + 2ac \end{pmatrix}.$$

It follows that $\mathrm{Tr}(M_0) = 3$, $\mathrm{Tr}(M_1) = -a$, $\mathrm{Tr}(M_2) = a^2 - 2b$, $\mathrm{Tr}(M_3) = -a^3 + 3ab - 3c$, and $\mathrm{Tr}(M_4) = a^4 + 2b^2 + 4ac - 4a^2b$. From this we get the Hermite matrix associated to f to be

$$H = \begin{pmatrix} 3 & -a & a^2 - 2b \\ -a & a^2 - 2b & -a^3 + 3ab - 3c \\ a^2 - 2b & -a^3 + 3ab - 3c & a^4 + 2b^2 + 4ac - 4a^2b \end{pmatrix}.$$

We find that

$$\det(H) = a^2b^2 - 4a^3c - 4b^3 - 27c^2 + 18abc,$$

and the 2×2 leading principal minor is

$$\det(H(3,3)) = 3a^2 - 6b - a^2 = 2a^2 - 6b = 2(a^2 - 3b).$$

Note that $\det(H) = \mathrm{Disc}(f)$, and $\det(H(3,3)) \propto \mathrm{Disc}(f')$. So H is positive definite, if and only if, $\mathrm{Disc}(f) > 0$ and $\mathrm{Disc}(f') > 0$. By Example 2.18, it follows that H being positive definite is equivalent to f having 3 distinct real roots. •

We can actually say more. The *signature* of a symmetric matrix $M \in \mathbb{R}^{n \times n}$ is defined to be

$$\text{signature}(M) = \#\text{positive eigenvalues} - \#\text{negative eigenvalues}.$$

Hence, $\text{signature}(M) = n$, if and only if, M is positive definite. So the above criterion is equivalent to the condition $\text{signature}(H) = n$, and an even stronger result holds.

Proposition 2.23. *Let H be the Hermite matrix associated to f. Then f has* $\text{signature}(H)$ *distinct real roots.*

2.3 Conditions for real-rootedness

While it is good that the Hermite–Sylvester Criterion is an "if and only if"-statement, it can be computationally cumbersome to verify the condition given in the criterion. Therefore, in this section, we consider simpler conditions of real-rootedness.

Newton's inequalities give a necessary condition for a polynomial to have all roots real.

Theorem 2.24 (Newton's inequalities)**.** *Let $f = a_n x^n + \cdots + a_1 x + a_0$ be a real polynomial. If f is real-rooted, then*

$$a_i^2 - \frac{i+1}{i}\frac{n+1-i}{n-i} a_{i-1} a_{i+1} \geq 0, \quad i = 1, \ldots, n-1.$$

While this condition might seem slightly complicated at first glance, the case for quadratic polynomials is well-known already in high school.

Example 2.25. For $n = 2$, this is $a_1^2 - 4a_0a_2 \geq 0$, which is just the usual discriminant condition. For $n = 3$, the inequalities become

$$\begin{aligned} a_1^2 - 3a_0a_2 &\geq 0, \\ a_2^2 - 3a_1a_3 &\geq 0. \end{aligned}$$

Consider the polynomial $f(x) = x^3 + 2x^2 + x - 1$, which satisfies both inequalities. We have that $\mathrm{Disc}(f) = -31 < 0$, and $\mathrm{Disc}(f') = 1 > 0$. It follows, from Example 2.18, that f has two non-real roots. Thus, we see that the converse of Theorem 2.24 does not hold. •

We want something, which is close to being a converse of Newton's inequalities. This leads us to the following theorem due to Kurtz, which instead gives a sufficient condition for all roots to be real.

Theorem 2.26 ([Kur92, Theorem 1])**.** *Let $f = \sum_{i=0}^{n} a_i x^i$ be a polynomial of degree $n \geq 2$ with positive coefficients. If*

$$a_i^2 - 4a_{i-1}a_{i+1} > 0, \quad i = 1, 2, \ldots, n-1, \tag{2.3}$$

then all the roots of f are real and distinct.

Example 2.27. For $n = 2$, Equation (2.3) is simply the usual discriminant condition. For $n = 3$, the inequalities from the theorem are $a_1^2 - 4a_2a_0 > 0$, and $a_2^2 - 4a_3a_1 > 0$. Thus, for $n > 2$, the the coefficient in Equation (2.3) is no longer the same as the coefficient in Theorem 2.24. •

To prove Theorem 2.26, we will need the following technical lemma.

Lemma 2.28 ([Kur92, Lemma 1])**.** *Let $f = \sum_{i=0}^{n} a_i x^i$ be a polynomial of degree $n \geq 2$ with real coefficients. Assume that all roots of f have negative real part. If f has a multiple real root, then there exists some $i \in \{1, \ldots, n-1\}$, such that*

$$a_i^2 - 4a_{i-1}a_{i+1} \leq 0.$$

Proof. We proceed by induction on n. For $n = 2$, the result is immediate. Suppose $n > 2$, and that f has a repeated root $-a \in \mathbb{R}_{<0}$. It follows that we can write f on the form

$$f(x) = (x + a)^2(b_{n-2}x^{n-2} + \cdots + b_0)$$

for some $b_i \in \mathbb{R}$. As all roots of f are in the left half-plane, all b_i are positive. We can rewrite f as

$$f(x) = \sum_{i=0}^{n} \left(b_{i-2} + 2ab_{i-1} + a^2b_i\right) x^i,$$

where $b_i = 0$ for all $i < 0$ and $i > n - 2$.

Assume, seeking a contradiction, that $a_k^2 - 4a_{k-1}a_{k+1} > 0$ for all $k \in \{1, \ldots, n-1\}$. I.e., for all $k \in \{1, \ldots, n-1\}$, we have that

$$\begin{aligned} 0 < a_k^2 - 4a_{k-1}a_{k+1} &= \left(b_{k-2} + 2ab_{k-1} + a^2b_k\right)^2 - 4\left(b_{k-3} + 2ab_{k-2} + a^2b_{k-1}\right)\left(b_{k-1} + 2ab_k + a^2b_{k+1}\right) \\ &= -4a^4b_{k+1}b_{k-1} + a^4b_k^2 - 8a^3b_{k+1}b_{k-2} - 4a^3b_{k-1}b_k - 4a^2b_{k-3}b_{k+1} \\ &\quad - 14a^2b_{k-2}b_k - 8ab_{k-3}b_k - 4ab_{k-2}b_{k-1} - 4b_{k-3}b_{k-1} + b_{k-2}^2. \end{aligned}$$

As $a > 0$ and $b_i > 0$ for all i, we must also have that

$$a^4b_k^2 + b_{k-2}^2 - 4a^3b_{k-1}b_k - 4ab_{k-2}b_{k-1} > 0.$$

For each k, let

$$\begin{aligned} q_k &= ab_k - 4b_{k-1} \\ r_k &= b_{k-1} - 4ab_k. \end{aligned}$$

Then

$$0 < a^4b_k^2 + b_{k-2}^2 - 4a^3b_{k-1}b_k - 4ab_{k-2}b_{k-1} = a^3b_kq_k + b_{k-2}r_{k-1} \tag{2.4}$$

for all $k \in \{1, \ldots, n-1\}$. So $0 < a^3b_1q_1 + b_{1-2}r_{1-1} = a^3b_1q_1 + b_{-1}r_0 = a^3b_1q_1$. Hence, we must have $q_1 > 0$, which implies that $r_1 < 0$. Additionally, $0 < a^3b_{n-1}q_{n-1} + b_{n-1-2}r_{n-1-1} = b_{n-3}r_{n-2}$, so $r_{n-2} > 0$.

Let ν be the smallest integer such that $r_\nu > 0$. From the above, we get that $1 < \nu < n - 1$. By Equation (2.4), we have that

$$a^3b_\nu q_\nu + b_{\nu-2}r_{\nu-1} > 0.$$

As $r_{\nu-1} < 0$, we must have $q_\nu > 0$. But then $r_\nu < 0$, which is a contradiction. Hence, there must exist some k, such that $a_k^2 - 4a_{k-1}a_{k+1} \leq 0$. ■

We are now ready to prove Theorem 2.26.

Proof of Theorem 2.26. We proceed by induction on n. For $n = 2$, the result is well-known. Let $n > 2$, and suppose the theorem holds for $n - 1$.

Let $f(x) = a_n x^n + \cdots + a_0$ with positive coefficients satisfying Equation (2.3). We define

$$q(x) := f(x) - a_0 = x(a_n x^{n-1} + \cdots + a_1),$$

and let $r(x) := a_n x^{n-1} + \cdots + a_1$. So r is of degree $n - 1 \geq 2$, has positive coefficients, and satisfies Equation (2.3). It follows, from the induction hypothesis, that r has $n - 1$ distinct real roots. By Remark 2.10, all roots of r are negative. As $q(x) = xr(x)$, q has n distinct real non-positive roots — including a root at zero. Thus, zero is the largest root of q.

Let $q_\lambda := q + \lambda$, where $0 \leq \lambda$, and denote by $N(\lambda)$ the number of distinct real roots of q_λ. Clearly, $N(0) = n$. We set

$$S := \{\lambda \mid \lambda > 0, N(\lambda) < n\},$$

which is clearly non-empty and bounded below by 0. Let $\lambda_0 := \inf S$ denote the greatest lower bound of S. If $\lambda_0 > a_0$, then $a_0 \notin S$. Hence, $N(a_0) \not< n$ (as $a_0 > 0$), so we must have $N(a_0) = n$. As $q_{a_0} = f$, we are done. So suppose, seeking a contradiction, that $\lambda_0 \leq a_0$.

It is well-known that the roots of a polynomial vary continuously as its coefficients vary.

As q_0 has n distinct real roots, this continuity implies that there exists $\varepsilon > 0$, such that for all $\lambda \in (0, \varepsilon)$, we have $N(\lambda) = n$, so $(0, \varepsilon) \cap S = \emptyset$, i.e., we must have $\lambda_0 > 0$. By the same argument, if $N(\lambda_0) = n$, then there exists $\varepsilon > 0$, such that for all $\lambda \in (\lambda_0, \lambda_0 + \varepsilon)$, we have $N(\lambda) = n$, so $(\lambda_0, \lambda_0 + \varepsilon) \cap S = \emptyset$, which would imply $\inf S \geq \lambda_0 + \frac{\varepsilon}{2} > \lambda_0$. Hence, we must have $N(\lambda_0) < n$.

It follows that q_{λ_0} has either some non-real root or a repeated real root.

Suppose first that q_{λ_0} has a non-real root. By the continuity of the roots, there exists an $\varepsilon > 0$, such that $0 < \lambda_0 - \varepsilon$, and $q_{\lambda_0 - \varepsilon}$ also has some non-real roots. But then $\lambda_0 - \varepsilon \in S$, so $\inf S \leq \lambda_0 - \varepsilon < \lambda_0$, which is a contradiction.

Suppose instead that q_{λ_0} has a repeated real root. We know that all roots of q are real and non-positive. As the constant term increasing does not change the sign of the roots (because all coefficients are positive), q_{λ_0} has all roots negative. Let a'_i denote the coefficients of q_{λ_0}. Then $a'_i = a_i$ for all $i \in \{1, \ldots, n\}$, but $a'_0 = \lambda_0$. By assumption on f, we have that $a_i^2 - 4a_{i-1}a_{i+1} > 0$ for all $i \in \{1, \ldots, n-1\}$. As we assumed $a_0 \geq \lambda$, we also find that $a_1^2 - 4\lambda_0 a_2 \geq a_1^2 - 4a_0 a_2 > 0$. Thus, q_{λ_0} satisfies the condition from Equation (2.3). As q_{λ_0} has a repeated real root, it follows, from Lemma 2.28, that there exists some $i \in \{1, \ldots, n-1\}$, such that $a_i'^2 - 4a'_{i-1}a'_{i+1} \leq 0$, which is a contradiction.

Thus, we reach a contradiction in both cases, so we must have $a_0 < \lambda_0$. It follows that f has n distinct real roots. ■

While Kurtz only proves the above result in the case of positive coefficients, we can extend the result to also include negative coefficients and coefficients alternating in sign.

Proposition 2.29. *Theorem 2.26 also holds, when the coefficients are negative or alternate in sign.*

Proof. Follows by applying Theorem 2.26 to $-f(x)$ and $f(-x)$, and observing that these operations do not change $a_i^2 - 4a_{i-1}a_{i+1}$. ■

Just like with Newton's inequalities, the converse of Theorem 2.26 does not hold.

Example 2.30. Consider the polynomial $p = x^3 + 6x^2 + 11x + 6$, which has the roots -1, -2 and -3. Both $a_1^2 - 4a_2a_0 < 0$, and $a_2^2 - 4a_3a_1 < 0$, so it does not satisfy the conditions of Theorem 2.26. Thus, the converse of Theorem 2.26 does not hold. •

Inspired by Newton's inequalities, one could hope that we could find some coefficient lying between 3 and 4, such that the condition from Equation (2.3) became both necessary and sufficient — at least for cubic polynomials. However, the following theorem shows that this is not possible.

Theorem 2.31 ([Kur92, Theorem 2])**.** *Given $\varepsilon > 0$ and an integer $n \geq 2$, there is a polynomial with positive coefficients of degree n, which has some non-real roots and whose coefficients, $a_0, \ldots, a_n$, satisfy*

$$a_i^2 - (4 - \varepsilon)a_{i-1}a_{i+1} > 0, \quad 1 \leq i \leq n-1. \tag{2.5}$$

Proof. Let $f(x) = \sum_{i=0}^{n} a_i x^i$ and define

$$S(f, i) = \frac{a_i^2}{a_{i-1}a_{i+1}}, \quad i = 1, \ldots, n-1.$$

So $S(f, i) > 4 - \varepsilon$ is equivalent to $a_i^2 - (4 - \varepsilon)a_{i-1}a_{i+1} > 0$.

We proceed by induction on $n \geq 2$. First, suppose $n = 2$, and fix $\varepsilon > 0$. Take $f = x^2 + x + a_0$ with $\frac{1}{4} < a_0 < \frac{1}{4-\varepsilon}$, then $4 - \varepsilon < S(f, 1) < 4$.

Now let $n > 2$ and fix $\varepsilon > 0$. Consider a polynomial $f_{n-1}(x) = \sum_{i=0}^{n-1} a_i x^i$, which satisfies $S(f_{n-1}, i) > 4 - \frac{\varepsilon}{2}$ for all $i = 1, \ldots, n-2$.

Define $f_\mu(x) = (\mu x + 1) f_{n-1}(x)$ for some $\mu > 0$. Then f_μ is of degree n. Note that the roots of f_{n-1} are also roots of f_μ. In particular, f_μ has some non-real roots. Note that the coefficients of f_μ are all positive. We have left to show that we can choose μ, such +that $S(f_\mu, i) > 4 - \varepsilon$ for all $i = 1, \ldots, n-1$. We can rewrite f_μ as

$$f_\mu(x) = \mu a_{n-1}x^n + (\mu a_{n-2} + a_{n-1})x^{n-1} + \cdots + (\mu a_1 + a_2)x^2 + (\mu a_0 + a_1)x + a_0.$$

Hence, we have that

$$\begin{aligned}
S(f_\mu, 1) &= \frac{(\mu a_0 + a_1)^2}{a_0(\mu a_1 + a_2)} = \frac{\mu^2 a_0^2 + \mu 2 a_0 a_1 + a_1^2}{\mu a_0 a_1 + a_0 a_2} = \frac{\mu a_0^2 + 2a_0 a_1 + \mu^{-1} a_1^2}{a_0 a_1 + \mu^{-1}} \overset{\mu\to\infty}{\to} \infty \\
S(f_\mu, n-1) &= \frac{(\mu a_{n-2} + a_{n-1})^2}{\mu a_{n-1}(\mu a_{n-3} + a_{n-2})} = \frac{\mu^2 a_{n-2}^2 + \mu 2 a_{n-2}a_{n-1} + a_{n-1}^2}{\mu^2 a_{n-3}a_{n-1} + \mu a_{n-2}a_{n-1}} \\
&= \frac{a_{n-2}^2 + \mu^{-1} 2 a_{n-2}a_{n-1} + \mu^{-2} a_{n-1}^2}{a_{n-3}a_{n-1} + \mu^{-1} a_{n-2}a_{n-1}} \overset{\mu\to\infty}{\to} \frac{a_{n-2}^2}{a_{n-3}a_{n-1}} = S(f_\mu, n-2),
\end{aligned}$$

and for $i = 2, \ldots, n-2$, we get that

$$\begin{aligned}
S(f_\mu, i) &= \frac{(\mu a_{i-1} + a_i)^2}{(\mu a_{i-2} + a_{i-1})(\mu a_i + a_{i+1})} = \frac{\mu^2 a_{i-1}^2 + \mu 2 a_{i-1}a_i + a_i^2}{\mu^2 a_{i-2}a_i + \mu a_{i-2}a_{i+1} + \mu a_{i-1}a_i + a_{i-1}a_{i+1}} \\
&= \frac{a_{i-1}^2 + \mu^{-1} 2 a_{i-1}a_i + \mu^{-2} a_i^2}{a_{i-2}a_i + \mu^{-1} a_{i-2}a_{i+1} + \mu^{-1} a_{i-1}a_i + \mu^{-2} a_{i-1}a_{i+1}} \overset{\mu\to\infty}{\to} \frac{a_{i-1}^2}{a_{i-2}a_i} = S(f_\mu, i-1).
\end{aligned}$$

It follows that we can choose μ large enough that $S(f_\mu, i) > 4 - \varepsilon$ for all $i = 1, \ldots, n-1$. ■

We close the chapter on univariate polynomials with a computer experiment and an implementation related to the theorems of this section.

2.3.1 Experiment: How often is Theorem 2.26 satisfied?

We know from Example 2.30 that the converse of Theorem 2.26 does not hold. But if we are given a real-rooted polynomial with positive coefficients, then how often will the condition from Equation (2.3) be satisfied? As the condition from Equation (2.3) is always satisfied for $n = 2$, it would be intuitive to think that the condition is also often satisfied for $n > 2$. This, however, turns out not to be the case.

We investigate this by testing the conditions of Theorem 2.26 on polynomials with real roots chosen uniformly in an interval. If the coefficients are positive, then any real root must be negative, by Remark 2.10. We consider different intervals of the forms $(-n, 0)$ and $(-n, -1)$. All experiments are run with 100,000,000 samples. The results are listed as the ratios between the number of instances satisfying the conditions and the sample size. The code can be found at [Kal25].

Interval	Degree 3	Degree 4	Degree 5	Degree 6
$(-100, 0)$	0.11938581	0.00486293	$5.144 \times e^{-5}$	$8.0 \times e^{-8}$
$(-1000, 0)$	0.11937622	0.00485887	$5.231 \times e^{-5}$	$1.6 \times e^{-7}$
$(-10000, 0)$	0.11934174	0.00487097	$5.08 \times e^{-5}$	$1.3 \times e^{-7}$
$(-100000, 0)$	0.11938009	0.00485577	$5.278 \times e^{-5}$	$1.6 \times e^{-7}$
$(-1000000, 0)$	0.11938332	0.00487046	$5.13 \times e^{-5}$	$1.2 \times e^{-7}$
$(-100, -1)$	0.09765286	0.00157997	0.0	0.0
$(-1000, -1)$	0.11685948	0.00442414	$3.061 \times e^{-5}$	$3.0 \times e^{-8}$
$(-10000, -1)$	0.11908152	0.00481732	$4.943 \times e^{-5}$	$9.0 \times e^{-8}$
$(-100000, -1)$	0.1193003	0.00485834	$5.111 \times e^{-5}$	$1.3 \times e^{-7}$
$(-1000000, -1)$	0.11941458	0.00486441	$5.016 \times e^{-5}$	$1.4 \times e^{-7}$

Table 2.5

The results seem to differ quite a lot depending on whether $(-1, 0)$ is included in the interval. However, it seems there is a tendency for the ratios to increase, when the intervals become bigger. This is much more pronounced for the intervals of the form $(-n, -1)$, and it is also much clearer in the higher degree cases. But why does the ratio increase, when the interval becomes bigger? What happens in the interval $(-1, 0)$? We investigate the situation in the degree three case.

Example 2.32. For $n = 3$, the inequalities from Theorem 2.26 are

$$\begin{aligned} a_1^2 - 4a_2a_0 > 0, \\ a_2^2 - 4a_3a_1 > 0. \end{aligned} \tag{2.6}$$

Consider $f(x) = (x - r_1)(x - r_2)(x - r_3)$, where $r_1, r_2, r_3 \in \mathbb{R}_{<0}$. Then

$$f(x) = x^3 + (-r_1 - r_2 - r_3)x^2 + (r_1r_2 + r_1r_3 + r_2r_3)x + (-r_1r_2r_3).$$

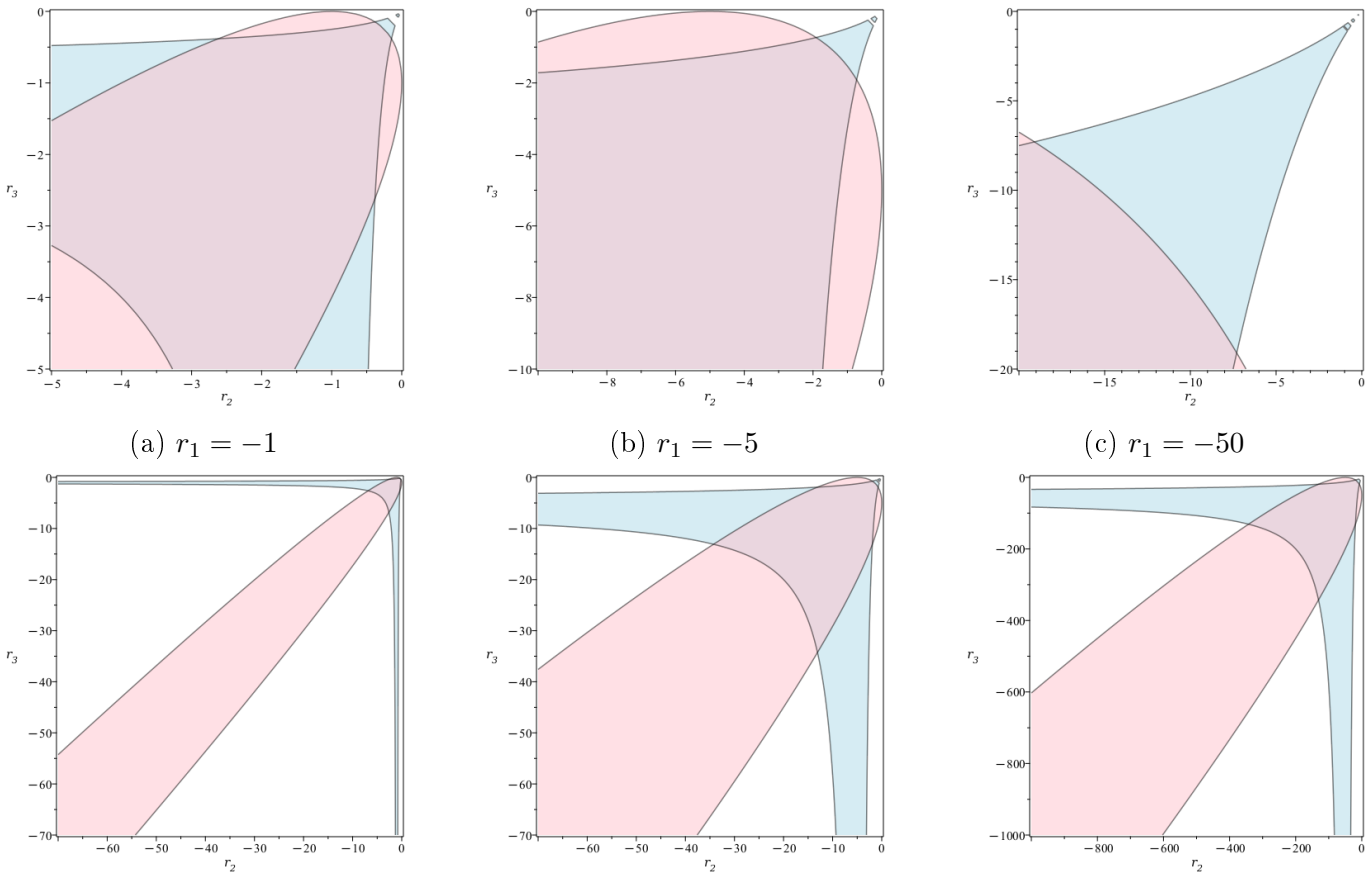
So, if we write $f(x) = \sum_{i=0}^{3} a_i x^i$, the coefficients are $a_0 = -r_1 r_2 r_3$, $a_1 = r_1 r_2 + r_1 r_3 + r_2 r_3$, $a_2 = -r_1 - r_2 - r_3$, and $a_3 = 1$. Then the inequalities from Equation (2.6) are

$$0 < a_1^2 - 4a_2 a_0 = r_1^2 r_2^2 + r_1^2 r_3^2 + r_2^2 r_3^2 - 2r_1^2 r_2 r_3 - 2r_1 r_2^2 r_3 - 2r_1 r_2 r_3^2, \tag{2.7}$$

$$0 < a_2^2 - 4a_3 a_1 = r_1^2 + r_2^2 + r_3^2 - 2r_1 r_2 - 2r_1 r_3 - 2r_2 r_3. \tag{2.8}$$

In Figure 2.6, we consider specific values of r_1 and plot the inequalities. These illustrations give us a hint to explain the results of the experiment. From the figure, we see that at least one root must be relative small compared to the other two. If we choose (r_2, r_3) to be a point below/to the left of the blue area, then r_1 will be small relative to the others. If we choose (r_2, r_3) to be a point above/to the right of the blue area, then either r_2 or r_3 (or both) will be small relatively to r_1. Thus, the roots cannot all be close together, which explains why we get a higher ratio, when we have bigger intervals.

We also notice that close to the origin (compared to the chosen value of r_1), the situation looks quite different, than it does in the big picture. This also explains why our results changed, when we changed the end point from 0 to -1. For a smaller interval, the change is more significant. This is of course due to the size of the area "close to" the origin becoming much larger, when r_1 is bigger. So this change of endpoint does not have much effect, when choosing one root to be very big.



(a) $r_1 = -1$ (b) $r_1 = -5$ (c) $r_1 = -50$

(d) $r_1 = -1$ (e) $r_1 = -5$ (f) $r_1 = -50$

Figure 2.6: The light blue (resp. pink) area is where Equation (2.7) (resp. Equation (2.8)) does *not* hold. The white area is where both inequalities hold.

•

2.3.2 Implementation: Finding examples based on Theorem 2.31

As the proof of Theorem 2.31 is constructive, we can implement an algorithm to find instances of the examples described in Theorem 2.31. The code can be found at [Kal25].

Let $\varepsilon > 0$ and $n \geq 2$ be given. We will build an example recursively from the degree two case, by following the proof. For $n = 2$, we let $f = x^2 + x + a_0$, where for $\varepsilon < 4$, we pick a_0 to be the midpoint in the interval $\left(\frac{1}{4}, \frac{1}{4-\varepsilon}\right)$, and for $\varepsilon \geq 4$, we just need $\frac{1}{4} < a_0$, so we pick $a_0 = 1$.

If we run the algorithm with $\varepsilon = \frac{1}{2}$ and $n = 5$, we get the following output:

$$2086080x^5 + 2155756x^4 + [595960.047619047619 \pm 1.89e^{-13}]x^3 + [18183.2460317460317 \pm 5.07e^{-14}]x^2 + [158.4900793650793651 \pm 6.14e^{-17}]x + [0.2519841269841269841 \pm 5.24e^{-20}]$$

The output involves intervals due to the use of the BigFloat datatype, so these intervals are, in fact, error bounds. However, after testing, we find that

$$2086080x^5 + 2155756x^4 + 595960.047619047619x^3 + 18183.2460317460317x^2 + 158.4900793650793651x + 0.2519841269841269841$$

also satisfies all the conditions. We can go further and round the coefficients to the nearest integer (except for the constant term, which we round to two decimal places)

$$2086080x^5 + 2155756x^4 + 595960x^3 + 18183x^2 + 158x + 0.25,$$

and it still satisfies the conditions from the theorem. Trivially, the polynomial (obtained by multiplying by four)

$$8344320x^5 + 8623024x^4 + 2383840x^3 + 72732x^2 + 632x + 1$$

will also satisfy the desired conditions. Hence, we have found an integral polynomial of degree five, which satisfies all conditions of Theorem 2.31 with $\varepsilon = \frac{1}{2}$.

3 Bounds on the number of complex roots in the multivariate case

To generalize the bounds on the number of real roots from the univariate case to the multivariate case, it is helpful to first find generalizations of the Fundamental Theorem of Algebra, i.e., generalize the bound on the number of complex roots. The most immediate generalization to the multivariate situation is Bézout's Theorem.

Theorem 3.1 (Bézout, [Bé79]). *Given a system of n polynomials in n variables with total degrees $d_1, \ldots, d_n$, the system has at most $d_1 \cdots d_n$ isolated solutions in $\mathbb{C}^n$.*

While Bézout's Theorem does give a bound on the number of complex solutions, other results can improve this bound — especially for sparse polynomials. But what other possibilities may we consider? Let f be a univariate polynomial of degree n with nonzero constant term. Then $\mathrm{NP}(f)$ is the line segment from zero to n, so $\mathrm{vol}(\mathrm{NP}(f)) = n = \deg(f)$. We see that, in the univariate case, the volume of the Newton polytope and the degree of the polynomial are equal (when the constant term is nonzero). With this in mind, we devote the chapter to bounds based on the volume of the Newton polytope associated to the system of polynomials.

Let $\mathcal{A} = \{a_0, \ldots, a_r\} \subset \mathbb{Z}^n$ be an ordered set of integer vectors, which affinely spans $\mathbb{R}^n$. We let $C \in \mathbb{C}^{n \times r}$ be some coefficient matrix. For any such matrix,

$$\begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = C \cdot x^A = 0 \tag{3.1}$$

is the corresponding sparse polynomial system in n variables $x = (x_1, \ldots, x_n)$ with support $\mathcal{A}$. The results in this chapter apply to all polynomials in $\mathbb{C}[x_1, \ldots, x_n]$, so we do not restrict the coefficients to $\mathbb{R}$. As $\mathrm{NP}(f_i) = \mathrm{conv}(\mathcal{A})$ for all $i = 1, \ldots, n$, we denote this polytope by $\mathrm{NP}(\mathcal{A})$. Later, we extend to the case of systems of so-called *mixed support*, but for now we consider non-mixed systems.

3.1 Kushnirenko's Theorem

The primary goal of this chapter is to prove the following theorem.

Theorem 3.2 (Kushnirenko's Theorem, [BKK76]). *Consider a system of n polynomials in n variables with support $\mathcal{A}$. This system has at most $n!\,\mathrm{vol}(\mathrm{NP}(\mathcal{A}))$ isolated solutions in $\mathbb{T}^n$. If the polynomials are generic given their support $\mathcal{A}$, then there are exactly $n!\,\mathrm{vol}(\mathrm{NP}(\mathcal{A}))$ isolated solutions in $\mathbb{T}^n$.*

Remark 3.3. Note that there might still be some (possibly infinite) component in the solution set of the system. The theorem only gives a bound on the number of *isolated* solutions. •

To prove Kushnirenko's Theorem, it will be helpful to consider linear forms in projective space, which correspond to our polynomial system. For this, we will use the map $\varphi_{\mathcal{A}}$, which we use to define a toric variety. We will need to understand this map quite well — an important step in this will be determining the size of its kernel. We will then show that the zero set of our polynomial system corresponds to the intersection of our toric variety with a linear space, which is cut out by the linear

forms corresponding to our original system. Finally, the Hilbert polynomial of our toric variety will give us the final piece of the puzzle.

This process will take up most of this chapter, and it is quite technical. We begin by defining the map

$$\begin{aligned}\varphi_{\mathcal{A}} \colon \mathbb{T}^n &\to \mathbb{P}^{\mathcal{A}},\\ x &\mapsto [x^a \mid a \in \mathcal{A}]\,.\end{aligned}$$

Then $X_{\mathcal{A}} := \overline{\varphi_{\mathcal{A}}(\mathbb{T}^n)} \subseteq \mathbb{P}^{\mathcal{A}}$ is a *toric variety*.

To aid in our computation of $\ker(\varphi_{\mathcal{A}})$, we first consider a factorization of $\varphi_{\mathcal{A}}$. For this, we need the following definition.

Definition 3.4. The *diagonal torus* in $\mathbb{T}^{\mathcal{A}}$ is $\delta(\mathbb{T}) = \{(a, \dots, a) \in \mathbb{T}^{\mathcal{A}} \mid a \in \mathbb{T}\} \subset \mathbb{T}^{\mathcal{A}}$. The *dense torus in projective space* is $\mathbb{T}^{\mathcal{A}}/\delta(\mathbb{T}) \subset \mathbb{P}^{\mathcal{A}}$. •

If we think about $\mathbb{T}^n$ and $\mathbb{T}^{\mathcal{A}}$ as multiplicative groups, then this quotient is indeed well-defined. Quotienting out with the diagonal torus corresponds to the points being invariant under scaling. Hence, the dense torus can be thought of as a kind of "projectivization" of the torus. The proof of the following lemma is then straightforward.

Lemma 3.5. *Let $\rho \colon \mathbb{T}^n \to \mathbb{T}^{\mathcal{A}}$ be given by $x \mapsto (x^a \mid a \in \mathcal{A})$, and let $\pi \colon \mathbb{T}^{\mathcal{A}} \to \mathbb{T}^{\mathcal{A}}/\delta(\mathbb{T})$ be the projection onto the dense torus. Then the following diagram commutes.*

$$\begin{array}{ccc} \mathbb{T}^n & \xrightarrow{\varphi_{\mathcal{A}}} & \mathbb{P}^{\mathcal{A}} \\ {\scriptstyle\rho}\downarrow & & \uparrow \\ \mathbb{T}^{\mathcal{A}} & \xrightarrow{\pi} & \mathbb{T}^{\mathcal{A}}/\delta(\mathbb{T}) \end{array}$$

With this in mind, we can now consider $\varphi_{\mathcal{A}} \colon \mathbb{T}^n \to \mathbb{T}^{\mathcal{A}}/\delta(\mathbb{T})$ as a group homomorphism.

We have not made any assumptions about $\mathcal{A}$ other than the fact that it affinely spans $\mathbb{R}^n$. But we can actually make another assumption without loss of generality.

Remark 3.6. If $0 \notin \mathcal{A}$, we translate $\mathcal{A}$ to $\mathcal{A}' = \mathcal{A} + a_0$, such that $0 \in \mathcal{A}'$. This translation corresponds to multiplying each polynomial in F by x^{a_0}, which does not change the zero set in $\mathbb{T}^n$ of our system. A point in $\varphi_{\mathcal{A}}(\mathbb{T}^n)$ must be of the form $[x^a \mid a \in \mathcal{A}]$ for some $x \in \mathbb{T}^n$. Then $\varphi_{\mathcal{A}'}(x) = [x^{a+a_0} \mid a \in \mathcal{A}] = [x^{a_0}x^a \mid a \in \mathcal{A}] = [x^a \mid a \in \mathcal{A}] = \varphi_{\mathcal{A}}(x)$. Hence, without loss of generality, we can assume that $0 \in \mathcal{A}$. For the rest of the chapter, we assume that $a_0 = 0 \in \mathcal{A}$. •

We are now able to determine the size of the kernel of $\varphi_{\mathcal{A}}$.

Lemma 3.7. $|\ker \varphi_{\mathcal{A}}| = [\mathbb{Z}^n : \mathbb{Z}\mathcal{A}]$.

Proof. By Remark 3.6, we have $a_0 = 0 \in \mathcal{A}$. As $x^{a_0} = 1$, it follows, by Lemma 3.5, that $\rho(\mathbb{T}^n) = 1 \times \mathbb{T}^{|\mathcal{A}|-1}$. Note that the composition (which is just π restricted to $\rho(\mathbb{T}^n)$)

$$1 \times \mathbb{T}^{|\mathcal{A}|-1} \hookrightarrow \mathbb{T}^{\mathcal{A}} \twoheadrightarrow \mathbb{T}^{\mathcal{A}}/\delta(\mathbb{T})$$

is an isomorphism. Thus, $\ker \varphi_{\mathcal{A}} = \ker \rho$. We assumed that $\mathcal{A}$ affinely spans $\mathbb{R}^n$, so the sublattice $\mathbb{Z}\mathcal{A} \subseteq \mathbb{Z}^n$ has full rank. Therefore, $\mathbb{Z}^n/\mathbb{Z}\mathcal{A}$ is a finite abelian group of order $[\mathbb{Z}^n : \mathbb{Z}\mathcal{A}] = |\mathbb{Z}^n/\mathbb{Z}\mathcal{A}|$. Consider

$$0 \to \mathbb{Z}\mathcal{A} \xrightarrow{f} \mathbb{Z}^n \xrightarrow{g} \mathbb{Z}^n/\mathbb{Z}\mathcal{A} \to 0,$$

which is a short exact sequence of abelian groups. Applying the $\mathrm{Hom}(-, \mathbb{T})$ functor, we get the exact sequence

$$0 \to \mathrm{Hom}(\mathbb{Z}^n/\mathbb{Z}\mathcal{A}, \mathbb{T}) \xrightarrow{g_*} \mathrm{Hom}(\mathbb{Z}^n, \mathbb{T}) \xrightarrow{f_*} \mathrm{Hom}(\mathbb{Z}\mathcal{A}, \mathbb{T}).$$

It is well-known that $\mathrm{Hom}(\mathbb{Z}, G) \cong G$ for all abelian groups G. Hence, $\mathrm{Hom}(\mathbb{Z}, \mathbb{T}) \cong \mathbb{T}$. From homological algebra, we know that $\mathrm{Hom}(X \times Y, Z) \cong \mathrm{Hom}(X, Z) \times \mathrm{Hom}(Y, Z)$. Thus, we can make the identification $\mathbb{T}^n \cong \mathrm{Hom}(\mathbb{Z}^n, \mathbb{T})$. Then, as the sequence is exact,

$$\begin{aligned} \mathrm{Hom}(\mathbb{Z}^n/\mathbb{Z}\mathcal{A}, \mathbb{T}) &\cong \mathrm{im}\,(g_*) \\ &\cong \ker\,(f_*)\,, \end{aligned}$$

and, by the identification $\mathbb{T}^n \cong \mathrm{Hom}(\mathbb{Z}^n, \mathbb{T})$, we get that

$$\begin{aligned} \ker\,(f_*) &\cong \{x \in \mathbb{T}^n \mid x \mapsto (x^a = 1), \forall a \in \mathbb{Z}\mathcal{A}\} \\ &= \{x \in \mathbb{T}^n \mid x^a = 1, \forall a \in \mathbb{Z}\mathcal{A}\} \\ &= \{x \in \mathbb{T}^n \mid x^a = 1, \forall a \in \mathcal{A}\} \\ &= \ker \varphi_{\mathcal{A}}. \end{aligned}$$

Note that $\mathbb{Z}\mathcal{A} \subset \mathbb{Z}^n$ has full rank, so $\mathbb{Z}^n/\mathbb{Z}\mathcal{A}$ is a finite abelian group. We can consider elements of $\mathrm{Hom}(\mathbb{Z}^n/\mathbb{Z}\mathcal{A}, \mathbb{T})$ as characters from the finite abelian group $\mathbb{Z}^n/\mathbb{Z}\mathcal{A}$ into $\mathbb{C}^\times$. From a general result from character theory (see, e.g., [MV06, Theorem 4.4]), $\mathrm{Hom}(\mathbb{Z}^n/\mathbb{Z}\mathcal{A}, \mathbb{T}) \cong \mathbb{Z}^n/\mathbb{Z}\mathcal{A}$. Therefore, we have that

$$|\ker \varphi_{\mathcal{A}}| = |\mathrm{Hom}(\mathbb{Z}^n/\mathbb{Z}\mathcal{A}, \mathbb{T})| = |\mathbb{Z}^n/\mathbb{Z}\mathcal{A}| = [\mathbb{Z}^n : \mathbb{Z}\mathcal{A}].$$

■

The coordinate ring of $\mathbb{P}^{\mathcal{A}}$ is $\mathbb{C}[z_a \mid a \in \mathcal{A}]$, and the coordinate ring of $\mathbb{T}^n$ is $\mathbb{C}[x_1^{\pm}, \ldots, x_n^{\pm}]$. We define the *pullback* along $\varphi_{\mathcal{A}}$ as the map

$$\varphi_{\mathcal{A}}^*\colon \mathbb{C}[z_a \mid a \in \mathcal{A}] \to \mathbb{C}[x_1, \ldots, x_n], \tag{3.2}$$

which is given by precomposition by $\varphi_{\mathcal{A}}$. The pullback of a linear form $\Lambda = \sum_{a \in \mathcal{A}} c_a z_a$ is then

$$\varphi_{\mathcal{A}}^*(\Lambda)(x) = (\Lambda \circ \varphi_{\mathcal{A}})(x) = \sum_{a \in \mathcal{A}} c_a x^a.$$

I.e., we simply evaluate Λ in $(x^a \mid a \in \mathcal{A})$. In this way, $\varphi_{\mathcal{A}}^*$ gives us a straightforward bijective correspondence between sparse polynomials with support $\mathcal{A}$ and linear forms on $\mathbb{P}^{\mathcal{A}}$. Consider a sparse polynomial $f = \sum_{a \in \mathcal{A}} c_a x^a$. The corresponding linear form is, of course, $\Lambda_f(z) = \sum_{a \in \mathcal{A}} c_a z_a$, and $H_f = V(\Lambda_f)$ is the hyperplane cut out by Λ_f. Thus, we get a bijective map between the nonzero solutions $V(f) \subset \mathbb{T}^n$ and $\varphi_{\mathcal{A}}^{-1}(H_f \cap \varphi_{\mathcal{A}}(\mathbb{T}^n))$. We can also consider a system $C \cdot x^A = 0$ of n polynomials in n variables, where $C \in \mathbb{C}^{n \times |\mathcal{A}|}$ is some coefficient matrix. The corresponding linear forms $(\Lambda_1, \ldots, \Lambda_n)$ are cut out the hyperplanes $H_i = V(\Lambda_i)$. Then $L = \cap_{i=1}^n H_i$ is a linear space of codimension equal to the rank of C, and we get that $\varphi_{\mathcal{A}}^{-1}(L \cap \varphi_{\mathcal{A}}(\mathbb{T}^n)) = V(C \cdot x^A)$. If $\mathbb{Z}\mathcal{A} = \mathbb{Z}^n$, then, by Lemma 3.7, $|\ker \varphi_{\mathcal{A}}| = 1$, so, in this case, the solutions to F and the points of $L \cap \varphi_{\mathcal{A}}(\mathbb{T}^n)$ are in bijective correspondance.

We summarize this discussion in the following proposition.

Proposition 3.8 ([Sot11, Lemma 3.5])**.** *There is a bijective correspondance between zero sets of sparse polynomials with support $\mathcal{A}$ and preimages $\varphi_{\mathcal{A}}^{-1}(H \cap \varphi_{\mathcal{A}}(\mathbb{T}^n))$, where H is a hyperplane. The zero set of a system of polynomials $C \cdot x^A = 0$ in n variables with support $\mathcal{A}$ is the preimage $\varphi_{\mathcal{A}}^{-1}(L) = \varphi_{\mathcal{A}}^{-1}(L \cap \varphi_{\mathcal{A}}(\mathbb{T}^n))$, where L is a linear space of codimension equal to the rank of the coefficient matrix C. When $\mathbb{Z}\mathcal{A} = \mathbb{Z}^n$, solutions to $C \cdot x^A = 0$ are in bijective correspondence via $\varphi_{\mathcal{A}}$ with the points of $\varphi_{\mathcal{A}}(\mathbb{T}^n) \cap L$.*

The following definition introduces a central concept.

Definition 3.9. Let X be a subvariety of $\mathbb{P}^m$ of dimension n, and let L be a general linear subspace of codimension n. The *degree* of X, denoted $\deg(X)$, is the number of points in $L \cap X$. •

Recall that $X_{\mathcal{A}} = \overline{\varphi_{\mathcal{A}}(\mathbb{T}^n)}$. Let $L \subset \mathbb{P}^{\mathcal{A}}$ be a general linear subspace of codimension $\dim X_{\mathcal{A}}$. Then $\deg(X_{\mathcal{A}}) = |X_{\mathcal{A}} \cap L|$.

Lemma 3.10. *The number of isolated solutions in $\mathbb{T}^n$ to a generic system of n polynomials in n variables with support $\mathcal{A}$ is $|\ker \varphi_{\mathcal{A}}| \deg(X_{\mathcal{A}})$.*

Proof. Let $T := \overline{X_{\mathcal{A}} \setminus \varphi_{\mathcal{A}}(\mathbb{T}^n)}$. By the Closure Theorem, $\dim(T) < \dim(X_{\mathcal{A}})$. Let L be a general linear space of codimension $\dim(X_{\mathcal{A}})$. As L is general, and $\dim(T) < \operatorname{codim}(L)$, we have that $T \cap L = \emptyset$. So $L \cap X_{\mathcal{A}} = L \cap \varphi_{\mathcal{A}}(\mathbb{T}^n)$. By Proposition 3.8, the solution set of a generic system of polynomials (corresponding to the chosen general linear space L) is $\varphi_{\mathcal{A}}^{-1}(L \cap \varphi_{\mathcal{A}}(\mathbb{T}^n))$. Consider a point $y \in \varphi_{\mathcal{A}}(\mathbb{T}^n) \subseteq \mathbb{T}^{\mathcal{A}}/\delta(\mathbb{T})$. As $\varphi_{\mathcal{A}} \colon \mathbb{T}^n \to \mathbb{T}^{\mathcal{A}}/\delta(\mathbb{T})$ is a group homomorphism, $\varphi_{\mathcal{A}}^{-1}(y)$ is a coset of $\ker \varphi_{\mathcal{A}}$. It follows that the number of points in $\varphi_{\mathcal{A}}^{-1}(y)$ is equal to $|\ker \varphi_{\mathcal{A}}|$. By Lemma 3.7, the kernel of $\varphi_{\mathcal{A}}$ has finitely many points. Hence, $\left|\varphi_{\mathcal{A}}^{-1}(X_{\mathcal{A}} \cap L)\right| = |\ker \varphi_{\mathcal{A}}| \deg(X_{\mathcal{A}})$. ■

This lemma shows that the *generic root count* is well-defined. We let $d_{\mathcal{A}}$ denote the number of solutions in $\mathbb{T}^n$ to a generic system of polynomials.

Lemma 3.11. *The number of isolated solutions in $\mathbb{T}^n$ to any system of n polynomials in n variables with support $\mathcal{A}$ is at most $d_{\mathcal{A}}$.*

Lemma 3.11 can be proven by considering the intersection $X_{\mathcal{A}} \cap L$ of $X_{\mathcal{A}}$ by a linear subspace $L \subset \mathbb{P}^{\mathcal{A}}$ with codimension equal to $\dim X_{\mathcal{A}}$. One can then use Hilbert polynomials to show that $\deg(X_{\mathcal{A}})$ is an upper bound to the number of isolated points in $X_{\mathcal{A}} \cap L$. The full proof is outside the scope of this thesis.

It turns out to be helpful to work with the lift of $\mathcal{A}$ to a homogenized set of exponent vectors

$$\mathcal{A}^+ = \{(1, a) \mid a \in \mathcal{A}\} \subseteq 1 \times \mathbb{Z}^n.$$

The lift to $\mathcal{A}^+$ gives us the map

$$\begin{aligned} \varphi_{\mathcal{A}^+} \colon \mathbb{T}^{1+n} &\to \mathbb{P}^{\mathcal{A}} \\ (t, x) &\mapsto [tx^a \mid a \in \mathcal{A}]. \end{aligned}$$

Note that $\varphi_{\mathcal{A}^+}(\mathbb{T}^{1+n}) = \varphi_{\mathcal{A}}(\mathbb{T}^n)$ as $\varphi_{\mathcal{A}^+}(t, x) = [tx^a \mid a \in \mathcal{A}] = [x^a \mid a \in \mathcal{A}] = \varphi_{\mathcal{A}}(x)$ for all $(t, x) \in \mathbb{T}^{1+n}$. These notions allow us to determine the homogeneous coordinate ring of $X_{\mathcal{A}}$, which we denote by $S_{\mathcal{A}}$.

Proposition 3.12. *The homogeneous coordinate ring of* $X_{\mathcal{A}} = \overline{\varphi_{\mathcal{A}^+}(\mathbb{T}^{1+n})}$ *is* $S_{\mathcal{A}} \simeq \mathbb{C}[\mathbb{N}\mathcal{A}^+]$*, and the* d*'th graded piece of* $S_{\mathcal{A}}$ *has basis* $\{t^d \cdot x^a \mid (d,a) \in \mathbb{N}\mathcal{A}^+\}$.

Proof. In Equation (3.2), we defined the pullback

$$\varphi_{\mathcal{A}}^*\colon \mathbb{C}[z_a \mid a \in \mathcal{A}] \to \mathbb{C}[x_1^{\pm}, \ldots, x_n^{\pm}].$$

Similarly, by the above discussion on $\varphi_{\mathcal{A}^+}$, the pullback of $\varphi_{\mathcal{A}^+}$ is

$$\varphi_{\mathcal{A}^+}^*\colon \mathbb{C}[z_a \mid a \in \mathcal{A}] \to \mathbb{C}[t, x_1, \ldots, x_n].$$

We have that

$$\varphi_{\mathcal{A}^+}^*(\mathbb{C}[z_a \mid a \in \mathcal{A}]) = \mathbb{C}[tx^a \mid a \in \mathcal{A}].$$

It is a well-known result that the coordinate ring of $\overline{\varphi_{\mathcal{A}^+}(\mathbb{T}^n)} = X_{\mathcal{A}}$ is isomorphic to $\varphi_{\mathcal{A}^+}^*(\mathbb{C}[z_a \mid a \in \mathcal{A}])$. Thus,

$$S_{\mathcal{A}} \cong \mathbb{C}[tx^a \mid a \in \mathcal{A}] \cong \mathbb{C}[y^{(1,a)} \mid a \in \mathcal{A}] = \mathbb{C}[y^{a'} \mid a' \in \mathcal{A}^+] = \mathbb{C}[\mathbb{N}\mathcal{A}^+].$$

Let $f \in (\mathbb{C}[z_a \mid a \in \mathcal{A}])_d$ be in the d'th graded piece of $\mathbb{C}[z_a \mid a \in \mathcal{A}]$. Then f must be of the form

$$f = \sum_j c_j \prod_{a \in \mathcal{A}, \sum i_{a,j} = d} z_a^{i_{a,j}}.$$

The pullback of f is

$$\begin{aligned}\varphi_{\mathcal{A}}^*(f) &= \sum_j c_j \prod_{a \in \mathcal{A}, \sum i_{a,j} = d} (tx^a)^{i_{a,j}} \\ &= \sum_j c_j t^d \prod_{a \in \mathcal{A}, \sum i_{a,j} = d} x^{a \cdot i_{a,j}}.\end{aligned}$$

We see that the natural grading on the coordinate ring $S_{\mathcal{A}}$ is by the exponent of the variable t. It follows that the d'th graded piece $(S_{\mathcal{A}})_d := \mathbb{C}_d[X_{\mathcal{A}}]$ of $S_{\mathcal{A}}$ has basis

$$\{t^d x^a \mid (d,a) \in \mathbb{N}\mathcal{A}^+\}.$$

■

In preparation for our next lemma, we need to perform a bit of construction work. Let

$$\mathcal{B} = \left\{ b \in \mathbb{Z}\mathcal{A} \;\middle|\; b = \sum_{a \in \mathcal{A}} \beta_a a,\ \beta_a \in [0,1) \cap \mathbb{Q} \right\}.$$

To better understand this set, we illustrate it with a small example.

Example 3.13. Consider the set $\mathcal{A} = \{0, (1,0), (0,1), (1,1), (3,1)\}$. In Figure 3.1, the purple points are the set $S = \{b \mid b = \sum_{a \in \mathcal{A}} b_a a,\ b_a \in \{0,1\}\}$. By definition, $\mathcal{B}$ must be a subset of $\mathrm{conv}(S)$. The

points in $\mathcal{B}$ are

$$\begin{aligned}
(0,0) &= 0(1,1) + 0(1,0) + 0(0,1) + 0(3,1),\\
(1,1) &= \frac{1}{2}(1,1) + \frac{1}{2}(1,0) + \frac{1}{2}(0,1) + 0(3,1),\\
(2,1) &= \frac{1}{2}(1,1) + \frac{1}{2}(1,0) + \frac{1}{6}(0,1) + \frac{1}{3}(3,1),\\
(3,1) &= \frac{1}{3}(1,1) + \frac{2}{3}(1,0) + \frac{2}{3}(0,1) + \frac{2}{3}(3,1),\\
(2,2) &= \frac{5}{6}(1,1) + \frac{1}{6}(1,0) + \frac{1}{6}(0,1) + \frac{1}{3}(3,1),\\
(3,2) &= \frac{5}{6}(1,1) + \frac{1}{6}(1,0) + \frac{1}{2}(0,1) + \frac{2}{3}(3,1),\\
(4,2) &= \frac{5}{6}(1,1) + \frac{2}{3}(1,0) + \frac{1}{3}(0,1) + \frac{5}{6}(3,1).
\end{aligned}$$

These points are marked by green crosses in Figure 3.1. •

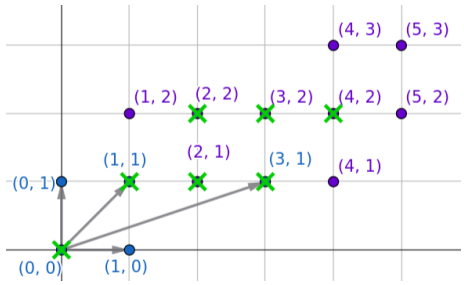


Figure 3.1: Example of the set $\mathcal{B}$, where $\mathcal{A} = \{0, (1,0), (0,1), (1,1), (3,1)\}$.

Clearly, $\mathcal{B}$ must be finite, as it is the intersection of a bounded set with a lattice. For each $b \in \mathcal{B}$, fix an integral expression

$$b = \sum_{a \in \mathcal{A}} b_a a,$$

such that $b_a \in \mathbb{Z}$. As there are only finitely many, we can find $\nu \in \mathbb{N}$, such that $-\nu \leq b_a$ for all b_a in such fixed integral expressions. We assumed that $0 \in \mathcal{A}$ (by Remark 3.6), so the value of b_0 does not affect the sum. By adjusting the value of the b_0's, we can find an integer $\mu \geq 0$, such that for all $b \in \mathcal{B}$, $\mu = \sum_{a \in \mathcal{A}} b_a$, and

$$(\mu, b) = \sum_{a \in \mathcal{A}} b_a (1, a), \tag{3.3}$$

where all b_a's still satisfy $-\nu \leq b_a \in \mathbb{Z}$, and we have $\nu, \mu \in \mathbb{N}$.

Lemma 3.14. *For $d \geq \nu |\mathcal{A}| + \mu$, the translation $v \mapsto v + \nu \sum_{a \in \mathcal{A}} (1, a) + (\mu, 0)$ maps elements of $\mathbb{Z}\mathcal{A}^+ \cap (d - \nu |\mathcal{A}| - \mu) \operatorname{NP}(\mathcal{A}^+)$ to elements of $\mathbb{N}\mathcal{A}^+ \cap d \operatorname{NP}(\mathcal{A}^+)$.*

Proof. Let $v \in \mathbb{Z}\mathcal{A}^+ \cap (d - \nu |\mathcal{A}| - \mu) \operatorname{NP}(\mathcal{A}^+)$. Let $x \in \operatorname{NP}(\mathcal{A}^+)$. We can write

$$x = \sum_{a \in \mathcal{A}^+} \alpha_a a, \quad \text{where } \alpha_a \in [0,1], \text{ and } \sum_{a \in \mathcal{A}^+} \alpha_a = 1.$$

As v is, in particular, in $(d - \nu\,|\mathcal{A}| - \mu)\,\mathrm{NP}(\mathcal{A}^+)$, we get that

$$v = \sum_{a\in\mathcal{A}} \alpha_a(1,a), \quad \text{where } \alpha_a \in \mathbb{Q}_{\geq 0}, \text{ and } d - \nu\,|\mathcal{A}| - \mu = \sum_{a\in\mathcal{A}} \alpha_a,$$

where $\alpha_a \in \mathbb{Q}$ follows from the fact that $v \in \mathbb{Z}\mathcal{A}^+$. For each $a \in \mathcal{A}$, let $\beta_a \in [0,1) \cap \mathbb{Q}$ and $\gamma_a \in \mathbb{N}$, such that $\alpha_a = \beta_a + \gamma_a$. Let $\beta = \sum_{a\in\mathcal{A}} \beta_a$. Then

$$v = \sum_{a\in\mathcal{A}} \beta_a(1,a) + \sum_{a\in\mathcal{A}} \gamma_a(1,a) = (\beta, b) + \sum_{a\in\mathcal{A}} \gamma_a(1,a)$$

for $b = \sum_{a\in\mathcal{A}} \beta_a a \in \mathcal{B}$. As $d \geq \nu\,|\mathcal{A}| + \mu$, we get that $d - \nu\,|\mathcal{A}| - \mu \in \mathbb{N}$. For all $a \in \mathcal{A}$, we have

$$\beta = \sum_{a\in\mathcal{A}} \alpha_a - \sum_{a\in\mathcal{A}} \gamma_a.$$

As $\alpha_a \geq \gamma_a \in \mathbb{N}$, and $\sum_{a\in\mathcal{A}} \alpha_a = d - \nu\,|\mathcal{A}| - \mu \in \mathbb{N}$, we get that $\beta \in \mathbb{N}$. By Equation (3.3), we get that

$$\begin{aligned} v &= (\beta, 0) + (0, b) + \sum_{a\in\mathcal{A}} \gamma_a(1,a) = (\beta,0) - (\mu,0) + (\mu, b) + \sum_{a\in\mathcal{A}} \gamma_a(1,a) \\ &= \beta(1,0) - (\mu,0) + \sum_{a\in\mathcal{A}} b_a(1,a) + \sum_{a\in\mathcal{A}} \gamma_a(1,a) = \beta(1,0) - (\mu,0) + \sum_{a\in\mathcal{A}} (b_a + \gamma_a)(1,a). \end{aligned}$$

It follows that

$$\begin{aligned} v &\mapsto v + \nu \sum_{a\in\mathcal{A}} (1,a) + (\mu,0) \\ &= \beta(1,0) - (\mu,0) + \sum_{a\in\mathcal{A}} (b_a + \gamma_a)(1,a) + \nu \sum_{a\in\mathcal{A}} (1,a) + (\mu,0) \\ &= \beta(1,0) + \sum_{a\in\mathcal{A}} (b_a + \nu + \gamma_a)(1,a). \end{aligned}$$

As $b_a \in \mathbb{Z}$, $\nu \in \mathbb{N}$, and $-\nu \leq b_a$ for all $a \in \mathcal{A}$, then $b_a + \nu \in \mathbb{N}$ for all $a \in \mathcal{A}$. Hence, $\beta, b_a + \nu + \gamma_a \in \mathbb{N}$ for all $a \in \mathcal{A}$, so the image of v is in $\mathbb{N}\mathcal{A}^+$. As $b_a \leq \mu$, $\gamma_a \leq \alpha_a \leq \sum_{a\in\mathcal{A}} \alpha_a = d - \nu\,|\mathcal{A}| - \mu$, and $\nu \leq \nu\,|\mathcal{A}|$, we must have $b_a + \nu + \gamma_a \leq d$. Therefore, the image of v is in $\mathbb{N}\mathcal{A}^+ \cap \mathrm{NP}(\mathcal{A}^+)$, which completes the proof. ■

We define Hilbert polynomials and recall some basic facts.

Definition 3.15. Let X be a projective variety. The *Hilbert function* H_X of X is $H_X(d) = \dim_{\mathbb{C}} \mathbb{C}_d[X]$. The *Hilbert polynomial* h_X of X is a polynomial, such that for all $d \in \mathbb{N}$ large enough $h_X(d) = H_X(d)$. •

Proposition 3.16. *The Hilbert polynomial h_X of a projective variety X satisfies the following properties:*

(1) $\deg(h_X) = \dim(X)$.

(2) $\mathrm{LC}(h_X) = \frac{\deg(X)}{\dim(X)!}$.

We let $H_{\mathcal{A}}$ and $h_{\mathcal{A}}$ denote the Hilbert function and polynomial, respectively, of $X_{\mathcal{A}}$. From Proposition 3.12, a basis of $(S_{\mathcal{A}})_d$ is $\{t^d x^a \mid (d,a) \in \mathbb{N}\mathcal{A}^+\}$. Note that this basis is indexed over $d\mathcal{A} = d\mathbb{N}\mathcal{A} \cap \mathrm{NP}(\mathcal{A})$. Then, from Proposition 3.12, we get that

$$H_{\mathcal{A}}(d) = \dim_{\mathbb{C}}(S_{\mathcal{A}})_d = \left|\{t^d x^a \mid (d,a) \in \mathbb{N}\mathcal{A}^+\}\right| = |d\mathcal{A}| = |\mathbb{N}\mathcal{A} \cap d\,\mathrm{NP}(\mathcal{A})|\,. \tag{3.4}$$

Ehrhart polynomials are the final ingredient we need to prove Kushnirenko's Theorem.

Definition 3.17. Let $\Lambda \subseteq \mathbb{Z}^n$ be a lattice, and let Δ be a lattice polytope in Λ (i.e., a polytope with vertices in Λ). The *Ehrhart polynomial* of Δ is $P_\Delta(d) = |\Lambda \cap d\Delta|$ for $d \in \mathbb{N}$. •

Proposition 3.18. *The Ehrhart polynomial P_Δ of a lattice polytope Δ satisfies the following properties:*

(1) $\deg(P_\Delta) = \dim(\mathrm{Aff}(\Delta))$.

(2) If Δ has dimension n, then $\mathrm{LC}(P_\Delta) = \frac{\mathrm{vol}(\Delta)}{[\mathbb{Z}^n : \mathbb{Z}\Lambda]}$.

The following lemma describes the relation between Ehrhart polynomials and Hilbert polynomials.

Lemma 3.19. $\mathrm{LC}(h_{\mathcal{A}}) = \mathrm{LC}(P_{\mathrm{NP}(\mathcal{A})}) = \frac{\mathrm{vol}(\mathrm{NP}(\mathcal{A}))}{|\mathbb{Z}^n/\mathbb{Z}\mathcal{A}|}$ *and* $\deg(h_{\mathcal{A}}) = n$.

Proof. Letting $\Lambda = \mathbb{Z}\mathcal{A}$ and $\Delta = \mathrm{NP}(\mathcal{A})$, we can apply Proposition 3.18 to the Ehrhart polynomial, $P_{\mathrm{NP}(\mathcal{A})}$, of $\mathrm{NP}(\mathcal{A})$. Recall that $d\mathcal{A} = d\,\mathrm{NP}(\mathcal{A}) \cap \mathbb{N}\mathcal{A} \subset d\,\mathrm{NP}(\mathcal{A}) \cap \mathbb{Z}\mathcal{A}$. So

$$P_{\mathrm{NP}(\mathcal{A})}(d) = |\mathbb{Z}\mathcal{A} \cap d\,\mathrm{NP}(\mathcal{A})| \geq |\mathbb{N}\mathcal{A} \cap d\,\mathrm{NP}(\mathcal{A})| = H_{\mathcal{A}}(d).$$

By Lemma 3.14, we can map elements from $\mathbb{Z}\mathcal{A}^+ \cap (d - \nu\,|\mathcal{A}| - \mu)\,\mathrm{NP}(\mathcal{A}^+)$ to $\mathbb{N}\mathcal{A}^+ \cap d\,\mathrm{NP}(\mathcal{A}^+)$ by a translation. Hence, the map ψ is, in particular, injective. Thus, by Equation (3.4), we get that

$$H_{\mathcal{A}}(d) = |\mathbb{N}\mathcal{A} \cap d\,\mathrm{NP}(\mathcal{A})| \geq |\mathbb{Z}\mathcal{A} \cap (d - \nu\,|\mathcal{A}| - \mu)\,\mathrm{NP}(\mathcal{A})| = P_{\mathrm{NP}(\mathcal{A})}(d - \nu\,|\mathcal{A}| - \mu).$$

We have now established that $P_{\mathrm{NP}(\mathcal{A})}(d - \nu\,|\mathcal{A}| - \mu) \leq H_{\mathcal{A}}(d) \leq P_{\mathrm{NP}(\mathcal{A})}(d)$. Hence, for d large enough,

$$P_{\mathrm{NP}(\mathcal{A})}(d - \nu\,|\mathcal{A}| - \mu) \leq h_{\mathcal{A}}(d) \leq P_{\mathrm{NP}(\mathcal{A})}(d).$$

As $P_{\mathrm{NP}(\mathcal{A})}$ and $h_{\mathcal{A}}$ are both polynomials, it follows, from Proposition 3.18, that

$$\deg(h_{\mathcal{A}}) = \deg(P_{\mathrm{NP}(\mathcal{A})}) = \dim(\mathrm{Aff}(\mathrm{NP}(\mathcal{A}))) = n,$$

as $\mathcal{A}$ is assumed to affinely span $\mathbb{R}^n$, and that $\mathrm{LC}(h_{\mathcal{A}}) = \mathrm{LC}(P_{\mathrm{NP}(\mathcal{A})}) = \frac{\mathrm{vol}(\mathrm{NP}(\mathcal{A}))}{[\mathbb{Z}^n : \mathbb{Z}\mathcal{A}]}$. ■

We are now finally ready to prove Kushnirenko's Theorem.

Proof of Kushnirenko's Theorem (Theorem 3.2). From Proposition 3.16, we have that $\deg(h_{\mathcal{A}}) = \dim(X_{\mathcal{A}})$, and $\mathrm{LC}(h_{\mathcal{A}}) = \frac{\deg(X_{\mathcal{A}})}{\dim(X_{\mathcal{A}})!}$. By Lemmas 3.7, 3.10 and 3.19, it follows that

$$\begin{aligned} d_{\mathcal{A}} &= |\ker \varphi_{\mathcal{A}}| \deg(X_{\mathcal{A}}) = [\mathbb{Z}^n : \mathbb{Z}\mathcal{A}] \deg(X_{\mathcal{A}}) = [\mathbb{Z}^n : \mathbb{Z}\mathcal{A}] \,\mathrm{LC}(h_{\mathcal{A}})(\dim(X_{\mathcal{A}})!) \\ &= [\mathbb{Z}^n : \mathbb{Z}\mathcal{A}] \frac{\mathrm{vol}(\mathrm{NP}(\mathcal{A}))}{[\mathbb{Z}^n : \mathbb{Z}\mathcal{A}]} n! = n!\,\mathrm{vol}(\mathrm{NP}(\mathcal{A})). \end{aligned}$$

Observe that we in Lemma 3.10 use the fact that our system of polynomials is generic.

By Lemma 3.11, $d_{\mathcal{A}}$ is an upper bound on the number of solutions for a non-generic choice of polynomials. ■

3.2 Bernstein's Theorem

As promised, we now cover the case, when the polynomials in our system $F = (f_1, \ldots, f_n)$ have different supports. If this is the case, we call F a *mixed system*. Let $\mathcal{A}_i$ denote the support of f_i. We recall the following result about the mixed volume.

Proposition 3.20 ([CLO05, Theorem 4.12]). *Let* $\mathrm{vol}_n(P)$ *be the* n*-dimensional volume of the polytope* P*. The* n*-dimensional mixed volume of the polytopes* $P_1, \ldots, P_n$ *is*

$$\mathrm{MV}_n(P_1, \ldots, P_n) = \sum_{k=1}^{n} (-1)^{n-k} \sum_{I \subseteq [n], |I|=k} \mathrm{vol}_n \left(\sum_{i \in I} P_i \right).$$

The following result due to Bernstein is a generalization of Kushnirenko's Theorem, which is why it is sometimes referred to as the Bernstein–Kushnirenko Theorem. The most common name for the result, however, is the BKK bound for Bernstein, Khovanskii, and Kushnirenko (Khovanskii has found more than a dozen different proofs of the theorem).

Theorem 3.21 (Bernstein's Theorem, [Ber75]). *Let* $F = (f_1, \ldots, f_n)$ *be a system of* n *polynomials in* n *variables with supports* $\mathcal{A}_1, \ldots, \mathcal{A}_n$*. Then* F *has at most* $\mathrm{MV}_n(\mathrm{NP}(\mathcal{A}_1), \ldots, \mathrm{NP}(\mathcal{A}_n))$ *isolated solutions in* $\mathbb{T}^n$*. If* F *is generic given* $\mathcal{A}_1, \ldots, \mathcal{A}_n$*, equality holds.*

To prove Bernstein's Theorem, one can apply Kushnirenko's Theorem, and then show multilinearity of the mixed volume and of the generic root count (see, e.g., [Sot11, Chapter 3]). The following two examples illustrate that while Bernstein's Theorem will often give a better bound on the number of isolated solutions in $\mathbb{T}^n$, this bound is not always an improvement over the one from Bezout's Theorem.

Example 3.22. Consider the system

$$f_1 = x^{27} + 42y^{15} = 0,$$
$$f_2 = y^{43} - 5x^{20} = 0,$$

where $\mathcal{A}_1 = \{(27, 0), (0, 15)\}$ and $\mathcal{A}_2 = \{(0, 43), (20, 0)\}$. If we just use Bezout's Theorem, we find that there can be at most $27 \cdot 43 = 1161$ isolated solutions in $\mathbb{C}^2$ to the system. Using Bernstein's Theorem brings this bound down quite a bit to $\mathrm{MV}_2(\mathrm{NP}(\mathcal{A}_1), \mathrm{NP}(\mathcal{A}_2)) = 861$ solutions in $\mathbb{T}^2$ (see [Kal25] for calculations). •

Example 3.23. We now consider a different system

$$f_1 = x^{57} - xy + 1 = 0,$$
$$f_2 = y^{63} - xy + 2 = 0.$$

Here, $\mathcal{A}_1 = \{(57, 0), (1, 1), (0, 0)\}$ and $\mathcal{A}_2 = \{(0, 63), (1, 1), (0, 0)\}$. Applying Bezout's Theorem tell us that there can be at most $57 \cdot 63 = 3591$ isolated solutions in $\mathbb{C}^2$ to the system. In this case, Bernstein's Theorem actually gives us the exact same bound: $\mathrm{MV}_2(\mathrm{NP}(\mathcal{A}_1), \mathrm{NP}(\mathcal{A}_2)) = 3591$ solutions in $\mathbb{T}^2$ (see [Kal25] for calculations). •

4 Fewnomial Theory

Having found generalizations for bounding the number of complex solutions in the multivariate case, we now want to find bounds on the number of real solutions as well. We are particularly interested in positive real solutions. Ultimately, we would like a bound, which does not depend on the degrees, nor on the volume(s) of the Newton polytope(s). Instead, we want to find bounds, which depend on the number of terms in the polynomials. An important step in this direction is Khovanskii's fewnomial theory. A *fewnomial* is a polynomial with *few* terms. In 1980, Khovanskii published his fewnomial bound. It was revolutionary, as it was the first bound, which was independent of the degrees of the polynomials.

Theorem 4.1 (Khovanskii's Fewnomial Bound, [Kho80])**.** *A system of n polynomials in n variables involving $n+l+1$ distinct monomials has strictly less than*

$$2^{\binom{n+l}{2}}(n+1)^{n+l}$$

nondegenerate positive solutions.

A solution $x \in \mathbb{R}$ of a system of polynomials is nondegenerate, if the derivatives of the polynomials evaluated at x span $\mathbb{R}^n$.

We consider an example to see how the different bounds, we have covered, compare to each other.

Example 4.2 (Continuation of Example 3.23)**.** Recall the system from Example 3.23:

$$f_1 = x^{57} - xy + 1 = 0,$$
$$f_2 = y^{63} - xy + 2 = 0.$$

In the previous chapter, we found the system to have at most 3591 isolated solutions in $\mathbb{T}^2$. However, we are interested in positive solutions. Hence, we want to apply Khovanskii's Fewnomial Bound. Here $n = 2$, and $l = 1$. So there are at most

$$2^{\binom{3}{2}}(2+1)^3 = 216$$

nondegenerate positive solutions to our system. Note that the same bound would also apply for Example 3.22. •

To find an even better bound, we will need a very different method called Gale dualization.

4.1 Gale duality

Using the terminology of fewnomial theory, we recall the setting from the previous chapter. Let $\mathcal{A} = \{0, a_1, \ldots, a_{n+l}\} \subset \mathbb{Z}^n$ be vectors, which span $\mathbb{R}^n$. As $0 \in \mathcal{A}$, this is equivalent to assuming that $\mathcal{A}$ affinely spans $\mathbb{R}^n$. Let $C \in \mathbb{C}^{n\times(n+l+1)}$ be some coefficient matrix. For any such matrix, we denote by

$$\begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = C \cdot x^A = 0$$

the corresponding fewnomial system in n variables $x = (x_1, \dots, x_n)$ with support $\mathcal{A}$. The solutions to this system may (like before) be interpreted as $\varphi_{\mathcal{A}}^{-1}(L)$, where $\varphi_{\mathcal{A}} \colon \mathbb{T}^n \to \mathbb{T}^{n+l} \subset \mathbb{C}^{n+l}$ is given by $x \mapsto (x^{a_1}, \dots, x^{a_{n+l}})$, and $L \subset \mathbb{C}^{n+l}$ is a codimension n plane given by some linear forms corresponding to the f_i. Note that we have restricted the codomain of $\varphi_{\mathcal{A}}$ to the affine hyperplane $[1 : z_{a_1} : \dots : z_{a_{n+l}}] \subset \mathbb{P}^{\mathcal{A}}$.

Definition 4.3. A set $\mathcal{A} \subset \mathbb{Z}^n$ of exponent vectors is *primitive*, if its $\mathbb{Z}$-affine span is all of $\mathbb{Z}^n$. •

Unless stated otherwise, henceforth, we assume that $\mathcal{A}$ is primitive, i.e., $\mathbb{Z}\mathcal{A} = \mathbb{Z}^n$. By Lemma 3.7, $|\ker \varphi_{\mathcal{A}}| = [\mathbb{Z}^n : \mathbb{Z}\mathcal{A}] = 1$, and it follows that $\varphi_{\mathcal{A}}$ is then injective. So $\mathbb{T}^n \supset V(f_1, \dots, f_n) \cong \varphi_{\mathcal{A}}(\mathbb{T}^n) \cap L \subset \mathbb{T}^{n+l}$.

In an attempt to make things simpler, we now want to parameterize L in a different way, which will give us the *Gale dual* of our original system above. Before we are able to define this Gale dual, we must introduce a lot of new terminology.

Construction 4.4. Let $p_1, \dots, p_{n+l} \in \mathbb{C}[y_1, \dots, y_l]$ be polynomials of degree one, which are pairwise non-proportional. Let

$$\mathcal{H} := \left\{ y \in \mathbb{C}^l \;\middle|\; \prod_{i=1}^{n+l} p_i(y) = 0 \right\} \subset \mathbb{C}^l$$

be a hyperplane arrangement. We can associate a collection of *weights* $\mathcal{B} = \{\beta_1, \dots, \beta_l\} \subset \mathbb{Z}^{n+l}$ to $\mathcal{H}$, which we then call a *weighted hyperplane arrangement*. A *master function* for the weighted arrangement $\mathcal{H}$ is of the form

$$p^{\beta} := p_1^{b_1} \dots p_{n+l}^{b_{n+l}},$$

where $\beta = (b_1, \dots, b_{n+l}) \in \mathcal{B}$. As some coordinates of β might be negative, we let p^{β} be defined on the domain $M_{\mathcal{H}} = \mathbb{C}^l \setminus \mathcal{H}$. If $\{1, p_1, \dots, p_{n+l}\}$ spans the space of degree one polynomials in $\mathbb{C}[y_1, \dots, y_l]$, we say that $\mathcal{H}$ is *essential*. This definition is equivalent to the normal vectors of the hyperplanes in $\mathcal{H}$ spanning $\mathbb{C}^l$.

A *system of master functions* in $M_{\mathcal{H}}$ with weights $\mathcal{B}$ is a system

$$p^{\beta_1} = \dots = p^{\beta_l} = 1.$$

We can assume the constant in these equations to be 1, as there are more polynomials than indeterminates, so any other constant can be suppressed.

The set $\mathcal{B}$ is *saturated*, if its elements are linearly independent and $\mathbb{Z}\mathcal{B} = \mathbb{Q}\mathcal{B} \cap \mathbb{Z}^{n+l}$ (note that $\mathbb{Q}\mathcal{B} \cap \mathbb{Z}^{n+l}$ is simply all integer points in the span of $\mathcal{B}$).

We assume, henceforth, that the system of master functions defines a zero-dimensional variety in $M_{\mathcal{H}}$, as then the elements of $\mathcal{B}$ are linearly independent, and the suppressed constants are sufficiently general. •

Now that we have this construction, we want to show its relation to our previous setting. We define the function $\psi_p \colon \mathbb{C}^l \to \mathbb{C}^{n+l}$ by

$$y \mapsto (p_1(y), \dots, p_{n+l}(y)).$$

Note that ψ_p is injective, if and only if, $\{1, p_1, \dots, p_{n+l}\}$ spans the space of degree one polynomials in $\mathbb{C}[y_1, \dots, y_l]$. Hence, ψ_p being injective is equivalent to $\mathcal{H}$ being essential. Recall that the pullback along ψ_p is the map $\psi_p^* \colon \mathbb{C}[z_1, \dots, z_{n+l}] \to \mathbb{C}[y_1, \dots, y_l]$ given by $g \mapsto g \circ \psi_p$.

Theorem 4.5 ([Sot11, Theorem 6.1]). *A system of master functions $p(y)^{\beta_1} = \cdots = p(y)^{\beta_l} = 1$ in $M_{\mathcal{H}}$ is the pullback along ψ_p of the linear space $\psi_p(\mathbb{C}^l)$ intersected with $\mathbb{G}$, where $\mathbb{G} \subset \mathbb{T}^{n+l}$ is a subgroup of dimension n. Any such pullback defines a system of master functions in $M_{\mathcal{H}}$. If ψ_p is injective, then the solutions to $p(y)^{\beta_1} = \cdots = p(y)^{\beta_l} = 1$ in $M_{\mathcal{H}}$ is isomorphic to $\psi_p(\mathbb{C}^l) \cap \mathbb{G}$.*

Proof. Let $h(z) = \prod_{i=1}^{n+l} z_i$ be a polynomial in $\mathbb{C}[z_1, \ldots, z_{n+l}]$, and let $H = V(h) \in \mathbb{C}^{n+l}$ denote the coordinate hyperplanes in $\mathbb{C}^{n+l}$. Note that $\mathbb{T}^{n+l} = \mathbb{C}^{n+l} \setminus H$. Then $\psi_p^*(h) = h \circ \psi_p(y) = \prod_{i=1}^{n+l} p_i(y)$, and we get that

$$\psi_p^{-1}(H) = \left\{ y \in \mathbb{C}^l \;\middle|\; \prod_i p_i(y) = 0 \right\} = \mathcal{H},$$

and $\psi_p^{-1}(\mathbb{T}^{n+l}) = M_{\mathcal{H}}$.

Define the subgroup $\mathbb{G} := \{z \in \mathbb{T}^{n+l} \mid z^{\beta_1} = \cdots = z^{\beta_l} = 1\}$ of $\mathbb{T}^{n+l}$. Note that $\mathbb{G} = V(z^{\beta_i} - 1 \mid i = 1, \ldots, n+l)$. For each i, we get that $\psi_p^*(z^{\beta_i} - 1) = p^{\beta_i} - 1$. So the pullback of the defining equations of $\mathbb{G}$ is the system of master functions $p(y)^{\beta_1} = \cdots = p(y)^{\beta_l} = 1$ in $M_{\mathcal{H}}$. Thus

$$\psi_p^{-1}\left(\psi_p(\mathbb{C}^l) \cap \mathbb{G}\right) = \{y \in \mathbb{C}^l \mid p(y)^{\beta_1} = \cdots = p(y)^{\beta_l} = 1\}.$$

The elements of $\mathcal{B}$ being linearly independent is equivalent to $\dim(\mathbb{G}) = n$. Additionally, the set $\mathcal{B}$ is saturated, if and only if, $\mathbb{G}$ is also connected. If ψ_p is assumed to be injective, the isomorphism follows. ■

We are now finally able to define the Gale dual. Suppose $\mathbb{G} \subset \mathbb{T}^{n+l}$ is a connected subgroup of dimension n. Suppose $L \subset \mathbb{C}^{n+l}$ is a linear subspace of dimension l, which is not parallel to any coordinate hyperplane. Then $\mathbb{G} \cap L$ is zero-dimensional.

Definition 4.6. Suppose we are given the following:

(i) A primitive set of exponent vectors, $\mathcal{A} = \{0, a_1, \ldots, a_{n+l}\} \subset \mathbb{Z}^n$, and equations $z^{\beta_1} = \cdots = z^{\beta_l} = 1$ defining $\mathbb{G} = \varphi_{\mathcal{A}}(\mathbb{T}^n)$ as a subset of $\mathbb{T}^{n+l}$. Then $\varphi_{\mathcal{A}} \colon \mathbb{T}^n \to \mathbb{G}$ is an isomorphism, and $\mathcal{B} = \{\beta_1, \ldots, \beta_l\}$ is saturated.

(ii) An affine-linear isomorphism $\psi_p \colon \mathbb{C}^l \to L$ and degree one polynomials $\Lambda_1, \ldots, \Lambda_n$ on $\mathbb{C}^{n+l}$ defining L.

Let $\mathcal{H} \subset \mathbb{C}^l$ be the zero set of $\psi_p^*\left(\prod_i z_i\right)$ (as seen in the above proof, this is equivalent to our usual definition of $\mathcal{H}$). The system of sparse polynomials on $\mathbb{T}^n$

$$\varphi_{\mathcal{A}}^*(\Lambda_1) = \cdots = \varphi_{\mathcal{A}}^*(\Lambda_n) = 0 \tag{4.1}$$

with support $\mathcal{A}$ is *Gale dual* to the system of master functions in $M_{\mathcal{H}}$ with weights $\mathcal{B}$

$$p(y)^{\beta_1} = \cdots = p(y)^{\beta_l} = 1, \tag{4.2}$$

and vice versa. •

The following result is immediate from Theorem 4.5.

Theorem 4.7 ([Sot11, Theorem 6.3]). *A pair of Gale dual systems are isomorphic.*

Having developed the terminology, we can now describe the actual process of finding the Gale dual of a system of polynomials.

Algorithm 4.8 (Gale dualization)**.**

Setup: Suppose $\mathcal{A} \subset \mathbb{Z}^n$ is primitive. Let Let $C \in \mathbb{C}^{n\times(n+l+1)}$ be a coefficient matrix, and let

$$\begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = C \cdot x^A = 0 \tag{4.3}$$

be the corresponding system of polynomials over $\mathbb{R}$ with support $\mathcal{A} = \{a_0, a_1, \ldots, a_{n+l}\}$. Assume additionally that $f_1, \ldots, f_n$ are linearly independent.

Step 1: As $\mathcal{A}$ is primitive, we can assume that $a_0 = 0$. We solve the system in Equation (4.3) for the monomials corresponding to $a_1, \ldots, a_n$, and get

$$\begin{aligned} x^{a_1} &= q_1(x), \\ &\vdots \\ x^{a_n} &= q_n(x), \end{aligned}$$

where the q_i's denote the polynomials resulting from solving for the x_i's. Set

$$\begin{aligned} p_1(x^{a_{1+n}}, \ldots, x^{a_{n+l}}) &:= q_1(x), \\ &\vdots \\ p_n(x^{a_{1+n}}, \ldots, x^{a_{n+l}}) &:= q_n(x). \end{aligned}$$

Furthermore, set $p_i(x^{a_{1+n}}, \ldots, x^{a_{n+l}}) := x^{a_i}$ for $i = 1+n, \ldots, n+l$. Note that we consider the p_i's as functions of the monomials $x^{a_{1+n}}, \ldots, x^{a_{n+l}}$. Then all p_i's are degree one polynomial functions in l arguments.

Step 2: Let $b_1, \ldots, b_{n+l} \in \mathbb{Z}$, such that $b_1a_1 + \cdots + b_{n+l}a_{n+l} = 0$. This is an integer linear combination of the exponent vectors in $\mathcal{A}$. Equivalently, we have that $(x^{a_1})^{b_1} \ldots (x^{a_{n+l}})^{b_{n+l}} = 0$. Let $\beta = (b_1, \ldots, b_{n+l})$. It follows that

$$(p_1(x^{a_{1+n}}, \ldots, x^{a_{n+l}}))^{b_1} \ldots (p_{n+l}(x^{a_{1+n}}, \ldots, x^{a_{n+l}}))^{b_{n+l}} = 1.$$

Then, for $y = (y_1, \ldots, y_l) \in \mathbb{C}^l$, we get the master functions

$$p(y)^\beta = (p_1(y_1, \ldots, y_l))^{b_1} \cdots (p_{n+l}(y_1, \ldots, y_l))^{b_{n+l}} = 1.$$

Step 3: The equations $p_i(y) = 0$ each define a hyperplane in a hyperplane arrangement $\mathcal{H}$ in $\mathbb{C}^l$. Let $\mathcal{B} := \{\beta_1, \ldots, \beta_l\} \subset \mathbb{Z}^{n+l}$ be the weights of the arrangement $\mathcal{H}$, such that the $\mathbb{Z}$-module of integral linear relations between the nonzero vectors in $\mathcal{A}$ has basis $\mathcal{B}$. From $\mathcal{B}$, we then get a system of master functions

$$p(y)^{\beta_1} = \cdots = p(y)^{\beta_l} = 1 \tag{4.4}$$

in $M_\mathcal{H}$. •

Remark 4.9. Just as the system of sparse polynomials in Equation (4.1) is Gale dual to the system of master functions in Equation (4.2), we also say that $\mathcal{B}$ is Gale dual to $\mathcal{A}$. In particular, we call $B = \begin{pmatrix}\beta_1 & \dots & \beta_l\end{pmatrix} \in \mathbb{Z}^{(n+l)\times l}$ the *Gale transform* of $A^+ = \begin{pmatrix}1 & \dots & 1\\ a_0 & \dots & a_{n+1}\end{pmatrix}$.

Consider coefficient matrix C. If we have a solution $x \in \mathbb{R}^n$ to the system $C \cdot x^A = 0$, then the vector $(x^{a_0}, x^{a_1}, \dots, x^{a_{n+l}})$ is in the kernel of C. Hence, by the above algorithm, the vectors in the kernel are of the form $(1, p_1(y), \dots, p_{n+l}(y))$ for $y \in \mathbb{R}^l$. For each $i = 1, \dots, n+l$, let $P_i \in \mathbb{R}^{l+1}$ be row vectors, such that $p_i(y) = \langle P_i, (1, y)\rangle$, and let $P_0 = (1, 0)$. Then $P_0, \dots, P_{n+l}$ span the kernel of C. We say that $P = (P_0, \dots, P_{n+l})^T$ is Gale dual to the coefficient matrix C. Similarly, the columns of B span the kernel of A^+ (this is clear from Step 2 of Algorithm 4.8). •

Similarly to $\mathbb{T}^{\mathbb{R}}$, we let $M_{\mathcal{H}}^{\mathbb{R}}$ denote $M_{\mathcal{H}}$ restricted to $\mathbb{R}$. The solutions in $M_{\mathcal{H}}^{\mathbb{R}}$ of the system of master functions does, in fact, correspond to the real solutions of the system from Equation (4.3). As we are particularly interested in *positive* real solutions, we need a notion corresponding to this in the realm of the system of master functions.

Definition 4.10. The *positive chamber* of the hyperplane complement $M_{\mathcal{H}}^{\mathbb{R}}$ is

$$\Delta_p = \{y \in \mathbb{R}^l \mid p_i(y) > 0, i \in [n+l+1]\}.$$

•

Finally, we can characterize the positive real solutions of our original system of polynomials.

Theorem 4.11. *The solutions in Δ_p (respectively $M_{\mathcal{H}}^{\mathbb{R}}$) of the system from Equation* (4.4) *correspond to positive real solutions (respectively real solutions) of the original system from Equation* (4.3).

We now present an example to make the process of Gale dualization clearer.

Example 4.12. We follow the steps of Algorithm 4.8. Let $n = 3$ and $l = 2$. Consider the system of polynomials

$$\begin{aligned}
x_2^2x_3^3 - 11x_1x_2x_3^3 - 33x_1x_2^2x_3 + 4x_2^2x_3 + 15x_1^2x_2 + 7 &= 0,\\
x_2^2x_3^3 \qquad\qquad + \;\; 5x_1x_2^2x_3 - 4x_2^2x_3 - \;\; 3x_1^2x_2 + 1 &= 0,\\
x_2^2x_3^3 - 11x_1x_2x_3^3 - 31x_1x_2^2x_3 + 2x_2^2x_3 + 13x_1^2x_2 + 8 &= 0,
\end{aligned}$$

which is, indeed, linearly independent. The support of this system is

$$\mathcal{A} = \{0, (0,2,3), (0,2,1), (1,1,3), (2,1,0), (1,2,1)\}.$$

By solving for $x_2^2x_3^3$, $x_2^2x_3$, and $x_1x_2x_3^3$ we get that

$$\begin{aligned}
x^{(0,2,3)} &= x_2^2x_3^3 = 1 - x_1^2x_2 - x_1x_2^2x_3,\\
x^{(0,2,1)} &= x_2^2x_3 = \frac{1}{2} - x_1^2x_2 + x_1x_2^2x_3,\\
x^{(1,1,3)} &= x_1x_2x_3^3 = \frac{10}{11}\left(1 + x_1^2x_2 - 3x_1x_2^2x_3\right).
\end{aligned}$$

So let

$$\begin{aligned} p_1(x_1^2x_2, x_1x_2^2x_3) &:= 1 - x_1^2x_2 - x_1x_2^2x_3, \\ p_2(x_1^2x_2, x_1x_2^2x_3) &:= \frac{1}{2} - x_1^2x_2 + x_1x_2^2x_3, \\ p_3(x_1^2x_2, x_1x_2^2x_3) &:= \frac{10}{11}\left(1 + x_1^2x_2 - 3x_1x_2^2x_3\right), \end{aligned}$$

$p_4(x_1^2x_2, x_1x_2^2x_3) := x_1^2x_2$, and $p_5(x_1^2x_2, x_1x_2^2x_3) := x_1x_2^2x_3$. As

$$\begin{aligned} (0,2,3) - 3(0,2,1) - (1,1,3) - (2,1,0) + 3(1,2,1) = 0, \\ 3(0,2,3) - (0,2,1) - 2(1,1,3) + 2(2,1,0) - 2(1,2,1) = 0, \end{aligned}$$

we have that

$$\begin{aligned} (x_1x_2^2x_3)^3(x_2^2x_3^3) &= (x_1^2x_2)(x_2^2x_3)^3(x_1x_2x_3^3), \\ (x_1^2x_2)^2(x_2^2x_3^3)^3 &= (x_1x_2^2x_3)^2(x_2^2x_3)(x_1x_2x_3^3)^2. \end{aligned}$$

Substituting in p_1, p_2, and p_3, we get that

$$\begin{aligned} (x_1x_2^2x_3)^3(1 - x_1^2x_2 - x_1x_2^2x_3) &= (x_1^2x_2)\left(\frac{1}{2} - x_1^2x_2 + x_1x_2^2x_3\right)^3\left(\frac{10}{11}\left(1 + x_1^2x_2 - 3x_1x_2^2x_3\right)\right), \\ (x_1^2x_2)^2(1 - x_1^2x_2 - x_1x_2^2x_3)^3 &= (x_1x_2^2x_3)^2\left(\frac{1}{2} - x_1^2x_2 + x_1x_2^2x_3\right)\left(\frac{10}{11}\left(1 + x_1^2x_2 - 3x_1x_2^2x_3\right)\right)^2. \end{aligned}$$

We let $y_1 = x_1^2x_2$ and $y_2 = x_1x_2^2x_3$, so we can consider the p_i's as polynomials in $y = (y_1, y_2)$. Then the above relations can be written as

$$\begin{aligned} p_5(y)^3p_1(y) &= p_4(y)p_2(y)^3p_3(y), \\ p_4(y)^2p_1(y)^3 &= p_5(y)^2p_2(y)p_3(y)^2. \end{aligned}$$

These equations are equivalent to the master functions

$$\begin{aligned} p(y)^{\beta_1} &= p_1(y)^1p_2(y)^{-3}p_3(y)^{-1}p_4(y)^{-1}p_5(y)^3 = 1, \\ p(y)^{\beta_2} &= p_1(y)^3p_2(y)^{-1}p_3(y)^{-2}p_4(y)^2p_5(y)^{-2} = 1, \end{aligned}$$

where $\beta_1 = (1, -3, -1, -1, 3)$ and $\beta_2 = (3, -1, -2, 2, -2)$. The equations $p_i(y) = 0$ define a hyperplane arrangement $\mathcal{H}$ in $\mathbb{C}^2$. Let $\mathcal{B} = \{\beta_1, \beta_2\} \subset \mathbb{Z}^{3+2}$. Then we get a system of master functions

$$p(y)^{\beta_1} = p(y)^{\beta_2} = 1$$

in $M_{\mathcal{H}}$. The system can be rearranged to the equivalent system of polynomials

$$\begin{aligned} q_1(y) &:= p_5(y)^3p_1(y) - p_4(y)p_2(y)^3p_3(y) = 0, \\ q_2(y) &:= p_4(y)^2p_1(y)^3 - p_5(y)^2p_2(y)p_3(y)^2 = 0. \end{aligned}$$

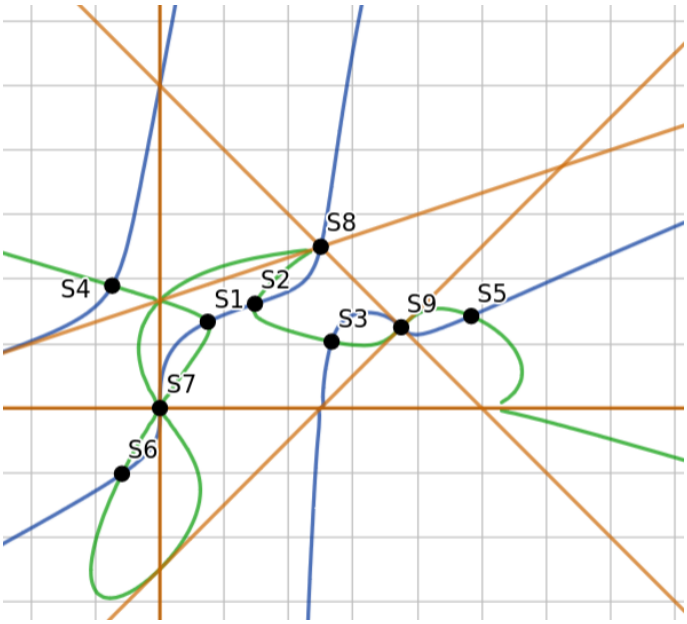


Figure 4.1: The hyperplane arrangement $\mathcal{H}^{\mathbb{R}}$ (orange) and the zero sets of q_1 (blue) and q_2 (green).

The pentagon area of Figure 4.1 is the positive chamber, Δ_p. From the figure, we see that the system of master functions has three solutions in Δ_p, six solutions in $M_{\mathcal{H}}^{\mathbb{R}}$, and nine solutions in all of $\mathbb{R}^2$. One can show that there are no solutions outside of the depicted area. The solutions in $M_{\mathcal{H}}^{\mathbb{R}}$ correspond to real solutions of the original system. In particular, the points in Δ_p correspond to the positive real solutions of the original system, by Theorem 4.11. We now consider the solutions in $\mathcal{H}^{\mathbb{R}}$ and explore their relation to the original system.

First, consider $S7 = (0, 0)$. Reverse engineering the process, we get that $0 = y_1 = x_1^2 x_2$ and $0 = y_2 = x_1 x_2^2 x_3$. Hence, by definition of the p_i's, we get that

$$\begin{aligned} x_2^2 x_3^3 &= 1, \\ x_2^2 x_3 &= \frac{1}{2}, \\ x_1 x_2 x_3^3 &= \frac{10}{11}. \end{aligned}$$

However, as $0 = x_1^2 x_2$, we must have either $x_1 = 0$ or $x_2 = 0$. This creates a contradiction. Hence, $S7$ does not correspond to any solution of the original system.

Similarly, for $S8 = \left(\frac{1}{2}, \frac{1}{2}\right)$, we get that

$$\begin{aligned} x_2^2 x_3^3 &= 1 - \frac{1}{2} - \frac{1}{2} = 0, \\ x_2^2 x_3 &= \frac{1}{2} - \frac{1}{2} + \frac{1}{2} = \frac{1}{2}, \\ x_1 x_2 x_3^3 &= \frac{10}{11}\left(1 + \frac{1}{2} - 3\frac{1}{2}\right) = 0. \end{aligned}$$

But $x_2^2 x_3^3 = 0$ requires either $x_2 = 0$ or $x_3 = 0$, which would imply $x_2^2 x_3 = 0$. Hence, $S8$ does not correspond to any solution of the original system. Analogously, $S9$ also does not correspond to any

solution of the original system. Hence, the solutions in $\mathcal{H}^{\mathbb{R}}$ do not correspond to any solutions of the original system.

By similar reverse engineering, we could find the solutions of the original system corresponding to each of the solutions in $M_{\mathcal{H}}^{\mathbb{R}}$ of the system of master functions. •

4.1.1 Special case: Circuits

In the next chapter, we will use Gale duality for the special case, where $\mathcal{A}$ is a circuit.

Definition 4.13. A *circuit* is a set $\mathcal{A} \subset \mathbb{Z}^n$ of $n+2$ exponent vectors, which affinely spans $\mathbb{R}^n$. •

Observe that the definition of a circuit does not require $\mathcal{A}$ to be primitive.

Remark 4.14. From Remark 3.6, the assumption that $0 \in \mathcal{A}$ is "free", as we already assumed that $\mathcal{A}$ affinely spans $\mathbb{R}^n$. It turns out that our assumption that $\mathcal{A}$ is primitive is also not needed.

Suppose $\mathcal{A} = \{a_0, \dots, a_{n+1}\}$ is not primitive, but that it still affinely spans $\mathbb{R}^n$. Let $A = \begin{pmatrix} a_0 & \dots & a_{n+1} \end{pmatrix} \in \mathbb{Z}^{n\times(n+2)}$. As A has full rank, there exists invertible matrices $S \in \mathbb{Z}^{n\times n}$ and $U \in \mathbb{Z}^{(n+2)\times(n+2)}$, such that the Smith normal form of A is

$$SAU = \begin{pmatrix} \alpha_1 & & & 0 & 0 & 0 \\ & \ddots & & \vdots & \vdots & \\ 0 & & \alpha_n & 0 & 0 & \end{pmatrix}.$$

Define the map $\zeta\colon \mathbb{C}^n \to \mathbb{C}^n$ as $\zeta(x) = x^{S^{-1}}$. Note that if $x \in \mathbb{R}^n_{>0}$, then $\zeta(x) \in \mathbb{R}^n_{>0}$. Similarly, if $y \in \mathbb{R}^n_{>0}$, then $\zeta^{-1}(y) = y^S \in \mathbb{R}^n_{>0}$. I.e., ζ restricts to a bijection on $\mathbb{R}^n_{>0}$. Given a coefficient matrix $C \in \mathbb{R}^{n\times(n+2)}$, we have that

$$0 = C \cdot x^A = C \cdot (y^S)^A = C \cdot y^{SA}.$$

It follows that the positive solutions of $C \cdot x^A = 0$ are in bijective correspondence with the positive solutions of $C \cdot y^{SA} = 0$. Let $\theta\colon \mathbb{C}^{n+2} \to \mathbb{C}^{n+2}$ be defined as $\theta(v) = v^U$. Similarly to the argument for ζ, we see that θ restricts to a bijection on $\mathbb{R}^{n+2}_{>0}$. Thus, the positive solutions of $C \cdot y^{SA} = 0$ are in bijective correspondence with the positive solutions of $C \cdot y^{SAU} = 0$. Note that

$$y^{SAU} = \begin{pmatrix} y^{\alpha_1} \\ \vdots \\ y^{\alpha_n} \\ 1 \\ 1 \end{pmatrix}.$$

Consider the coordinate change, where for $i \leq n$ we let $z_{a_i} = y^{\alpha_i}$, and for $i > n$ we let $z_{a_i} = 1$. This change is bijective and preserves solutions in $\mathbb{R}_{>0}$.

Putting everything together, we have now shown that transforming $\mathcal{A}$ to a primitive set preserves positive solutions. Therefore, the results of Gale duality concerning positive solutions also holds for non-primitive exponent sets, when we restrict to $\mathbb{R}$ and only consider positive solutions. •

When $\mathcal{A}$ is not primitive, the p_0 and P_0 from Remark 4.9 are no longer trivial, and they are included in the theory alongside the rest of the p_i's and P_i's. To make it easier to apply the results, we summarize Gale duality in this special case.

Proposition 4.15. *Let $\mathcal{A} = \{a_0, \ldots, a_{n+1}\} \subset \mathbb{Z}^n$ be a circuit, and let $C \in \mathbb{R}^{n\times(n+2)}$ be some coefficient matrix. Assume that* $\operatorname{rank}(C) = n$, *and* $\dim(\mathrm{NP}(\mathcal{A})) = n$. *Let $B = (b_0, \ldots, b_{n+1})^T$ be a Gale dual matrix of A, and let*

$$P = \begin{pmatrix} P_0 \\ \vdots \\ P_{n+1} \end{pmatrix} \in \mathbb{R}^{(n+2)\times 2}$$

be a Gale dual configuration of C, as in Remark 4.9. Finally, let $p_i(y) = \langle P_i, (1, y)\rangle$ for each $i \in [n+2]$. We define the function $g\colon \Delta_p \to \mathbb{R}$ by

$$g(y) = \prod_{j=0}^{n+1} p_j(y)^{b_j}.$$

Then the number of solutions (counted with multiplicity) of $g(y) = 1$ in Δ_p is equal to the number of positive solutions to the system $C \cdot x^A = 0$.

5 Descartes' rule of signs for circuits

Fix an ordered set of exponent vectors $\mathcal{A} = \{a_0, \dots, a_{n+1}\} \subset \mathbb{Z}^n$. Note that the cardinality of $\mathcal{A}$ is fixed at $|\mathcal{A}| = n + 2$. To put this in the context of the previous chapter, we have simply let $l = 1$, so there are $n + l + 1$ exponent vectors. Hence, we now let $C = (c_{i,j}) \in \mathbb{R}^{n\times(n+2)}$ be some coefficient matrix, and for any such matrix, we denote by

$$\begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = C \cdot x^A = 0 \tag{5.1}$$

the corresponding fewnomial system in n variables $x = (x_1, \dots, x_n)$ with support $\mathcal{A}$. Throughout, we let $n_{\mathcal{A}}(C)$ be the number of positive real solutions to the system in Equation (5.1). Define the matrix

$$A^+ = \begin{pmatrix} 1 & \dots & 1 \\ a_0 & \dots & a_{n+1} \end{pmatrix}.$$

Note that $\dim(\text{conv}(\mathcal{A})) = n$ is equivalent to having $\text{rank}(A^+) = n + 1$. Assume, henceforth, that $\text{rank}(C) = n$.

Remark 5.1. While the definition of a circuit is quite constraining, the results of this setting also apply in a more general setting. We still assume that $|\mathcal{A}| = n + 2$. Let $m = \dim(\text{conv}(\mathcal{A}))$. Then there exists some subconfiguration $\mathcal{A}' \subset \mathcal{A}$, such that $|\mathcal{A}'| = m+2$, and $\dim(\text{conv}(\mathcal{A}')) = m$. Without loss of generality, assume that $\mathcal{A}' = \{a_0, \dots, a_{m+1}\}$. Let $C \in \mathbb{R}^{n\times(n+2)}$ be some coefficient matrix with $\text{rank}(C) = n$, and let F denote the corresponding polynomial system, as in Equation (5.1). As C has full rank, we can find an equivalent system of the form

$$\begin{aligned} \tilde{f}_i &= \sum_{j=0}^{m+1} \tilde{c}_{i,j} x^{a_j}, \quad i = 1, \dots, m, \\ \tilde{f}_i &= \sum_{j=0}^{m+1} \tilde{c}_{i,j} x^{a_j} + x^{a_{i+1}}, \quad i = m+1, \dots, n, \end{aligned}$$

which has coefficient matrix $\tilde{C}$. Then, the system of the first m equations has either finitely many solutions, or the number of solutions to both systems is infinite. Since $\dim(\text{conv}(\mathcal{A})) = m$, we can consider the first m polynomials as a square system in m variables. Note that any solution $x \in \mathbb{R}^n_{>0}$ to F will correspond to a solution $x' \in \mathbb{R}^m_{>0}$ in the square system. Suppose $x' \in \mathbb{R}^m_{>0}$ is a solution to this square system. If x' extends to a solution $x \in \mathbb{R}^n_{>0}$ of the original system, then this extension is unique. Therefore, $n_{\mathcal{A}}(C) \leq n_{\mathcal{A}'}(C')$. Hence, we assume $m = n$. If $m \neq n$, then m should be written in place of n in the results in this chapter. •

By Remark 5.1, we can assume without losing generality that $\mathcal{A}$ is, in fact, a circuit. This is equivalent to adding the assumption that $\mathcal{A}$ is minimally affinely dependent, i.e., all maximal minors of A^+ are nonzero. Let $C_0, \dots, C_{n+1} \in \mathbb{R}^n$ denote the columns of C. Let $P \in \mathbb{R}^{(n+2)\times 2}$ be a Gale dual of C, i.e., P has columns, which are a basis of the kernel of C (see Remark 4.9). Let $P_0, \dots, P_{n+1} \in \mathbb{R}^2$ denote the rows of P. Then $\{P_0, \dots, P_{n+1}\}$ is a Gale dual configuration of the configuration $\{C_0, \dots, C_{n+1}\}$, and it is unique up to linear transformation.

Remark 5.2. Suppose we are given some coefficient matrix $C \in \mathbb{R}^{n\times(n+2)}$. If $x \in \mathbb{R}^n_{>0}$ is a positive solution to Equation (5.1), then $(x^{a_0}, \ldots, x^{a_{n+1}})$ must be in $\ker(C)$. This is equivalent to having the linear combination

$$x^{a_0}C_0 + \cdots + x^{a_{n+1}}C_{n+1} = 0,$$

i.e., to having $0 \in \operatorname{pos}(C_0, \ldots, C_{n+1})$. As the columns of P generate the kernel of C, any vector in $\ker(C)$ is of the form Pv for some $v \in \mathbb{R}^2$. Note that Pv is positive, if and only if, $\langle P_j, v\rangle > 0$ for all $j \in [n+2]$. It follows that the existence of a positive solutions to Equation (5.1) is equivalent to the existence of a vector $v \in \mathbb{R}^2$ satisfying $\langle P_j, v\rangle > 0$ for all $j \in [n+2]$. The existence of such a vector v also requires that all the P_j's lie in the same open halfspace. •

The condition outlined in the above remark is equivalent to $n_{\mathcal{A}}(C) > 0$. Therefore, in this chapter, we assume that this condition is satisfied. Let $C(i,j)$ denote the submatrix of C, where we have removed the i'th row and the j'th column. The following is a well-known linear algebra result.

Lemma 5.3. *The vectors P_i and P_j are colinear, if and only if, the maximal minor $\det(C(i,j))$ is zero.*

In the following definition, it is crucial that all P_j's lie in the same open half-space (see Remark 5.2).

Definition 5.4. Let C be of rank n and satisfying $0 \in \operatorname{pos}(C_0, \ldots, C_{n+1})$, and fix a Gale dual configuration $\{P_0, \ldots, P_{n+1}\}$. Define an equivalence relation $\sim$ on $[n+2]$ to be

$$i \sim j \iff \det(P_i, P_j) = 0.$$

Let k be the number of equivalence classes ($1 \le k \le n+2$). We denote by

$$[n+2]/\sim = \{K_0, \ldots, K_{k-1}\}$$

the set of equivalence classes, where the classes are ordered, such that for all $j_1, j_2 \in [k]$ and for any $i_1 \in K_{j_1}$ and $i_2 \in K_{j_2}$, we have that $\det(P_{i_1}, P_{i_2}) > 0$, if $j_1 < j_2$. Henceforth, we refer to this ordering as the *canonical ordering* of the equivalence classes. •

Example 5.5. Let $n = 6$. In Figure 5.1, we see an example of a configuration of P_i's. Clearly, $k = 4$, and the equivalence classes are $K_0 = \{2,7\}, K_1 = \{3\}, K_2 = \{1,4,0,6\}$, and $K_3 = \{5\}$. Here we number the equivalence classes by going counterclockwise around the unit circle from one "end" of the blue line to the other. This is what we in the above definition refer to as the canonical ordering of the equivalence classes. •

We pick some arbitrary ordering within each equivalence class and combine it with the canonical ordering of the equivalence classes to get a permutation $\sigma \in S_{n+2}$, such that

$$\det(P_{\sigma(i)}, P_{\sigma(j)}) \ge 0, \quad \text{for all } i < j.$$

Let $K \subset [n+2]$ be a set of representatives of the classes in $[n+2]/\sim$. Then σ induces a bijection $\tau\colon [k] \to K$, which associates each equivalence class to a representative of said equivalence class, such that $\det(P_{\tau(i)}, P_{\tau(j)}) > 0$.

Definition 5.6. We call σ an *ordering* for C and τ a *strict ordering* for C. •

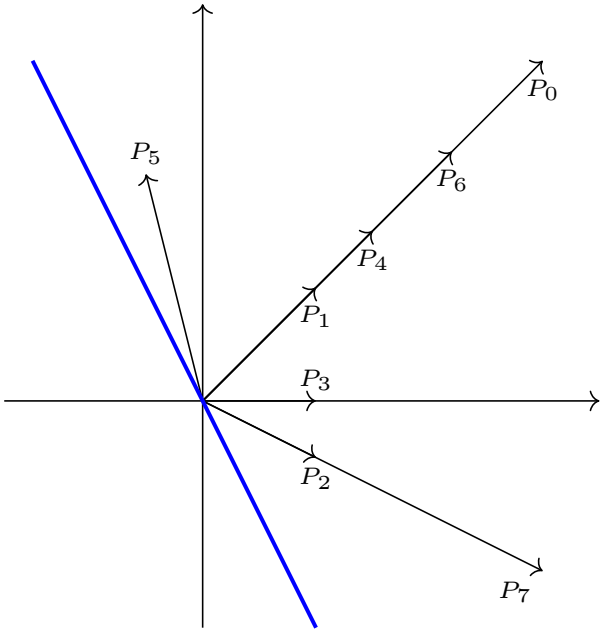


Figure 5.1: The configuration $\{P_0, \ldots, P_6\}$.

Example 5.7 (Continuation of Example 5.5)**.** Let $K = \{2, 3, 1, 5\}$ be a set of representatives of the classes in $[n+2]/\sim$. Choose $\sigma \in S_{n+2}$, such that within the equivalence classes, we order by magnitude of the vector. We get that e.g., $\det(P_{\sigma(1)}, P_{\sigma(4)}) = \det(P_2, P_1) > 0$, but $\det(P_{\sigma(1)}, P_{\sigma(2)}) = \det(P_2, P_7) = 0$. Hence, $\det(P_{\sigma(i)}, P_{\sigma(j)}) \geq 0$ for all $i < j$, so σ is an ordering. We can now define $\tau\colon [4] \to K$ to be

$$\tau(j) = \begin{cases} 2, & \text{if } j = 0, \\ 3, & \text{if } j = 1, \\ 1, & \text{if } j = 2, \\ 5, & \text{if } j = 3. \end{cases}$$

Clearly, $\det(P_{\tau(i)}, P_{\tau(j)}) > 0$ for all $i, j \in [4]$ with $i < j$. It follows that τ is a strict ordering. •

Let $B \in \mathbb{Z}^{(n+2)\times 1}$ be any matrix, which is Gale dual to A^+. This means that the column vector $b = (b_0, \ldots, b_{n+1})^T$ of B is a basis of $\ker(A^+)$. Hence, up to multiplication by some constant,

$$b_j = (-1)^j \det(A^+(j)),$$

where $A^+(j)$ is the matrix A^+ without the j'th column.

Definition 5.8. We keep the assumptions from Definition 5.4. Let $L_\ell = K_0 \cup \cdots \cup K_\ell$ for each $\ell \in [k]$. We define the terms $\lambda_\ell = \sum_{j \in K_\ell} b_j$ and $\mu_\ell = \sum_{j \in L_\ell} b_j$, which gives us the sequences $\lambda = (\lambda_\ell)_{\ell \in [k]}$ and $\mu = (\mu_\ell)_{\ell \in [k]}$. •

Remark 5.9. Observe that the K_i's are disjoint. We have that

$$\mu_\ell = \sum_{j \in L_\ell} b_j = \sum_{i \in [\ell+1]} \sum_{j \in K_i} b_j = \sum_{i \in [\ell+1]} \lambda_i.$$

In particular, $\mu_{k-1} = \sum_{i\in[k]} \lambda_i$. We also have that $\mu_{k-1} = \sum_{j\in L_{k-1}} b_j = \sum_{j\in[n+2]} b_j$. As $(1,\dots,1)$ is a row of A^+, and $(b_0,\dots,b_{n+1})^T$ is in $\ker(A^+)$, we have that $(1,\dots,1)(b_0,\dots,b_{n+1})^T = 0$. Thus $\mu_{k-1} = 0$. It also follows that $\lambda_0 + \cdots + \lambda_{k-2} = -\lambda_{k-1}$. Note that both λ and μ depend on τ even though it is not obvious from the notation. •

We are now finally ready to state a generalized form of Descartes' rule of signs.

Theorem 5.10 (Descartes' rule of signs for circuits, [BDF22, Theorem 2.4])**.** *Assume $\mathcal{A} \subset \mathbb{Z}^n$ to be a circuit. Let $C \in \mathbb{R}^{n\times(n+2)}$ be a coefficient matrix. We assume that C has rank n and satisfies $0 \in \operatorname{pos}(C_0,\dots,C_{n+1})$. Let τ be a strict ordering for C, and let μ be defined as in Definition 5.8. If $n_{\mathcal{A}}(C)$ is finite, then $n_{\mathcal{A}}(C) \le 1 + \operatorname{sgnvar}(\mu)$. In particular, we get that $n_{\mathcal{A}}(C) \le k - 1 \le n + 1$.*

Observe that the upper bound $n_{\mathcal{A}}(C) \le n+1$ is the same as the classical Descartes' rule of signs gives in the univariate case (see Corollary 2.9). Additionally, the part of Descartes' rule of signs, which says "and the difference is even" also applies here. If λ_0 and λ_{k-1} are both nonzero, then $1 + \operatorname{sgnvar}(\mu) - n_{\mathcal{A}}(C)$ is even (see [BDF22, Proposition 2.14]). We can, in fact, go even further. Just like Descartes' rule of signs, Theorem 5.10 is a sharp bound.

Proposition 5.11 ([BDF22, Theorem 3.4])**.** *Let $\mathcal{A} \subset \mathbb{Z}^n$ be a circuit, and let $\{K_0,\dots,K_{k-1}\}$ be a partition of $[n+2]$. Pick an arbitrary bijection $\tau\colon [k] \to K$. Then there exists a coefficient matrix $C \in \mathbb{R}^{n\times(n+2)}$ satifying*

1) $\operatorname{rank}(C) = n$,

2) *τ is a strict ordering for C,*

3) $0 \in \operatorname{pos}(C_0,\dots,C_{n+1})$,

and which obtains the bound from Theorem 5.10:

$$n_{\mathcal{A}}(C) = 1 + \operatorname{sgnvar}(\mu).$$

The proof of Proposition 5.11 is outside the scope of this thesis. To be able to prove Theorem 5.10, we need some results from linear algebra. Most are stated without proofs.

Definition 5.12. Consider a sequence $h = (h_1,\dots,h_s)$ of real valued analytic functions defined on some open interval $\Delta \subset \mathbb{R}$. We say that the *sequence h satisfies Descartes' rule of signs on Δ*, if for any sequence $a = (a_1,\dots,a_s)$ of real numbers, the number of roots of $\sum_{i=1}^s a_i h_i$ in Δ counted with multiplicity is at most $\operatorname{sgnvar}(a)$. •

Note that the ordering of the sequence is only significant up to reversal. So $(h_1,\dots,h_s)$ satisfies Descartes' rule of signs on Δ, if and only if, $(h_s,\dots,h_1)$ also does.

Definition 5.13. Let $h = (h_1,\dots,h_s)$ be a sequence of real valued analytic functions defined on some open interval $\Delta \subset \mathbb{R}$. The *Wronskian* of $h_1,\dots,h_s$ is

$$\operatorname{Wr}(h_1,\dots,h_s) = \det\begin{pmatrix} h_1 & \dots & h_s \\ h_1' & \dots & h_s' \\ \vdots & \ddots & \vdots \\ h_1^{(s-1)} & \dots & h_s^{(s-1)} \end{pmatrix}.$$

•

The following result is well-known.

Proposition 5.14. *The real valued analytic functions $h_1, \dots, h_s$ are linearly independent, if and only if, $\mathrm{Wr}(h_1, \dots, h_s)$ is not identically zero.*

Proposition 5.15 ([BDF22, Proposition 2.6])**.** *Let $h_1, \dots, h_s$ be a sequence of functions. Then $h_1, \dots, h_s$ satisfies Descartes' rule of signs on $\Delta \subset \mathbb{R}$, if and only if, the following conditions hold:*

1) *Linear independence: For all collections of indices $1 \leq j_1 < j_2 < \cdots < j_\ell \leq s$, we have $\mathrm{Wr}(h_{j_1}, \dots, h_{j_\ell}) \neq 0$.*
2) *Same sign: For any two collections of indices $1 \leq i_1 < \cdots < i_\ell \leq s$ and $1 \leq j_1 < \cdots < j_\ell \leq s$ of the same size, we have $\mathrm{Wr}(h_{i_1}, \dots, h_{i_\ell}) \cdot \mathrm{Wr}(h_{j_1}, \dots, h_{j_\ell}) > 0$.*

We also need the following technical result.

Proposition 5.16 ([BDF22, Proposition 2.7])**.** *Let $M \in \mathbb{R}^{t \times r}$ with $t \leq r$. Assume that for any $s \leq t$, there exists $\varepsilon_s \in \{-1, 1\}$, such that for any choice of s consecutive indices $j_1, \dots, j_1 + s - 1$, ε_s equals the sign of the determinant of the submatrix of M, which consists of the first s rows and the s consecutive columns with the chosen indices. Then, the sign of any $s \times s$ minor of M, which consists of the first s rows and any s columns, equals ε_s.*

Proof. For any $1 \leq s \leq t$, and any $J = \{j_1, \dots, j_s\}$ with $1 \leq j_1 < \dots, < j_s \leq r$, define M_J to be the submatrix of M, which consists of the first s rows and the columns (in order) with indices in J. We define the dispersion number $d(J)$ to be the number of integer points in $[j_1, j_s]$ different from $j_1, \dots, j_s$, i.e., the number of skipped indices. We observe that $d(J) = 0$ means that $j_1, \dots, j_s$ are consecutive integers.

We proceed by induction over $s \geq 1$ and the dispersion number $d \geq 0$. For $s = 1$, the result is immediate. The case of $d = 0$ follows from the assumption of the proposition. So assume that $s \geq 2$ and $d \geq 1$. Assume, by induction, that the result holds for $s' \leq s$ and $d' \leq d$ with at least one of the inequalities being strict.

Let $J = \{j_1, \dots, j_s\}$ with $1 \leq j_1 < \cdots < j_s \leq r$, and $d(J) = d$. As the dispertion number is at least one, there exists an index j satisfying $j_1 < j < j_s$. We have that $d(J \cup \{j\}) = d - 1$. By simply applying [Pin09, Equation (1.2)] to our situation, we get that

$$\det(M_J) \det(M_{\{j_2,\dots,j,\dots,j_{s-1}\}}) = \det(M_{\{j_2,\dots,j,\dots,j_s\}}) \det(M_{\{j_1,\dots,j_{s-1}\}}) + \det(M_{\{j_1,\dots,j,\dots,j_{s-1}\}}) \det(M_{\{j_2,\dots,j_s\}}).$$

As the second factor of each sum is the determinant of a matrix of size $s - 1$, we have, from the induction assumption, that all of these factors have the same sign ε_{s-1}. Note that $d((J \setminus \{j_1\}) \cup \{j\}) < d(J \cup \{j\}) < d(J)$, and analogously for j_s. Hence, by the induction assumption, $\det(M_{\{j_2,\dots,j,\dots,j_s\}})$ and $\det(M_{\{j_1,\dots,j,\dots,j_{s-1}\}})$ both have the same sign ε_s. Thus, $\det(M_J)$ must also have sign ε_s. ∎

Consider a coefficient matrix $C \in \mathbb{R}^{n \times (n+2)}$. Assume $\mathrm{rank}(C) = n$ and that $0 \in \mathrm{pos}(C_0, \dots, C_{n+1})$. Let $\{P_0, \dots, P_{n+1}\}$ be a Gale dual configuration of C. As $\{P_0, \dots, P_{n+1}\}$ is unique up to linear transformation, we can assume without loss of generality that the vector v from Remark 5.2 is of the form $v = (1, y)$ for some $y \in \mathbb{R}$. Hence, we get the polynomials

$$p_j(y) = \langle P_j, (1, y) \rangle, \quad j \in [n + 2].$$

Thus the p_j's are univariate polynomials of degree one, so the positive chamber is $\Delta_p = \{y \in \mathbb{R} \mid p_0(y), \dots, p_{n+1}(y) > 0\}$. Note that P_i and P_j being colinear is equivalent to p_i and p_j being proportional. As we assumed that $n_{\mathcal{A}}(C) > 0$, it follows from Remark 5.2 that there exists $y \in \mathbb{R}$, such that $p_j(y) > 0$ for all $j \in [n+2]$. Hence, Δ_p is a non-empty open interval.

Lemma 5.17. *The functions* $\left(\frac{p'_{\tau(\ell)}}{p_{\tau(\ell)}} - \frac{p'_{\tau(\ell+1)}}{p_{\tau(\ell+1)}} \;\middle|\; \ell \in [k-1]\right)$ *satisfy properties (1) and (2) in Proposition 5.15.*

Proof. By Proposition 5.16, it is enough to consider consecutive indices. Let $j_0 \in [k-1]$ be a starting index and let ℓ be the length of the selected subsequence. We want to show that the sequence

$$\left(\frac{p'_{\tau(j_0)}}{p_{\tau(j_0)}} - \frac{p'_{\tau(j_0+1)}}{p_{\tau(j_0+1)}},\ \frac{p'_{\tau(j_0+1)}}{p_{\tau(j_0+1)}} - \frac{p'_{\tau(j_0+2)}}{p_{\tau(j_0+2)}},\ \dots,\ \frac{p'_{\tau(j_0+(\ell-1))}}{p_{\tau(j_0+(\ell-1))}} - \frac{p'_{\tau(j_0+\ell)}}{p_{\tau(j_0+\ell)}}\right)$$

satisfies properties (1) and (2) from Proposition 5.15. We have that

$$\left(\frac{p'_j}{p_j}\right)' = \frac{p''_j p_j - p'_j p'_j}{p_j^2} = -\left(\frac{p'_j}{p_j}\right)^2,$$

and

$$\left(\frac{p'_j}{p_j}\right)'' = -2\left(\frac{p'_j}{p_j}\right)\left(\frac{p'_j}{p_j}\right)' = 2\left(\frac{p'_j}{p_j}\right)^3.$$

A straighforward induction proof shows that, for all $i, j \in [k-1]$, we have

$$\left(\frac{p'_j}{p_j}\right)^{(i-1)} = (-1)^{i-1}(i-1)!\left(\frac{p'_j}{p_j}\right)^{i+1}.$$

Hence,

$$\begin{aligned}
&\operatorname{Wr}\left(\frac{p'_{\tau(j_0)}}{p_{\tau(j_0)}} - \frac{p'_{\tau(j_0+1)}}{p_{\tau(j_0+1)}},\ \frac{p'_{\tau(j_0+1)}}{p_{\tau(j_0+1)}} - \frac{p'_{\tau(j_0+2)}}{p_{\tau(j_0+2)}},\ \dots,\ \frac{p'_{\tau(j_0+(\ell-1))}}{p_{\tau(j_0+(\ell-1))}} - \frac{p'_{\tau(j_0+\ell)}}{p_{\tau(j_0+\ell)}}\right) \\
&\quad = \det\left(\left((-1)^{i-1}(i-1)!\left(\left(\frac{p'_{\tau(j_0+(j-1))}}{p_{\tau(j_0+(j-1))}}\right)^{i+1} - \left(\frac{p'_{\tau(j_0+j)}}{p_{\tau(j_0+j)}}\right)^{i+1}\right)\right)_{i,j=1,\dots,\ell}\right) \\
&\quad = \left(\prod_{i=1}^{\ell}(-1)^{i-1}(i-1)!\right)\det\left(\left(\left(\frac{p'_{\tau(j_0+(j-1))}}{p_{\tau(j_0+(j-1))}}\right)^{i+1} - \left(\frac{p'_{\tau(j_0+j)}}{p_{\tau(j_0+j)}}\right)^{i+1}\right)_{i,j=1,\dots,\ell}\right).
\end{aligned}$$

Consider the Vandermonde determinant

$$\begin{aligned}\det(V) &= \det\begin{pmatrix} 1 & \dots & 1 \\ \frac{p'_{\tau(j_0)}}{p_{\tau(j_0)}} & \dots & \frac{p'_{\tau(j_0+\ell)}}{p_{\tau(j_0+\ell)}} \\ \vdots & \ddots & \vdots \\ \left(\frac{p'_{\tau(j_0)}}{p_{\tau(j_0)}}\right)^\ell & \dots & \left(\frac{p'_{\tau(j_0+\ell)}}{p_{\tau(j_0+\ell)}}\right)^\ell \end{pmatrix} \\ &= \det\begin{pmatrix} 0 & \dots & 0 & 1 \\ \frac{p'_{\tau(j_0)}}{p_{\tau(j_0)}} - \frac{p'_{\tau(j_0+1)}}{p_{\tau(j_0+1)}} & \dots & \frac{p'_{\tau(j_0+(\ell-1))}}{p_{\tau(j_0+(\ell-1))}} - \frac{p'_{\tau(j_0+\ell)}}{p_{\tau(j_0+\ell)}} & \frac{p'_{\tau(j_0+\ell)}}{p_{\tau(j_0+\ell)}} \\ \vdots & \ddots & \vdots & \vdots \\ \left(\frac{p'_{\tau(j_0)}}{p_{\tau(j_0)}}\right)^\ell - \left(\frac{p'_{\tau(j_0+1)}}{p_{\tau(j_0+1)}}\right)^\ell & \dots & \left(\frac{p'_{\tau(j_0+(\ell-1))}}{p_{\tau(j_0+(\ell-1))}}\right)^\ell - \left(\frac{p'_{\tau(j_0+\ell)}}{p_{\tau(j_0+\ell)}}\right)^\ell & \left(\frac{p'_{\tau(j_0+\ell)}}{p_{\tau(j_0+\ell)}}\right)^\ell \end{pmatrix},\end{aligned}$$

where the second equality is obtained by subtracting the second column from the first, the third from the second, all the way up to subtracting the ℓ'th column from the $(\ell-1)$'th column. By applying a Laplace expansion to the first row, we get that

$$\det(V) = (-1)^\ell \det\left(\left(\left(\frac{p'_{\tau(j_0+(j-1))}}{p_{\tau(j_0+(j-1))}}\right)^{i+1} - \left(\frac{p'_{\tau(j_0+j)}}{p_{\tau(j_0+j)}}\right)^{i+1}\right)_{i,j=1,\dots,\ell}\right).$$

It follows that

$$\begin{aligned}&\operatorname{Wr}\left(\frac{p'_{\tau(j_0)}}{p_{\tau(j_0)}} - \frac{p'_{\tau(j_0+1)}}{p_{\tau(j_0+1)}},\ \frac{p'_{\tau(j_0+1)}}{p_{\tau(j_0+1)}} - \frac{p'_{\tau(j_0+2)}}{p_{\tau(j_0+2)}},\ \dots,\ \frac{p'_{\tau(j_0+(\ell-1))}}{p_{\tau(j_0+(\ell-1))}} - \frac{p'_{\tau(j_0+\ell)}}{p_{\tau(j_0+\ell)}}\right) \\ &= \left(\prod_{i=1}^{\ell}(-1)^{i-1}(i-1)!\right)(-1)^\ell \det(V) = \left(\prod_{i=1}^{\ell}(-1)^{i-1}(i-1)!\right)(-1)^\ell \prod_{1\le j<i\le \ell+1}\left(\frac{p'_{\tau(i)}}{p_{\tau(i)}} - \frac{p'_{\tau(j)}}{p_{\tau(j)}}\right),\end{aligned}$$

where the last equality follows from the well-known formula for the Vandermonde determinant

$$\det(V) = \prod_{1\le j<i\le \ell+1}\left(\frac{p'_{\tau(i)}}{p_{\tau(i)}} - \frac{p'_{\tau(j)}}{p_{\tau(j)}}\right).$$

By definition of τ, we have that

$$\left(\frac{p'_{\tau(i)}}{p_{\tau(i)}} - \frac{p'_{\tau(j)}}{p_{\tau(j)}}\right) = \frac{p'_{\tau(i)}p_{\tau(j)} - p_{\tau(i)}p'_{\tau(j)}}{p_{\tau(i)}p_{\tau(j)}} = \frac{\det(P_{\tau(j)}, P_{\tau(i)})}{p_{\tau(i)}p_{\tau(j)}} > 0$$

for all $j > i$. In particular, $\left(\frac{p'_{\tau(i)}}{p_{\tau(i)}} - \frac{p'_{\tau(j)}}{p_{\tau(j)}}\right) \neq 0$, so the Wronskian does not vanish, which gives us property (1). The sign of the Wronskian only depends on ℓ, as the $\frac{p'_{\tau(i)}}{p_{\tau(i)}} - \frac{p'_{\tau(j)}}{p_{\tau(j)}}$ all have the same sign. This gives us property (2). ■

We can now finally prove Descartes' rule of signs for polynomial systems supported on circuits.

Proof of Theorem 5.10. Let $\mathcal{A}$, C and the strict ordering τ be as in the statement of Theorem 5.10. Fix a Gale dual configuration $\{P_0, \dots, P_{n+1}\}$ of $\{C_0, \dots, C_{n+1}\}$, such that

$$p_j(y) = \langle P_j, (1, y)\rangle > 0, \quad j \in [n+2].$$

In this proof, we consider

$$g(y) := p(y)^\beta = \prod_{j\in[n+2]} p_j(y)^{b_j},$$

where we recall the notation used in the definition of master functions in Construction 4.4. By Proposition 4.15, it is enough to bound the number of solutions of $g(y) = 1$ in Δ_p. As we assume $n_{\mathcal{A}}(C)$ to be finite, we must have $p(y)^\beta \not\equiv 1$, as otherwise there would be infinitely many solutions even in Δ_p.

We recall the partition $[n+2]/\sim = \{K_0, \dots, K_{k-1}\}$ defined earlier. For any $\ell \in [k]$, we have that $j \in K_\ell$, if and only if, $\tau(\ell) \sim j$, i.e., $\det(P_{\tau(\ell)}, P_j) = 0$. This is equivalent to there existing a constant $s_{j,\ell}$, such that $P_j = s_{j,\ell} P_{\tau(\ell)}$. As all the P_i's are in the same halfspace, we must have $s_{j,\ell} > 0$. Hence, we get that

$$\begin{aligned} g(y) &= \prod_{j\in[n+2]} p_j(y)^{b_j} = \prod_{\ell\in[k]} \prod_{j\in K_\ell} p_j(y)^{b_j} = \prod_{\ell\in[k]} \prod_{j\in K_\ell} (s_{j,\ell} p_{\tau(\ell)}(y))^{b_j} \\ &= \prod_{\ell\in[k]} p_{\tau(\ell)}^{\sum_{j\in K_\ell} b_j} \prod_{j\in K_\ell} s_{j,\ell}^{b_j} = \prod_{\ell\in[k]} p_{\tau(\ell)}^{\lambda_\ell} s_\ell = s \prod_{\ell\in[k]} p_{\tau(\ell)}^{\lambda_\ell}, \end{aligned}$$

where $s_\ell = \prod_{j\in K_\ell} e_{j,\ell}^{b_j}$, and $s = \prod_{\ell\in[k]} e_\ell$. We also have that

$$\prod_{\ell\in[k-1]} \frac{p_{\tau(\ell)}^{\mu_\ell}}{p_{\tau(\ell+1)}^{\mu_\ell}} = \frac{p_{\tau(0)}^{\lambda_0}}{p_{\tau(1)}^{\lambda_0}} \frac{p_{\tau(1)}^{\lambda_0+\lambda_1}}{p_{\tau(2)}^{\lambda_0+\lambda_1}} \cdots \frac{p_{\tau(k-3)}^{\lambda_0+\cdots+\lambda_{k-3}}}{p_{\tau(k-2)}^{\lambda_0+\cdots+\lambda_{k-3}}} \frac{p_{\tau(k-2)}^{\lambda_0+\cdots+\lambda_{k-2}}}{p_{\tau(k-1)}^{\lambda_0+\cdots+\lambda_{k-2}}} = \left(\prod_{\ell\in[k-1]} p_{\tau(\ell)}^{\lambda_\ell}\right) \frac{1}{p_{\tau(k-1)}^{\lambda_0+\cdots+\lambda_{k-2}}} = \prod_{\ell\in[k]} p_{\tau(\ell)}^{\lambda_\ell},$$

as $\lambda_{k-1} = -\sum_{i=0}^{k-2} \lambda_i$, by Remark 5.9. Hence, we get that

$$g(y) = s \prod_{\ell\in[k]} p_{\tau(\ell)}^{\lambda_\ell} = s \prod_{\ell\in[k-1]} \frac{p_{\tau(\ell)}^{\mu_\ell}}{p_{\tau(\ell+1)}^{\mu_\ell}}.$$

Let $G(y) = \log(g(y))$. Clearly, it is well-defined over Δ_p. For any $y \in \Delta_p$, we have that $g(y) = 1$, if and only if, $G(y) = 0$. As $\mu_{k-1} = 0$, we can rewrite $G(y)$ as

$$G(y) = \log(s) + \sum_{\ell\in[k-1]} \mu_\ell \left(\log(p_{\tau(\ell)}(y)) - \log(p_{\tau(\ell+1)}(y))\right).$$

Over Δ_p, we have that

$$G'(y) = \sum_{\ell\in[k-1]} \mu_\ell \left(\frac{p'_{\tau(\ell)}(y)}{p_{\tau(\ell)}(y)} - \frac{p'_{\tau(\ell+1)}(y)}{p_{\tau(\ell+1)}(y)}\right).$$

By Lemma 5.17, the sequence

$$\left(\frac{p'_{\tau(\ell)}}{p_{\tau(\ell)}} - \frac{p'_{\tau(\ell+1)}}{p_{\tau(\ell+1)}} \;\middle|\; \ell \in [k-1]\right)$$

of real valued analytic functions satisfies the conditions of Proposition 5.15. Hence, the number of roots of G' in Δ_p counted with multiplicity is at most $\operatorname{sgnvar}(\mu)$. By Rolle's Theorem, G must have at most $\operatorname{sgnvar}(\mu)+1$ roots in Δ_p counted with multiplicity. It follows that there are at most $\operatorname{sgnvar}(\mu)+1$ solutions to $g(y)=1$ in Δ_p counted with multiplicity. Therefore, by Gale duality, $n_{\mathcal{A}}(C) \leq \operatorname{sgnvar}(\mu)+1$. The sequence μ is k terms long, so $\operatorname{sgnvar}(\mu) \leq k-1$. But $\mu_{k-1}=0$, so, in fact, $\operatorname{sgnvar}(\mu) \leq k-2$. As $k \leq n+2$, we get that $n_{\mathcal{A}}(C) \leq 1+\operatorname{sgnvar}(\mu) \leq k-1 \leq n+1$. ■

To illustrate this new bound, we return to our example from the previous chapter.

Example 5.18 (Continuation of Example 4.2)**.** In Example 4.2, we found that the system

$$f_1 = x^{57} - xy + 1 = 0,$$
$$f_2 = y^{63} - xy + 2 = 0.$$

has at most 216 nondegenerate positive solutions. Notice that $\mathcal{A} = \{(0,0),(1,1),(57,0),(0,63)\}$ is, in fact, a circuit. So we can apply Theorem 5.10. We have that

$$A^+ = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 57 & 0 \\ 0 & 1 & 0 & 63 \end{pmatrix}.$$

The kernel of A^+ is spanned by the column vector $b = (1157, -1197, 21, 19)^T$. The coefficient matrix of our system is

$$C = \begin{pmatrix} 1 & -1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{pmatrix}.$$

We calculate that kernel of C, and find that

$$P = \begin{pmatrix} 1 & -1 \\ 2 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

From Figure 5.2, it is clear that the equivalence classes $[4]/\sim$ (with respect to the canonical ordering)

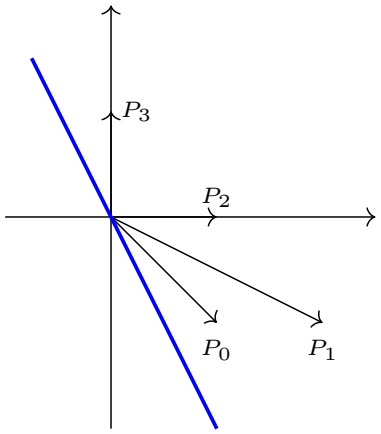


Figure 5.2: The vectors $P_0, \ldots, P_3$.

will be $K_i = \{i\}$ for $i \in [4]$. Hence, $\mu = (1157, -40, -19, 0)$. So $1+\operatorname{sgnvar}(\mu) = 2$. Thus, there are at most two positive solutions to our system, by Theorem 5.10. •

References

[BCR98] Jacek Bochnak, Michel Coste, and Marie-Francoise Roy. *Real Algebraic Geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer, 1998.

[BD16] Frédéric Bihan and Alicia Dickenstein. Descartes' Rule of Signs for Polynomial Systems supported on Circuits. 2016. arXiv:2010.09165v2.

[BDF22] Frédéric Bihan, Alicia Dickenstein, and Jens Forsgård. Optimal Descartes' Rule of Signs for Circuits. 2022. arXiv:1601.05826v2.

[Ber75] David N. Bernstein. The number of roots of a system of equations. *Functional Analysis and Its Applications*, 9:183–185, 1975.

[BKK76] David N. Bernstein, Anatoliy G. Kushnirenko, and Askold G. Khovanskii. Newton polytopes. *Uspehi Matematičeskih Nauk*, 31(3):201–202, 1976.

[Bud07] François Budan de Boislaurent. *Nouvelle Méthode pour la Résolution des Équations Numériques d'un degré quelconque*. 1807.

[Bé79] Étienne Bézout. *Théorie générale des équations algébriques*. 1779.

[CLO05] David A. Cox, John B. Little, and Donal O'Shea. *Using Algebraic Geometry*. Number 185 in Graduate Texts in Mathematics. Springer, 2nd edition, 2005.

[CLO15] David A. Cox, John Little, and Donal O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Undergraduate Texts in Mathematics. Springer, 4th edition, 2015.

[Des37] René Descartes. *La Géométrie*. 1637.

[Fou20] Jean-Baptiste Joseph Fourier. Sur l'usage du théorème de Descartes dans la recherche des limites des racines. *Bulletin des Sciences, Par la Société Philomatique de Paris*, pages 156–165, 1820.

[Gat14] Andreas Gathmann. Algebraic geometry, 2014. Class Notes TU Kaiserslautern.

[Kal25] Marie Kaltoft. Computer implementations used in master's thesis, 2025. `https://mariekaltoft.github.io/mastersthesis/`.

[Kho80] Askold G. Khovanskii. A class of systems of transcendental equations. *Doklady Akademii Nauk SSSR*, 255(4):804–807, 1980.

[Kur92] David C. Kurtz. A Sufficient Condition for All the Roots of a Polynomial To Be Real. *The American Mathematical Monthly*, 99(3):259–263, 1992.

[MV06] Hugh L. Montgomery and Robert C. Vaughan. *Multiplicative Number Theory I: Classical Theory*. Number 97 in Cambridge studies in advanced mathematics. Cambridge University Press, Cambridge, 1st edition, 2006.

[Nat19] Melvyn B. Nathanson. The Hermite-Sylvester criterion for real-rooted polynomials, 2019. arXiv:1911.01745v2.

[OSC25] The OSCAR Team. OSCAR – Open Source Computer Algebra Research system, Version 1.0.5, 2025.

[Pin09] Allan Pinkus. *Totally Positive Matrices*. Number 181 in Cambridge Tracts in Mathematics. Cambridge University Press, 2009.

[Sot11] Frank Sottile. *Real Solutions to Equations from Geometry*, volume 57 of *University Lecture Series*. American Mathematical Society, Providence, R.I, 2011.

[Stu29] Jacques Charles François Sturm. Mémoire sur la résolution des équations numériques. *Bulletin des Sciences de Férussac*, 11:419–425, 1829.