Discrete Comput Geom 34:475–495 (2005)
DOI: 10.1007/s00454-005-1179-x

Discrete & Computational
Geometry
© 2005 Springer Science+Business Media, Inc.

Some Bounds for the Number of Components of Real Zero Sets of Sparse Polynomials*

Daniel Perrucci

Departamento de Matemática,
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires,
Ciudad Universitaria, Pab. I (1428), Buenos Aires, Argentina
perrucci@dm.uba.ar

Abstract. We prove that the zero set of a 4-nomial in n variables in the positive orthant has at most three connected components. This bound, which does not depend on the degree of the polynomial, not only improves the best previously known bound (which was 10) but is optimal as well. In the general case we prove that the number of connected components of the zero set of an m-nomial in n variables in the positive orthant is lower than or equal to $(n+1)^{m-1}2^{1+(m-1)(m-2)/2}$, improving slightly the known bounds. Finally, we show that for generic exponents, the number of non-compact connected components of the zero set of a 5-nomial in three variables in the positive octant is at most 12. This strongly improves the best previously known bound, which was 10,384. All the bounds obtained in this paper continue to hold for real exponents.

1. Introduction

Descartes' Rule of Signs provides a bound for the number of positive roots of a given real univariate polynomial which depends on the number of sign changes among its coefficients but not on its degree. One of its consequences is that the number of positive roots of a polynomial with m monomials is bounded above by $m-1$.

Many attempts have been made to generalize Descartes' Rule of Signs (or its corollaries) to a larger class of functions. Even though this task has not yet been completed, important advances have been made [2]–[4], [8].

We introduce the notation and terminology we use throughout this paper. As usual, $\mathbb{N}$ denotes the set of positive integers. Let $n \in \mathbb{N}$. Given $x \in \mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_k > 0 \text{ for } 1 \le k \le n\}$ and $a := (a_1, \ldots, a_n) \in \mathbb{R}^n$, x^a denotes $x_1^{a_1} \cdots x_n^{a_n}$.

* This research was partially supported by the Argentinian Grant UBACyT 01-X198 and by CONICET.

Definition 1. Let $m \in \mathbb{N}$. An *m-nomial* in n variables is a function $f\colon \mathbb{R}^n_+ \to \mathbb{R}$ defined as

$$f(x) = \sum_{i=1}^{m} c_i x^{a_i},$$

where $c_i \in \mathbb{R}$, $c_i \neq 0$ and $a_i = (a_{i1}, a_{i2}, \ldots, a_{in}) \in \mathbb{R}^n$ for $i = 1, \ldots, m$.

An interesting fact is that Descartes' Rule of Signs continues to hold if one counts multiplicities and also if one allows real exponents (the adaptation of the proof given in Proposition 1.1.10 of [1], for instance, is straightforward).

Definition 2. Let $n, m \in \mathbb{N}$. We consider the functions $F\colon \mathbb{R}^n_+ \to \mathbb{R}^n$ of the form $F = (f_1, \ldots, f_n)$ with f_i an m_i-nomial, such that the total number of distinct exponent vectors in $f_1, \ldots, f_n$ is less than or equal to m. We then define $K(n,m)$ to be the maximum number of isolated zeros (in $\mathbb{R}^n_+$) an F of this type may have. Similarly, we define $K'(n,m)$ to be the maximum number of non-degenerate zeros (in $\mathbb{R}^n_+$) an F of this type may have.

A proof of the finiteness of $K(n,m)$ can be found in many sources, for instance, Corollary 4.3.8 of [1]. The finiteness of $K'(n,m)$ is a consequence of the fact that $K'(n,m)$ is always less than or equal to $K(n,m)$. A bound for $K'(n,m)$ is provided by Khovanski's theorem, which is the most important result in the theory of fewnomials:

Theorem 1. *Following the notations above*,

$$K'(n,m) \leq (n+1)^{m-1} 2^{(m-1)(m-2)/2}.$$

For a proof of Khovanski's theorem, see [1, Chapter 4], [3] or [4]. Nevertheless, the statement mentioned above is not exactly equal to any of those in the references. To prove Theorem 1 divide every equation in the system $F(x) = 0$ by x^a, where x^a ($a \in \mathbb{R}^n$) is one of the monomials of the system, to make the number of monomials drop and then use Theorem 4.1.1 [1] or Section 3.12, Corollary 6, of [4]. Another fact to be highlighted is that here we allow fewnomials with real exponents instead of integer exponents as in [1]. Nevertheless, the proof in the last reference does not make use of this fact.

Another way to generalize Descartes' Rule of Signs is to increase just the number of variables. In this case the problem is to find a bound for the number of connected components of the zero set of a single polynomial, which is expected to be a hypersurface. This paper is devoted to the study of this problem, both in particular cases and in the general one. The results presented here are inspired by a paper by Li et al. [6].

Definition 3. Given a subset X of $\mathbb{R}^n_+$, we denote by $\mathrm{Tot}(X)$, $\mathrm{Comp}(X)$ and $\mathrm{Non}(X)$ the number of connected components, compact connected components and non-compact connected components of X, respectively.

Given $n, m \in \mathbb{N}$, $P(n,m)$, $P_{\mathrm{comp}}(n,m)$ and $P_{\mathrm{non}}(n,m)$ are defined in the following way. First we define the set

$$\Omega(n,m) := \{f\colon \mathbb{R}^n_+ \to \mathbb{R} \mid f \text{ is a } k\text{-nomial with } 1 \leq k \leq m\}.$$

We then define

$$P(n,m) := \max\{\mathrm{Tot}(f^{-1}(0)) \mid f \in \Omega(n,m)\},$$

$$P_{\mathrm{comp}}(n,m) := \max\{\mathrm{Comp}(f^{-1}(0)) \mid f \in \Omega(n,m)\},$$

$$P_{\mathrm{non}}(n,m) := \max\{\mathrm{Non}(f^{-1}(0)) \mid f \in \Omega(n,m)\}.$$

It is clear from the definitions that, for all $n, m \in \mathbb{N}$,

$$P_{\mathrm{comp}}(n,m) \le P(n,m), \qquad P_{\mathrm{non}}(n,m) \le P(n,m),$$

$$P(n,m) \le P_{\mathrm{comp}}(n,m) + P_{\mathrm{non}}(n,m)$$

and that P, P_{comp} and P_{non} are increasing functions of their second parameter. For fixed $n, m \in \mathbb{N}$, the finiteness of $P(n,m)$ (and thus that of $P_{\mathrm{comp}}(n,m)$ and $P_{\mathrm{non}}(n,m)$) is a consequence of the fact that it is bounded from above by $n(n+1)^m 2^{n-1} 2^{m(m-1)/2}$ (see Corollary 2 of [6]). Strongly based on this paper, we derive a slightly better bound:

Theorem 2. *Using the previous notation*, $P(n,m) \le (n+1)^{m-1} 2^{1+(m-1)(m-2)/2}$.

Our approach is different from that in [6] in the way we bound the number of non-compact connected components. We state our result in the following theorem, which will also be useful in the last section, when dealing with 5-nomials:

Theorem 3. *Let us consider $m, n \ge 2$. If $Z := f^{-1}(0) \subset \mathbb{R}^n_+$ with f an m-nomial in n variables such that the dimension of the Newton polytope* (*see Definition* 4) *of f is n, then*

- $\mathrm{Non}(Z) \le 2n P(n-1, m-1)$,
- $\mathrm{Tot}(Z) \le \sum_{i=0}^{n-1} (2^i n!/(n-i)!) P_{\mathrm{comp}}(n-i, m-i)$.

We note that, due to the fact that $\mathbb{R}^n_+$ is not a closed set, a bounded connected component of the zero set of an m-nomial may be non-compact. This is the case, for example, when f is the 3-nomial in two variables defined by $f(x_1, x_2) = x_1^2 + x_2^2 - 1$.

The next proposition shows that, for a fixed number of monomials, a big number of variables will not increase the number of connected components:

Proposition 1 [6, Theorem 2]. *Given $m \in \mathbb{N}$, for all $n \in \mathbb{N}$,*

$$P(n,m) \le \begin{cases} m-1 & \text{if} \quad m \le 2, \\ P(m-2, m) & \text{if} \quad m \ge 3. \end{cases}$$

The reference given for the proposition above makes the additional assumption that $m \le n+1$. Nevertheless, the proof there does not make use of this fact. As we really need to eliminate this extra assumption, we give a brief proof of this proposition in the next section.

One of the goals of this paper is to find a sharp bound for $P(n, 4)$ and Proposition 1 shows that it is enough to find such a bound for $P(2, 4)$. Our result is stated in the following theorem:

Theorem 4. *Under the previous notation, we have:*

1. $P_{\text{comp}}(2, 4) = 1$.
2. $P_{\text{non}}(2, 4) = 3$.
3. $P(2, 4) = 3$ (*and thus* $P(n, 4) = 3$).
4. *If f is a* 4*-nomial in two variables and* $\dim \operatorname{Newt}(f) = 2$, *then* $\operatorname{Tot}(f^{-1}(0)) \leq 2$.

This theorem improves the best previously known bound for $P(n, 4)$, which was 10 [6, Theorems 2 and 3, and Example 2]. We state the results used to prove this last bound and sketch a brief proof of it in the next section. We note that in Theorem 3 of [6] the equality of the second item is proved in the smooth case.

The techniques we use to prove the previous theorems also allow us to prove the following theorem concerning 5-nomials.

Theorem 5. *Let f be a* 5*-nomial in three variables such that* $\dim \operatorname{Newt}(f) = 3$. *Let* $Z := f^{-1}(0) \subset \mathbb{R}^3_+$. *Then* $\operatorname{Non}(Z) \leq 12$.

This theorem significantly improves the best previously known bound of 10,384 (the proof of this bound is sketched briefly in the next section too).

This paper is organized as follows: Section 2 details some preliminaries. Section 3 concerns 4-nomials and contains the proof of Theorem 4. In Section 4 we deal with the general case of m-nomials in n variables and we prove Theorems 2 and 3. Finally, in Section 5, we prove Theorem 5.

2. Preliminaries

2.1. *Previously Known Bounds for Some Particular Cases*

The following result provides us with a bound for the number of non-degenerate roots in the positive quadrant for a fewnomial system having at most four different monomials.

Lemma 1 [6, Section 2, Proposition 1]. *Following the notation of Definition* 2, $K'(2, 4) \leq 5$.

The next theorem enables us to get a bound for the number of connected components in the positive orthant of the zero set of a single fewnomial.

Theorem 6 [6, Theorem 2]. *Following the notation of Definition* 3, *we have:*

- $P_{\text{comp}}(n, m) \leq 2\lfloor K'(n, m)/2 \rfloor \leq K'(n, m)$.
- $P_{\text{non}}(n, m) \leq 2P(n - 1, m)$.

With these results, we can easily prove that $P(n, 4) \leq 10$ in the following way:

$$P(n, 4) \leq P(2, 4) \leq P_{\text{comp}}(2, 4) + P_{\text{non}}(2, 4) \leq 2\lfloor K'(2, 4)/2 \rfloor + 2P(1, 4) \leq 10,$$

the last inequality being true because of Descartes' Rule of Signs. We improve this bound in Section 3.

In the same way,

$$\begin{aligned} P_{\text{non}}(3,5) &\leq 2P(2,5) \leq 2P_{\text{comp}}(2,5) + 2P_{\text{non}}(2,5) \\ &\leq 4\lfloor K'(2,5)/2 \rfloor + 4P(1,5) \leq 10{,}384. \end{aligned}$$

We improve this bound for the generic case in Section 5.

2.2. *Monomial Changes of Variables and Newton Polytopes*

We start this section with some notation and definitions.

Notation 1. Given a non-singular matrix $B \in \mathbb{R}^{n\times n}$, $B = (b_{ij})_{1\leq i,j\leq n}$, we denote by $B_1, \ldots, B_n$ the columns of B. We call the *monomial change of variables associated to* B the function

$$h_B\colon \mathbb{R}^n_+ \longrightarrow \mathbb{R}^n_+, \qquad h_B(x) = (x^{B_1}, \ldots, x^{B_n}).$$

The following formulae hold for all $x \in \mathbb{R}^n_+$, $a \in \mathbb{R}^n$ and non-singular matrices $B, C \in \mathbb{R}^{n\times n}$:

- $h_B(x)^a = x^{Ba}$.
- $h_B \circ h_C = h_{CB}$.

Recall the Newton polytope of a polynomial f, denoted by $\mathrm{Newt}(f)$, which is a convenient combinatorial encoding of the monomial term structure of a polynomial.

Definition 4. Given an m-nomial f in n variables, $f(x) := \sum_{i=1}^m c_i x^{a_i}$, $\mathrm{Newt}(f)$ denotes the smallest convex set containing the set of exponent vectors $\{a_1, \ldots, a_m\}$. The dimension of $\mathrm{Newt}(f)$, $\dim \mathrm{Newt}(f)$, is defined as the dimension of the smallest translated linear subspace containing $\mathrm{Newt}(f)$.

Therefore, for any n-variate m-nomial f, $\dim \mathrm{Newt}(f) \leq \min\{m-1, n\}$.

Given an m-nomial $f(x) = \sum_{i=1}^m c_i x^{a_i}$ and a non-singular matrix $B \in \mathbb{R}^{n\times n}$, we have that

$$f \circ h_B(x) = \sum_{i=1}^m c_i h_B(x)^{a_i} = \sum_{i=1}^m c_i x^{Ba_i},$$

and thus:

1. $f \circ h_B$ is also an m-nomial.
2. $\mathrm{Newt}(f \circ h_B) = \{Bv \in \mathbb{R}^n \mid v \in \mathrm{Newt}(f)\}$, and then, as B is non-singular, $\dim \mathrm{Newt}(f) = \dim \mathrm{Newt}(f \circ h_B)$.
3. As h_B is an analytic automorphism of the positive orthant, then the zero sets of f and $f \circ h_B$ have the same number of compact and non-compact connected components and critical points.

Remark 1. Given an m-nomial f in n variables, $c \in \mathbb{R}, c \neq 0$ and $b \in \mathbb{R}^n$, the function $c^{-1}x^{-b}f$ is an m-nomial whose Newton polytope is a translation of $\mathrm{Newt}(f)$. Then $\dim \mathrm{Newt}(c^{-1}x^{-b}f) = \dim \mathrm{Newt}(f)$. On the other hand, the zero set of $c^{-1}x^{-b}f$ (included in $\mathbb{R}^n_+$ by definition) is equal to the zero set of f (also included in $\mathbb{R}^n_+$). In particular, by choosing c as one of the coefficients of f, we will get an m-nomial with a coefficient equal to 1. Moreover, by choosing b as one of the exponents of f, we will get an m-nomial with a non-zero constant term. So, these particularities can be assumed without loss of generality and not modifying the zero set of the m-nomial, or the dimension of its Newton polytope. It can also be proved that $p \in f^{-1}(0) \subset \mathbb{R}^n_+$ is a critical point of f if and only if it is a critical point of $c^{-1}x^{-b}f$.

Proposition 2. *Let f be an m-nomial in n variables, let $Z := f^{-1}(0) \subset \mathbb{R}^n_+$ and let $d := \dim \mathrm{Newt}(f)$. Then*:

1. *If* $d \leq n-1$, *then* $\mathrm{Comp}(Z) = 0$ *and* $\mathrm{Non}(Z) \leq P(d, m)$.
2. *If* $d = m-1$, *then* $\mathrm{Comp}(Z) = 0$ *and* $\mathrm{Non}(Z) \leq 1$.

Proof. The proof can be done exactly as the proof of Theorem 2, Assertion 1, of [6]. For instance, to prove the first assertion, suppose $f(x) = c_1 + \sum_{i=2}^m c_i x^{a_i}$. Let us consider a non-singular matrix $B \in \mathbb{R}^{n\times n}$, such that the first d columns of B^{-1} are a basis of $\langle a_2, \ldots, a_m\rangle$. As each of the vectors Ba_i, $i = 2, \ldots, m$, has its $n-d$ last coordinates equal to zero, the m-nomial $f \circ h_B$ actually involves only d variables and its zero set may be described as $Z' \times \mathbb{R}^{n-d}_+$, where Z' is the zero set of an m-nomial in d variables. Thus, $\mathrm{Comp}(Z) = 0$ and $\mathrm{Non}(Z) \leq P(d, m)$. □

We can now give a proof of Proposition 1.

Proof of Propostion 1. Let f be an m-nomial in n variables, let $Z := f^{-1}(0) \subset \mathbb{R}^n_+$ and let $d := \dim \mathrm{Newt}(f)$. For $m \leq 2$, the proof is easy. If $m \geq 3$, as d is always less than or equal to $m-1$, then we just need to consider the following cases:

- If $1 \leq n \leq m-2$, then $\mathrm{Tot}(Z) \leq P(m-2, m)$, because an m-nomial in n variables can be considered as an m-nomial in $m-2$ variables with the particularity that the last $m-2-n$ variables are not actually involved in its formula.
- If $m-1 \leq n$ and $d \leq m-2$, then $d \leq n-1$. By Proposition 2, $\mathrm{Tot}(Z) \leq 0 + P(d, m) \leq P(m-2, m)$.
- If $m-1 \leq n$ and $d = m-1$, again by Proposition 2, $\mathrm{Tot}(Z) \leq 0 + 1 \leq P(m-2, m)$. □

Finally, we recall two classical results from topology that will be quite useful in the next section.

Theorem 7 (Connected Curve Classification). *Let Γ be a differentiable manifold of dimension* 1. *Then Γ is diffeomorphic either to S^1 or to $\mathbb{R}$ depending on whether Γ is compact or not.*

The proof of this theorem can be found in [7].

We also use the next adaptation of Jordan's lemma to the positive quadrant, which can be easily proved from its original statement (see, for example, [5]) upon an application of the exponential function.

Lemma 2 (Adaptation of Jordan's Lemma). *Let Γ be a curve in $\mathbb{R}^2_+$ homeomorphic to S^1. Then $\mathbb{R}^2_+ \backslash \Gamma$ has two connected components, which we call* $\mathrm{Int}(\Gamma)$ *and* $\mathrm{Ext}(\Gamma)$, *such that they are both open sets*, $\mathrm{Int}(\Gamma)$ *is bounded*, $\overline{\mathrm{Int}(\Gamma)} = \mathrm{Int}(\Gamma) \cup \Gamma$ *is compact and* $\mathrm{Ext}(\Gamma)$ *is unbounded.*

3. On 4-Nomials in Two Variables

Most of the results we obtain in this section come from the study of the restriction of 4-nomials in two variables to curves of the type $\{x \in \mathbb{R}^2_+ \mid x^a = J\}$ with $a \in \mathbb{R}^2$ and $J \in \mathbb{R}_+$. We introduce the notation we use.

Notation 2. Let $f\colon \mathbb{R}^2_+ \to \mathbb{R}$ be an m-nomial in two variables, let $p = (p_1, p_2) \in \mathbb{R}^2_+$ and let $u = (u_1, u_2) \in \mathbb{R}^2$, $u \neq 0$. By $h_{(p,u)}$ we denote the following parametrization of $\{x \in \mathbb{R}^2_+ \mid x^u = p^u\}$:

$$h_{(p,u)}\colon \mathbb{R}_+ \to \mathbb{R}^2_+,$$

$$h_{(p,u)}(t) = (h^{(1)}_{(p,u)}(t), h^{(2)}_{(p,u)}(t)) = \begin{cases} (t, (p^u)^{1/u_2} t^{-u_1/u_2}) & \text{if} \quad u_2 \neq 0, \\ (p_1, t) & \text{if} \quad u_2 = 0. \end{cases}$$

By $f_{(p,u)}$ we denote the following function:

$$f_{(p,u)}\colon \mathbb{R}_+ \to \mathbb{R}, \qquad f_{(p,u)} = f \circ h_{(p,u)}.$$

Remark 2.

- If $u_2 \neq 0$, $h_{(p,u)}(p_1) = p$ and if $u_2 = 0$, then $h_{(p,u)}(p_2) = p$.
- $f_{(p,u)}$ is an m'-nomial in one variable, with $m' \leq m$. The exponents of $f_{(p,u)}$ are proportional to the projections of the exponent vectors of f on $\langle u\rangle^\perp$. For instance, if $u_2 \neq 0$ and $a = (a_1, a_2)$ is an exponent of f, then $a_1 - u_1 a_2/u_2 = \langle a, (u_2, -u_1)\rangle u_2^{-1}$ is an exponent vector of $f_{(p,u)}$. The inequality $m' \leq m$ is due to the fact that different exponent vectors of f may have the same projection on $\langle u\rangle^\perp$, and so some monomials in $f_{(p,u)}$ may be re-grouped together and make the number of monomials decrease.
- Suppose $p = (p_1, p_2)$ is a critical point of f satisfying $f(p) = 0$ and $u = (u_1, u_2) \in \mathbb{R}^2$. If $u_2 \neq 0$, then p_1 is a degenerate zero of $f_{(p,u)}$, and if $u_2 = 0$, then p_2 is a degenerate zero of $f_{(p,u)}$. This is a consequence of the chain rule.

Notice that for $p \in \mathbb{R}^2_+$ and $u \in \mathbb{R}^2$, $u \neq 0$, the image of $h_{(p,u)}$ is an unbounded curve containing p. The following lemma gives us some information about the intersection between this curve and a compact connected component of the zero set of an m-nomial.

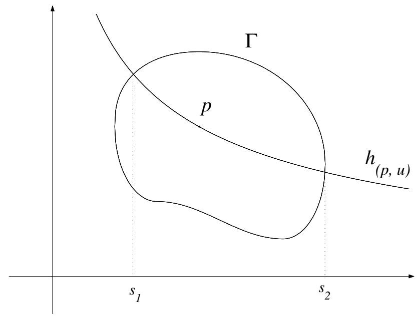

Fig. 1

Lemma 3. *Let f be an m-nomial in two variables and let $Z := f^{-1}(0) \subset \mathbb{R}^2_+$. Let Γ be a compact connected component of Z containing only regular points of f (so Γ is a differentiable submanifold of $\mathbb{R}^2_+$ diffeomorphic to S^1). Let $p = (p_1, p_2) \in \mathrm{Int}(\Gamma)$ and $u = (u_1, u_2) \in \mathbb{R}^2$, $u \neq 0$. Then, if $u_2 \neq 0$, $f_{(p,u)}$ has a zero $s_1 \in (0, p_1)$ and a zero $s_2 \in (p_1, +\infty)$ such that $h_{(p,u)}(s_i) \in \Gamma$ $(i = 1, 2)$. If $u_2 = 0$, $f_{(p,u)}$ has a zero $s_1 \in (0, p_2)$ and a zero $s_2 \in (p_2, +\infty)$ such that $h_{(p,u)}(s_i) \in \Gamma$ $(i = 1, 2)$ (Fig. 1).*

Proof. Let us suppose $u_2 \neq 0$. As Γ is a compact set and p lies in $\mathrm{Int}(\Gamma)$, there exist $x \in (0, p_1)$ and $y \in (p_1, \infty)$ such that both $h_{(p,u)}(x)$ and $h_{(p,u)}(y)$ lie in $\mathrm{Ext}(\Gamma)$ and the lemma follows. If $u_2 = 0$, a similar argument works. □

Suppose now that f is a 4-nomial in two variables. As explained before, by studying the restriction of f to curves of a certain type we obtain some information about its coefficients.

Lemma 4. *Let f be a 4-nomial in two variables and let $Z := f^{-1}(0) \subset \mathbb{R}^2_+$. Suppose that one of the following two conditions is satisfied:*

1. *Z has a critical point $p = (p_1, p_2)$ and $Z \backslash \{p\} \neq \emptyset$.*
2. *Z has a compact connected component Γ and $Z \backslash \Gamma \neq \emptyset$.*

Then two of the coefficients of f are positive and the other two are negative.

Proof. Suppose Z satisfies the first condition. Let $q = (q_1, q_2) \in Z \backslash \{p\}$.

If $p_1 \neq q_1$, then $p_1/q_1 \neq 1$. Let

$$v_1 := \frac{\log(q_2/p_2)}{\log(p_1/q_1)}.$$

Then $p_1^{v_1} p_2 = q_1^{v_1} q_2$. Let $v \in \mathbb{R}^2$, $v := (v_1, 1)$. As was explained in Remark 2, p_1 is a zero of $f_{(p,v)}$ with multiplicity at least 2. On the other hand,

$$f_{(p,v)}(q_1) = f(q_1, p^v q_1^{-v_1}) = f(q_1, q^v q_1^{-v_1}) = f(q) = 0,$$

because $q \in Z$. As $p_1 \neq q_1$, we know that $f_{(p,v)}$ has at least three zeros (counting multiplicities) in $\mathbb{R}_+$. We know that $f_{(p,v)}$ is an m'-nomial with $m' \leq 4$. By Descartes' Rule of Signs, we know that the number of sign changes in $f_{(p,v)}$ is at least three; thus, $m' = 4$ and among the four coefficients of $f_{(p,v)}$ there must be two positive and two negative. On the other hand, if

$$f(x) = \sum_{i=1}^{4} c_i x^{a_i},$$

then

$$f_{(p,v)}(x_1) = \sum_{i=1}^{4} c_i \, (p^v)^{a_{i2}} x_1^{a_{i1} - a_{i2} v_1}.$$

As the signs of the coefficients of $f_{(p,v)}$ are defined by the signs of the coefficients of f, then f must have two positive and two negative coefficients.

If $p_1 = q_1$, as $p \neq q$, we will have $p_2 \neq q_2$. In this case let us take $v := (1, 0)$ and proceed as above.

Let us suppose now that Z satisfies the second condition, which is having a compact connected component Γ, and $Z \neq \Gamma$. If Z has a critical point, then the first condition is also satisfied. If it does not have a critical point, we consider $\hat{p} := (\hat{p}_1, \hat{p}_2) \in \operatorname{Int}(\Gamma)$ and $\hat{q} := (\hat{q}_1, \hat{q}_2) \in Z \backslash \Gamma$.

If $\hat{p}_1 \neq \hat{q}_1$, in the same way as we did before, we can find a vector $w \in \mathbb{R}^2$, $w = (w_1, 1)$ such that $\hat{p}^w = \hat{q}^w$. Then $f_{(\hat{p},w)}$ has at least one zero s_1 in the interval $(0, \hat{p}_1)$ such that $h_{(\hat{p},w)}(s_1) \in \Gamma$ and at least one zero s_2 in the interval $(\hat{p}_1, +\infty)$ such that $h_{(\hat{p},w)}(s_2) \in \Gamma$. On the other hand,

$$f_{(\hat{p},w)}(\hat{q}_1) = f(\hat{q}_1, \hat{p}^w \hat{q}_1^{-w_1}) = f(\hat{q}_1, \hat{q}^w \hat{q}_1^{-w_1}) = f(\hat{q}) = 0,$$

because $\hat{q} \in Z$. Besides, due to the fact that $h_{(\hat{p},w)}(\hat{q}_1) = \hat{q} \in Z \backslash \Gamma$, $\hat{q}_1 \neq s_1$ and $\hat{q}_1 \neq s_2$. Then we deduce that $f_{(\hat{p},w)}$ has at least three zeros in $\mathbb{R}_+$, and then $f_{(\hat{p},w)}$ is also a 4-nomial with at least three sign changes. So, $f_{(\hat{p},w)}$ and f both have two coefficients with each sign.

If $\hat{p}_1 = \hat{q}_1$, then $\hat{p}_2 \neq \hat{q}_2$, and the same argument works. $\square$

Due to the lemma above, we focus our attention for a moment on 4-nomials with two coefficients of each sign. We start relating some properties of the zero set of a 4-nomial in two variables of this form with its Newton polytope.

Lemma 5. *Let f be a 4-nomial in two variables with two positive and two negative coefficients, such that* $\dim \operatorname{Newt}(f) = 2$. *Let $Z := f^{-1}(0) \subset \mathbb{R}^2_+$, and suppose one of the following two conditions is satisfied*:

1. *Z has a critical point $p = (p_1, p_2)$.*
2. *Z has a compact connected component Γ.*

Then $\operatorname{Newt}(f)$ *is a quadrilateral without parallel sides and coefficients corresponding to adjacent vertices have opposite signs.*

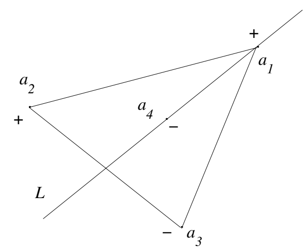

Fig. 2

Proof. Define $r = (r_1, r_2) \in \mathbb{R}_+^2$ as follows: if Z satisfies the first condition, then $r = p$ and if Z satisfies the second one but not the first one (so Γ is diffeomorphic to S^1), then r is any point in $\mathrm{Int}(\Gamma)$. By Remark 2 and Lemma 3, we know that for all $v = (v_1, v_2) \in \mathbb{R}^2$, $v \neq 0$, $f_{(r,v)}$ has at least two zeros (counting multiplicities) in $\mathbb{R}_+$.

Since $\dim \mathrm{Newt}(f) = 2$, the exponent vectors do not lie on a line. Suppose $f(x) = \sum_{i=1}^4 c_i x^{a_i}$. Then the vertices of $\mathrm{Newt}(f)$ are among the vectors a_1, a_2, a_3 and a_4 and $\mathrm{Newt}(f)$ might either be a triangle or a quadrilateral. We need to study four cases separately:

- Suppose $\mathrm{Newt}(f)$ is a triangle whose vertices are the vectors a_1, a_2 and a_3 and that the vector a_4 lies in the interior of $\mathrm{Newt}(f)$ (Fig. 2). Assume c_1 and c_2 are positive and c_3 and c_4 are negative (by multiplying f by -1 and re-ordering the monomials if necessary). Let $v := a_1 - a_4 \neq 0$ and let L be the line through a_1 and a_4. As a_1 and a_4 have the same projection on $\langle v \rangle^\perp$ and a_2 and a_3 are on opposite sides of the line L, we conclude that $f_{(r,v)}$ is a 3-nomial of the following type (if $v_2 \neq 0$):

$$\begin{aligned} f_{(r,v)}(x_1) \; = \; & c_3 (r^v)^{a_{32}/v_2} x_1^{a_{31} - a_{32} v_1/v_2} \\ & + (c_1 (r^v)^{a_{12}/v_2} + c_4 (r^v)^{a_{42}/v_2}) x_1^{a_{41} - a_{42} v_1/v_2} \\ & + c_2 (r^v)^{a_{22}/v_2} x_1^{a_{21} - a_{22} v_1/v_2}. \end{aligned}$$

  Even though we do not know at this point if in the above formula the terms are written in increasing or decreasing order, this 3-nomial has exactly one sign change, because the monomials of higher and lower exponent have distinct coefficient signs (we are supposing $c_2 > 0$ and $c_3 < 0$). By Descartes' Rule of Signs, it cannot have two zeros (counting multiplicities) as we know it does. Then we have a contradiction, and we conclude that the Newton polytope of f cannot be a triangle having the remaining exponent vector in its interior. If $v_2 = 0$, the same procedure works.
- Let us suppose now that $\mathrm{Newt}(f)$ is a triangle whose vertices are the exponent vectors a_1, a_2 and a_3; and that the vector a_4 lies on one of the edges of $\mathrm{Newt}(f)$. Without loss of generality, we suppose that a_4 lies on the segment $a_1 a_2$ (Fig. 3). By taking again $v := a_1 - a_4$, we have that a_1, a_2 and a_4 have the same projection on $\langle v \rangle^\perp$. Thus, $f_{(r,v)}$ is a 2-nomial (because its first, second and fourth term can be re-grouped together in a single monomial) and, by Descartes' Rule of Signs, $f_{(r,v)}$

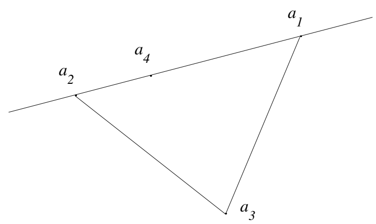

Fig. 3

cannot have two zeros (counting multiplicities) as we know it should. We then have a contradiction, which enables us to eliminate this case.

- Suppose $\mathrm{Newt}(f)$ is a quadrilateral with a pair of parallel opposite sides. Without loss of generality, we suppose that the segments a_1a_2 and a_3a_4 are parallel (Fig. 4). Let us take $v := a_1 - a_2$. As a_1 and a_2 have the same projection on $\langle v\rangle^{\perp}$, and a_3 and a_4 also do so, we can re-group the monomials in $f_{(r,v)}$ and form a 2-nomial, which again is impossible.
- Finally, suppose that $\mathrm{Newt}(f)$ is a quadrilateral and that coefficients of the same sign correspond to adjacent vertices. Without loss of generality, let us suppose that a_1 and a_2 are adjacent, a_3 and a_4 are adjacent too, c_1 and c_2 are positive and c_3 and c_4 are negative. Let $v := a_1 - a_2$ and let L be the line through a_1 and a_2 (Fig. 5). As a_1 and a_2 have the same projection on $\langle v\rangle^{\perp}$, then $f_{(r,v)}$ is a 3-nomial. However, as the two remaining exponent vectors (both corresponding to negative coefficients) lie in the same side of L, $f_{(r,v)}$ has just one sign change. For this reason, it cannot have two zeros, and we get a contradiction.

We conclude that the lemma follows. $\square$

The next lemma shows the existence of a convenient change of variables for certain bivariate 4-nomials.

Lemma 6. *Let f be a bivariate 4-nomial having two positive coefficients and two negative coefficients and such that* $\mathrm{Newt}(f)$ *is a quadrilateral with no parallel opposite sides and coefficients corresponding to adjacent vertices having opposite signs. Then*

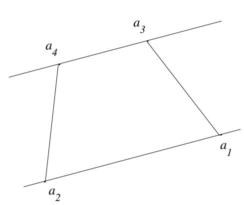

Fig. 4

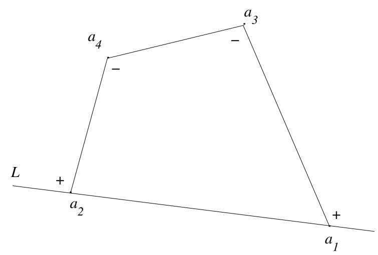

Fig. 5

there is an invertible change of variables h such that $f \circ h$ *is*

$$f \circ h(x_1, x_2) = 1 - x_1 - x_2 + Ax_1^c x_2^d,$$

with $A > 0$, $c, d > 1$, *and h is the composition of a monomial change of variables with a re-scaling of the variables.*

Proof. Suppose that we enumerate the vertices of $\mathrm{Newt}(f)$ in such a way that a_1 and a_4 are not adjacent. Because of Remark 1 we can suppose f is of the following type:

$$f(x_1, x_2) = 1 + \sum_{i=2}^{4} c_i x^{a_i},$$

i.e., $a_1 = (0, 0)$. As coefficients with the same sign correspond to non-adjacent vertices of $\mathrm{Newt}(f)$, we know that $c_2, c_3 < 0$ and $c_4 > 0$. Consider the four triangles that can be formed with three of the four vertices of $\mathrm{Newt}(f)$. Among these triangles there must be one having the minimal area. Suppose it is the triangle $a_1a_2a_3$ (this can be enforced by rotating indices if necessary). Because of the fact that $\mathrm{Newt}(f)$ does not have parallel opposite sides, this area is strictly less than the area of the triangles $a_1a_2a_4$ and $a_1a_3a_4$.

Let $B := \{a_2, a_3\}$ and let C be the matrix having the elements of B as columns. As $\mathrm{Newt}(f)$ is a quadrilateral, $a_1 = (0, 0)$, a_2 and a_3 do not lie on a line. Then B is a basis of $\mathbb{R}^2$ and C is non-singular. As in Lemma 1 of [6], let h be the composition of $h_{C^{-1}}$ and the linear re-scaling $(x_1, x_2) \mapsto (x_1/|c_2|, x_2/|c_3|)$. Let $(c, d) := C^{-1}a_4$. Then

$$f \circ h(x_1, x_2) = 1 - x_1 - x_2 + c_4 \frac{1}{|c_2|^c} \frac{1}{|c_3|^d} x_1^c x_2^d.$$

Let A be the last coefficient of the 4-nomial above. Then $A > 0$. On the other hand, the Newton polytope $\mathrm{Newt}(f \circ h_{C^{-1}})$ must also be a quadrilateral with vertices $(0, 0)$, $(1, 0)$, $(0, 1)$ and (c, d). As a_1 and a_4 are opposite vertices in $\mathrm{Newt}(f)$, then $(0, 0)$ and (c, d) must be opposite vertices in $\mathrm{Newt}(f \circ h_{C^{-1}})$. Thus, $c, d > 0$. As the area of triangle $a_1a_2a_3$ is smaller than that of triangle $a_1a_2a_4$, the area of triangle $(0, 0)(1, 0)(0, 1)$ should be smaller than that of triangle $(0, 0)(1, 0)(c, d)$, and thus $d > 1$. In an analogous way, we can prove that $c > 1$. $\square$

We recall that, as h is a diffeomorphism of $\mathbb{R}_+^2$, the zero sets of f and $f \circ h$ have the same number of compact and non-compact connected components and critical points.

The following lemma lets us deal with the case when the zero set of the 4-nomial has a critical point.

Lemma 7. *Let f be a 4-nomial in two variables such that* $\dim \mathrm{Newt}(f) = 2$ *and let* $Z := f^{-1}(0) \subset \mathbb{R}_+^2$. *Suppose that $p = (p_1, p_2) \in Z$ is a critical point of f, and also that $Z \setminus \{p\} \neq \emptyset$. Then Z is a connected non-compact set; that is to say,* $\mathrm{Non}(Z) = 1$ *and* $\mathrm{Comp}(Z) = 0$.

Proof. The proof of this lemma is done in four steps. In the first one we make use of the lemmata we proved before to make sure that it is enough to restrict our attention to 4-nomials of a very specific form. In the second one we study the sign of f on some curves we consider. In the third one we use the information we obtained to characterize a non-compact connected set W where f vanishes. In the last one we prove that, in fact, $W = Z$.

Step 1. By Lemma 4, we know that among the coefficients of f there must be two positive and two negative ones, and, by Lemma 5, $\mathrm{Newt}(f)$ must be a quadrilateral without parallel opposite sides, and with same sign coefficients corresponding to opposite vertices. Then, by Lemma 6, we can suppose f is of the following type:

$$f(x_1, x_2) = 1 - x_1 - x_2 + A x_1^c x_2^d,$$

with $A > 0$ and $c, d > 1$.

Step 2. Let $v := (c, d - 1)$. This vector has the nice property that $(0, 1)$ and (c, d), which are exponent vectors in f, have the same projection on $\langle v \rangle^\perp$. Because of this fact, $f_{(p,v)}$ is a 3-nomial. In fact,

$$f_{(p,v)}(x_1) = (-(p^v)^{1/(d-1)} + A\,(p^v)^{d/(d-1)}) x_1^{-c/(d-1)} + 1 - x_1.$$

Notice that $-c/(d-1) < 0$ because $c, d > 1$. By Remark 2, p_1 is a zero of $f_{(p,v)}$ of multiplicity greater than or equal to 2. By Descartes' Rule of Signs, the 3-nomial $f_{(p,v)}$ must have at least two sign changes, p_1 is a zero of multiplicity exactly 2 and $f_{(p,v)}$ does not have other zeros. As the unique zero of $f_{(p,v)}$ has an even multiplicity, and its leading exponent coefficient is negative, we know that $f_{(p,v)}(x_1) \leq 0$ for all $x_1 \in \mathbb{R}_+$ and $f_{(p,v)}(x_1) < 0$ if $x_1 \neq p_1$.

As $(0, 0)$ and $e_2 := (0, 1)$ are both vector exponents in f, it can be proved in an analogous way that $f_{(p,e_2)}(x_1) \geq 0$ for all $x_1 \in \mathbb{R}_+$ and $f_{(p,e_2)}(x_1) > 0$ if $x_1 \neq p_1$. Moreover,

$$f_{(p,e_2)}(x_1) = f(x_1, (p^{e_2})^1 x_1^0) = (1 - p_2) - x_1 + A p_2^d x_1^c,$$

and as it has a zero of multiplicity equal to 2, it has two sign changes and then we deduce that $p_2 < 1$. In the same way we can also prove that $p_1 < 1$.

It can easily be checked that for $x_1 \in (0, p_1)$, $h^{(2)}_{(p,e_2)}(x_1) < h^{(2)}_{(p,v)}(x_1)$, and for $x_1 \in (p_1, +\infty)$, $h^{(2)}_{(p,e_2)}(x_1) > h^{(2)}_{(p,v)}(x_1)$. To illustrate the situation, in Fig. 6 we have drawn the curves $h_{(p,e_2)}$ and $h_{(p,v)}$ indicating the sign of f on them.

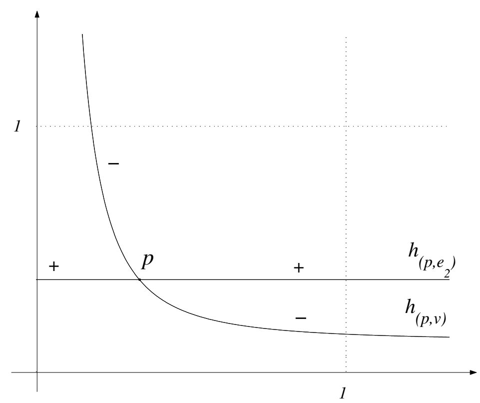

Fig. 6

Finally, for a fixed $\alpha \in \mathbb{R}_+$, we analyze the function $f(\alpha, x_2)$ in the variable x_2:

$$f(\alpha, x_2) = (1 - \alpha) - x_2 + A\alpha^c x_2^d.$$

Notice that for every fixed $\alpha \in (0, 1)$, $\lim_{x_2 \to 0+} f(\alpha, x_2) = 1 - \alpha > 0$.

Step 3. In order to study how many times the line $\{x_1 = \alpha\}$ intersects Z for a fixed $\alpha \in \mathbb{R}_+$, we continue studying the function $f(\alpha, x_2)$. As $A\alpha^c > 0$ and $d > 1$, if $\alpha < 1$, this function is a 3-nomial with two sign changes. Because of Descartes' Rule of Signs, it will have either no zeros or two (counted with multiplicity) in $\mathbb{R}_+$. If $\alpha = 1$, this function is a 2-nomial with just one sign change and, finally, if $\alpha > 1$, it is a 3-nomial with one sign change. In both cases it has exactly one zero in $\mathbb{R}_+$.

For a fixed $\alpha \in (0, p_1)$ the function (in the variable x_2) $f(\alpha, x_2)$ must have an odd number of zeros (counted with multiplicity) in the interval $(h^{(2)}_{(p,e_2)}(\alpha), h^{(2)}_{(p,v)}(\alpha))$. As it has at most two zeros in $\mathbb{R}_+$, then it has just one zero in that interval. We call it $g(\alpha)$.

In an analogous way, for a fixed $\alpha \in (p_1, 1)$ the function $f(\alpha, x_2)$ must have at least one zero in the interval $(0, h^{(2)}_{(p,v)}(\alpha))$, which we call $t(\alpha)$, and another one in the interval $(h^{(2)}_{(p,v)}(\alpha), h^{(2)}_{(p,e_2)}(\alpha))$, which we call $g(\alpha)$. As this function has at most two zeros in $\mathbb{R}_+$, then it has no other zeros.

For a fixed $\alpha \in [1, +\infty)$, the function $f(\alpha, x_2)$ must have an odd number of zeros (counted with multiplicity) in the interval $(h^{(2)}_{(p,v)}(\alpha), h^{(2)}_{(p,e_2)}(\alpha))$. As this function has at most one zero in $\mathbb{R}_+$, then it has just one zero in that interval. Again, we call it $g(\alpha)$.

Finally, we define $g(p_1) = p_2$, and we prove that the function $g\colon \mathbb{R}_+ \to \mathbb{R}_+$ we have just defined is continuous. As p is a critical point of f, we know that

$$\frac{\partial f}{\partial x_2}(p) = -1 + dAp_1^c p_2^{d-1} = 0,$$

and this implies that $p^v = 1/dA$.

Suppose there exists $x_1 \in \mathbb{R}_+$, $x_1 \neq p_1$ such that $(\partial f/\partial x_2)(x_1, g(x_1)) = 0$; then

$$g(x_1) = (1/dA)^{1/(d-1)} x_1^{-c/(d-1)} = (p^v)^{1/(d-1)} x_1^{-c/(d-1)} = h^{(2)}_{(p,v)}(x_1),$$

and this is impossible because of the definition of g. Then, for all $x_1 \neq p_1$, $(\partial f/\partial x_2)(x_1, g(x_1)) \neq 0$.

Let us fix $\alpha \neq p_1$ and see that g is continuous in α. Suppose that $\alpha > p_1$ (if $\alpha < p_1$ the proof can be done in the same way). We know that $h^{(2)}_{(p,v)}(\alpha) = (p^v)^{1/d-1}\alpha^{-c/d-1} < g(\alpha) < p_2 = h^{(2)}_{(p,e_2)}(\alpha)$ and $(\partial f/\partial x_2)(\alpha, g(\alpha)) \neq 0$. Then, by the Implicit Function Theorem, there is a continuous function, we call it s, defined in an interval $(\alpha-\varepsilon, \alpha+\varepsilon)$ with $\alpha - \varepsilon > p_1$, such that $s(\alpha) = g(\alpha)$ and $f(x_1, s(x_1)) = 0$ for all x_1 in the interval of definition. Moreover, choosing a suitable value of ε, we can suppose that, for all x_1 in $(\alpha-\varepsilon, \alpha+\varepsilon)$, $s(x_1)$ lies in $((p^v)^{1/(d-1)}x_1^{-c/(d-1)}, p_2)$. As $x_2 = g(x_1)$ is the unique value in this interval such that $f(x_1, x_2) = 0$, we have $g \equiv s$ in $(\alpha - \varepsilon, \alpha + \varepsilon)$ and therefore g is continuous in α.

To prove that g is continuous in p_1, notice that if $x_1 > p_1$, then

$$(p^v)^{1/(d-1)}x_1^{-c/(d-1)} < g(x_1) < p_2$$

and

$$\lim_{x_1 \to p_1^+} (p^v)^{1/(d-1)}x_1^{-c/(d-1)} = (p^v)^{1/(d-1)}p_1^{-c/(d-1)} = p_2.$$

So, we have $\lim_{x_1 \to p_1^+} g(x_1) = p_2$. Analogously, we prove that $\lim_{x_1 \to p_1^-} g(x_1) = p_2$.

Now, let us consider $w \in \mathbb{R}^2$, $w := (c-1, d)$. In the same way that we proved the existence of the function g, we can prove that there exists a function $k\colon \mathbb{R}_+ \to \mathbb{R}_+$ satisfying the following properties:

- For all positive x_2, we have $f(k(x_2), x_2) = 0$.
- k is continuous.
- If $x_2 < p_2$ then $k(x_2) \in (p_1, (p^w)^{1/(c-1)}x_2^{-d/(c-1)})$, and $x_1 = k(x_2)$ is the unique value in that interval such that $f(x_1, x_2) = 0$.
- $k(p_2) = p_1$.
- If $x_2 > p_2$ then $k(x_2) \in ((p^w)^{1/(c-1)}x_2^{-d/(c-1)}, p_1)$, and $x_1 = k(x_2)$ is the unique value in that interval such that $f(x_1, x_2) = 0$.

We define $W_1 = \{(x_1, g(x_1)) \mid x_1 \in \mathbb{R}_+\} \subset \mathbb{R}^2_+$, $W_2 = \{(k(x_2), x_2) \mid x_2 \in \mathbb{R}_+\} \subset \mathbb{R}^2_+$ and $W = W_1 \cup W_2$. As the functions g and k are continuous, W_1 and W_2 are connected. As $g(p_1) = p_2$ and $k(p_2) = p_1$, it follows that $p \in W_1 \cap W_2$, and then W is connected. Moreover, it is an unbounded set.

Step 4. We prove now that $W = Z$, and, therefore, that $\mathrm{Non}(Z) = 1$ and $\mathrm{Comp}(Z) = 0$.

Due to the fact that, for all x_1 and x_2 in $\mathbb{R}_+$, $f(x_1, g(x_1)) = 0$ and $f(k(x_2), x_2) = 0$, it is clear that $W \subset Z$. Let $q := (q_1, q_2) \in Z$.

Suppose that $q_1 < p_1$ and $q_2 < p_2$. Let

$$z_1 := \frac{\log(q_2/p_2)}{\log(p_1/q_1)}.$$

So, $p_1^{z_1}p_2 = q_1^{z_1}q_2$. As $p_1/q_1 > 1$ and $q_2/p_2 < 1$, then $z_1 < 0$. Let $z \in \mathbb{R}^2$, $z := (z_1, 1)$. We know that p_1 is a zero of multiplicity at least 2 of $f_{(p,z)}$. On the other hand,

$$f_{(p,z)}(q_1) = f(q_1, p^z q_1^{-z_1}) = f(q_1, q^z q_1^{-z_1}) = f(q_1, q_2) = 0,$$

because $q \in Z$. Then $f_{(p,z)}$ has at least three zeros (counted with multiplicity) and, by Descartes' Rule of Signs, at least three sign changes. As $c, d > 1$ and $z_1 < 0$, then $0 < -z_1 < c - dz_1$ and $0 < 1 < c - dz_1$, and we have that

$$f_{(p,z)}(x_1) = 1 - x_1 - p^z x_1^{-z_1} + A(p^z)^d x_1^{c-dz_1}$$

has just two sign changes. Then it cannot happen that $q_1 < p_1$ and $q_2 < p_2$ at the same time.

Suppose now that $q_1 \geq p_1$. Consider the following cases:

- $q_1 \geq 1$: as we have shown at the beginning of this lemma, the line $\{x_1 = q_1\}$ intersects Z in a single point, Which is $(q_1, g(q_1))$. Then it must be $q_2 = g(q_1)$ and then $q \in W_1$.
- $p_1 < q_1 < 1$: we know that the line $\{x_1 = q_1\}$ intersects Z in two points: $(q_1, g(q_1))$ and $(q_1, t(q_1))$, with $g(q_1) \in ((p^v)^{1/(d-1)} x_1^{-c/(d-1)}, p_2)$ and $t(q_1) \in (0, (p^v)^{1/(d-1)} q_1^{-c/(d-1)})$. If $q_2 = g(q_1)$, then $q \in W_1$. If $q_2 = t(q_1)$, then

  $$q_2 < (p^v)^{1/(d-1)} q_1^{-c/(d-1)} = (p_1/q_1)^{c/d-1} p_2 < p_2,$$

  and therefore $x_1 = k(q_2)$ is the unique value of x_1 in the interval $(p_1, (p^w)^{1/(c-1)} q_2^{-d/(c-1)})$ such that $f(x_1, q_2) = 0$. Since the previous inequalities imply

  $$q_1 < p_1 p_2^{(d-1)/c} q_2^{-(d-1)/c} < p_1 p_2^{d/(c-1)} q_2^{-d/(c-1)} = (p^w)^{1/(c-1)} q_2^{-d/c-1},$$

  we conclude that $q_1 = k(q_2)$ and so $q \in W_2$.
- If $q_1 = p_1$, we see that $\{x_1 = q_1\}$ intersects Z only in p. Let us consider $e_1 = (1, 0)$. We know that p_2 is a zero of multiplicity at least 2 of $f_{(p,e_1)}$, but

  $$f_{(p,e_1)}(x_2) = (1 - p_1) - x_2 + A p_1^c x_2^d$$

  is a 3-nomial with two sign changes. By Descartes' Rule of Signs, p_2 has multiplicity equal to 2 and $f_{(p,e_1)}$ has no other zeros. Then $\{x_1 = q_1\} \cap Z = \{p\}$, and then $q = p \in Z$.

If $q_2 \geq p_2$, we proceed in an analogous way. Thus, we conclude that $Z = W$, and that Z has a unique connected component, which is unbounded. □

We can now give a proof of the following theorem.

Theorem 8. *Let f be a 4-nomial in two variables and let $Z := f^{-1}(0) \subset \mathbb{R}^2_+$. If Z has a compact connected component Γ, then $Z = \Gamma$.*

Proof. Suppose $Z \backslash \Gamma \neq \emptyset$. By Lemma 4, we know that among the coefficients of f there are two positive and two negative. On the other hand, we know that $\dim \mathrm{Newt}(f) = 2$, otherwise Z could not have compact connected components. Then, by Lemma 5, $\mathrm{Newt}(f)$ is a quadrilateral without parallel sides and coefficients of the same sign correspond to opposite vertices. By Lemma 7, Z does not have critical points; otherwise it would have only a unique non-compact connected component. Finally, because of Lemma 6, we can suppose f is of the following type:

$$f(x_1, x_2) = 1 - x_1 - x_2 + Ax_1^c x_2^d,$$

with $A > 0$; $c, d > 1$.

Again, in order to study how many times the line $\{x_1 = \alpha\}$ intersects Z for a fixed α in $\mathbb{R}_+$, we define a function g_α in the variable x_2 as the restriction of f to that line, i.e.,

$$g_\alpha(x_2) = f(\alpha, x_2).$$

Then

$$g'_\alpha(x_2) = -1 + Ad\alpha^c x_2^{d-1} < 0 \qquad \Longleftrightarrow \qquad x_2 < \left(\frac{1}{Ad}\right)^{1/(d-1)} \alpha^{-c/(d-1)}.$$

Let $J := (1/Ad)^{1/(d-1)}$. Then $J > 0$ and the function g_α has a minimum in $x_2 = J\alpha^{-c/(d-1)}$. For $x_1 \in \mathbb{R}_+$, let $\ell_1(x_1)$ be the minimum of the function g_{x_1} and let $\ell(x_1) := (x_1, \ell_1(x_1))$. Then

$$f \circ \ell(x_1) = (-J + AJ^d)x_1^{-c/(d-1)} + 1 - x_1,$$

so $f \circ \ell$ turns out to be a 3-nomial.

As Γ is a compact set, the function x_1 reaches its minimum (we call it m) and its maximum (we call it M) on Γ. Let us prove that $m \neq M$: as Γ is a differentiable manifold of dimension 1, then Γ has an infinite number of points. If $m = M$, then $\Gamma \subset \{x_1 = m\}$, and the 3-nomial g_m has infinitely many zeros, which is impossible.

Let $p := (p_1, p_2)$ and $q := (q_1, q_2)$ in Γ such that $m = p_1$ and $M = q_1$. Let us see that p_2 is a zero of multiplicity 2 of g_{p_1}. As $p \in \Gamma$, $g_{p_1}(p_2) = f(p_1, p_2) = 0$. On the other hand, as p is the minimum of x_1 in Γ, using Lagrange multipliers,

$$g'_{p_1}(p_2) = \frac{\partial f}{\partial x_2}(p) = 0.$$

By Descartes' Rule of Signs, we know that g_{p_1} must have at least two sign changes. As we know that

$$g_{p_1}(x_2) = (1 - p_1) - x_2 + Ap_1^c x_2^d,$$

then it must be $p_1 < 1$, and so g_{p_1} has no zeros other than p_2. As the unique zero of g_{p_1} has an even multiplicity and its leading coefficient is positive, for all $x_2 \neq p_2$, $g_{p_1}(x_2) > 0$. Then $f \circ \ell(p_1) = 0$. In an analogous way, we can prove that q_2 is the unique zero of the function g_{q_1} and $f \circ \ell(q_1) = 0$. We conclude that $p_1 = m$ and $q_1 = M$ are two different zeros of $f \circ \ell$.

As $f \circ \ell(x_1) = (-J + AJ^d)x_1^{-c/(d-1)} + 1 - x_1$ is a 3-nomial, it has no zeros other than m and M which have multiplicity 1. As its leading coefficient is negative, we know that $f \circ \ell(x_1) < 0$ for all $x_1 \in (0, m) \cup (M, +\infty)$ and $f \circ \ell(x_1) > 0$ for all $x_1 \in (m, M)$. Let $s \in (m, M)$. Then, for all $x_2 \in \mathbb{R}_+$, $f(s, x_2) \geq f \circ \ell(s) > 0$. Then $\Gamma \cap \{x_1 = s\} = \emptyset$ and the open sets $\{x_1 < s\}$ and $\{x_1 > s\}$ disconnect Γ, which is a contradiction. $\square$

Now we can give a proof of Theorem 4, which is the main goal of this section.

Proof of Theorem 4. 1. The inequality $P_{\mathrm{comp}}(2,4) \le 1$ is a consequence of Theorem 8. In the following example the equality holds:

$$f_1(x_1,x_2) = x_2^2 - 4x_1^3x_2 + x_1^8 + 3x_1^4.$$

In fact, $f_1(x_1,x_2) = 0$ if and only if $x_2 = 2x_1^3 \pm x_1^2\sqrt{1-(x_1^2-2)^2}$, and the set of positive values of x_1 where the polynomial under the square root symbol is non-negative is the interval $[1,\sqrt{3}]$.

2. Let f be a 4-nomial in two variables and let $Z := f^{-1}(0) \subset \mathbb{R}^2_+$.

If $\dim \mathrm{Newt}(f) = 1$, then, by Proposition 2 and Descartes' Rule of Signs, we know that $\mathrm{Non}(Z) \le P(1,4) \le 3$.

If $\dim \mathrm{Newt}(f) = 2$ and 0 is a regular value of f, then, by Theorem 3 of [6], $\mathrm{Non}(Z) \le 2$.

If $\dim \mathrm{Newt}(f) = 2$ and 0 is not a regular value of f, there is a critical point p in Z. If $Z = \{p\}$, then $\mathrm{Non}(Z) = 0$. If $Z \ne \{p\}$, then, by Lemma 7, $\mathrm{Non}(Z) = 1$.

The equality holds in the following example:

$$f_2(x_1,x_2) = (x_1-1)(x_1-2)(x_1-3) = x_1^3 - 6x_1^2 + 11x_1 - 6.$$

3. Let f be a 4-nomial in two variables, and let $Z := f^{-1}(0) \subset \mathbb{R}^2_+$.

If Z has any compact connected component Γ, by Theorem 8, $Z = \Gamma$ and then $\mathrm{Tot}(Z) = 1$. If it does not, because of the previous item we have that $\mathrm{Tot}(Z) \le 3$ and the same example shows that the equality holds.

4. Let f be a 4-nomial in two variables such that $\dim \mathrm{Newt}(f) = 2$, and let $Z := f^{-1}(0) \subset \mathbb{R}^2_+$.

If Z has any compact connected component Γ, again by Theorem 8, $Z = \Gamma$ and then $\mathrm{Tot}(Z) = 1$. If it does not, as was shown in the second item of this theorem, $\mathrm{Non}(Z) \le 2$, and $\mathrm{Tot}(Z) \le 2$.

The equality holds in the following example:

$$f_3(x_1,x_2) = x_1x_2 - 2x_1 - x_2 + 1.$$

In fact, $f(x_1,x_2) = 0$ is an implicit equation for the hyperbola $x_2 = 1/(x_1-1)+2$. □

4. On *m*-Nomials in *n* Variables

In this section we prove Theorems 2 and 3. Theorem 2 gives us an explicit upper bound for the number of connected components of the zero set of an m-nomial in n variables in the positive orthant and Theorem 3 is an auxiliary theorem for Theorem 2, but it is also used in the next section. We now give a proof of Theorem 3.

Proof of Theorem 3. As observed earlier in Remark 1, we can assume f is of the following form:

$$f(x) = c_m + \sum_{i=1}^{m-1} c_i x^{a_i}.$$

As $\dim\langle a_1, \ldots, a_{m-1}\rangle = n$, without loss of generality, we can suppose that $B := \{a_1, \ldots, a_n\}$ is a basis of $\mathbb{R}^n$. Let A be the matrix having the elements of B as columns, let g be the m-nomial $f \circ h_{A^{-1}}$ and let $W := g^{-1}(0) \subset \mathbb{R}^n_+$. As $h_{A^{-1}}$ is a diffeomorphism, $\mathrm{Non}(W) = \mathrm{Non}(Z)$ and $\dim \mathrm{Newt}(g) = \dim \mathrm{Newt}(f)$. Moreover, we have that

$$g(x) = c_m + \sum_{i=1}^{m-1} c_i x^{A^{-1}a_i} = c_m + \sum_{i=1}^{n} c_i x_i + \sum_{i=n+1}^{m-1} c_i x^{A^{-1}a_i}.$$

Suppose W has t non-compact connected components and let $\{p_1, \ldots, p_t\}$ be a set of points intersecting each and every non-compact connected component of W. Suppose, for $1 \le i \le t$, $p_i = (p_{i1}, \ldots, p_{in})$. For $1 \le j \le n$, we consider $M_j, m_j \in \mathbb{R}_+$ such that $M_j > \max\{p_{ij},\ 1 \le i \le t\}$ and $m_j < \min\{p_{ij},\ 1 \le i \le t\}$, and define $S_j = \{x \in \mathbb{R}^n_+ \mid x_j = M_j\}$ and $T_j = \{x \in \mathbb{R}^n_+ \mid x_j = m_j\}$. Let us prove that each non-compact connected component of W intersects at least one of the sets $S_1, \ldots, S_n, T_1, \ldots, T_n$.

Let X be a non-compact connected component of W. If X is not bounded, then there exists j_0, $1 \le j_0 \le n$, such that $X \cap S_{j_0}$ is not empty.

If X is bounded, then it is not closed. Let $T := \bigcap_{j=1}^{n} \{x \in \mathbb{R}^n_+ \mid x_j \ge m_j\}$. If $X \subseteq T$, then it is a connected component of $W \cap T$. As $W = g^{-1}(0) \subset \mathbb{R}^n_+$ and g is a continuous function, there exists a closed set $F \subset \mathbb{R}^n$ such that $W = F \cap \mathbb{R}^n_+$. Then

$$W \cap T = F \cap \mathbb{R}^n_+ \cap T = F \cap T,$$

and $W \cap T \subset \mathbb{R}^n_+$ is closed because it is an intersection of closed sets. It follows that X is closed because it is a connected component of a closed set. This is a contradiction, and then $X \nsubseteq T$, and this implies that there exists j_1, $1 \le j_1 \le n$, such that $X \cap T_{j_1} \ne \emptyset$.

In this way we have found $2n$ sets $(S_1, \ldots, S_n, T_1, \ldots, T_n)$ such that each non-compact connected component of W has a non-empty intersection with one of them. Thus,

$$\mathrm{Non}(W) \le \sum_{j=1}^{n} \mathrm{Tot}(W \cap S_j) + \sum_{j=1}^{n} \mathrm{Tot}(W \cap T_j).$$

Each of these $2n$ intersections has at most $P(n-1, m-1)$ connected components, because they can be regarded as zero sets of m'-nomials in $n-1$ variables, with $1 \le m' \le m-1$. For example, the set $W \cap S_n$ can be described as the zero set of the following function:

$$\hat{g}\colon \mathbb{R}^{n-1}_+ \to \mathbb{R},$$

$$\hat{g}(x_1, \ldots, x_{n-1}) = (c_m + c_n M_n) + \sum_{i=1}^{n-1} c_i x_i + \sum_{i=n+1}^{m-1} c_i (x_1, \ldots, x_{n-1}, M_n)^{A^{-1}a_i}.$$

We have thus proved that

$$\mathrm{Non}(Z) = \mathrm{Non}(W) \le 2n\, P(n-1, m-1),$$

which is our first assertion.

Finally, note that for the function $\hat{g}$ defined above, $\dim \mathrm{Newt}(\hat{g}) = n - 1$. Proceeding inductively, we get the second inequality. $\square$

We prove now Theorem 2.

Proof of Theorem 2. We proceed by induction on n.

If $n = 1$ by Descartes' Rule of Signs, we know that

$$P(1, m) \leq m - 1 < 2^{m-1} 2^{1+(m-1)(m-2)/2}.$$

Suppose now that $n > 1$. Given an m-nomial f in n variables, let $d := \dim \mathrm{Newt}(f)$ and $Z := f^{-1}(0) \subset \mathbb{R}^n_+$.

If $d < n$, by the first item of Proposition 2 and the induction hypothesis,

$$\mathrm{Tot}(Z) \leq P(d, m) \leq (d+1)^{m-1} 2^{1+(m-1)(m-2)/2} \leq (n+1)^{m-1} 2^{1+(m-1)(m-2)/2}.$$

If $d = n$, as $m - 1 \geq d$, we have that $m \geq n + 1$. If $m = n + 1$, by the second item of Proposition 2, $\mathrm{Tot}(Z) \leq 1$. If $m \geq n + 2$, by the second item of Theorem 3, the first item of Theorem 6 and Theorem 1, we have

$$\mathrm{Tot}(Z) \leq \sum_{i=0}^{n-1} 2^i \frac{n!}{(n-i)!} (n - i + 1)^{m-i-1} 2^{(m-i-1)(m-i-2)/2}.$$

Now we use the following inequality, valid for all $i, n, m \in \mathbb{N}$ such that $m \geq n + 2$ and $0 \leq i \leq n - 1$, that can be easily proved by induction on i:

$$2^i \frac{n!}{(n-i)!} (n - i + 1)^{m-i-1} 2^{(m-i-1)(m-i-2)/2} \leq \frac{1}{2^i} (n+1)^{m-1} 2^{(m-1)(m-2)/2}.$$

Then we conclude that

$$\mathrm{Tot}(Z) \leq \sum_{i=0}^{n-1} \frac{1}{2^i} (n+1)^{m-1} 2^{(m-1)(m-2)/2} < (n+1)^{m-1} 2^{1+(m-1)(m-2)/2},$$

which completes the proof. □

5. On 5-Nomials in Three Variables

As a consequence of what has been proved in the previous sections, we get Theorem 5:

Proof of Theorem 5. By Theorem 3, $\mathrm{Non}(Z) \leq 6P(2, 4) = 18$. Nevertheless, in the proof of that theorem, we have shown the existence of six 4-nomials in two variables, we call them $g_1, \ldots, g_6$, such that for $i = 1, \ldots, 6$, $\dim \mathrm{Newt}(g_i) = 2$ and

$$\mathrm{Non}(Z) \leq \sum_{i=1}^{6} \mathrm{Tot}(g_i^{-1}(0)).$$

By the fourth item of Theorem 4, we know that $\mathrm{Tot}(g_i^{-1}(0)) \leq 2$, and then we conclude that $\mathrm{Non}(Z) \leq 12$. □

This bound is significantly sharper than the best previously known one, which was 10,384.

Acknowledgments

The author thanks Teresa Krick and Juan Sabia for their guidance in this work, Fernando López García for his valuable cooperation, Maurice Rojas for some nice talks and Gabriela Jeronimo and the referees for their very helpful corrections and suggestions.

References

1. R. Benedetti and J. J. Risler, *Real Algebraic and Semi-Algebraic Sets*, Hermann, Paris, 1990.
2. A. Gabrielov and T. Zell, On the number of connected components of the relative closure of a semi-Pfaffian family, in *Algorithmic and Quantitative Real Algebraic Geometry*, AMS, Providence, RI, 2003, pp. 65–75.
3. A. G. Khovanski, On a class of systems of trascendental equations, *Dokl. Akad. Nauk SSSR* **255**(4), (1980), 804–807; English transl. in *Soviet Math. Dokl.* **22**(3) (1980), 762–765.
4. A. G. Khovanski, *Fewnomials*, Translations of Mathematical Monographs, Volume 88, AMS, Providence, RI, 1991.
5. S. Lefschetz, *Introduction to Topology*, Princeton University Press, Princeton, NJ, 1945.
6. T. Li, M. Rojas and X. Wang, Counting real connected components of trinomial curve intersections and m-nomial hypersurfaces, *Discrete Comput. Geom.* **30**(3) (2003), 379–414.
7. J. W. Milnor, *Topology from the Differentiable Viewpoint*, Princeton University Press, Princeton, NJ, 1965.
8. M. Voorhoeve, A generalization of Descartes' rule, *J. London Math. Soc.* (2) **20**(3) (1979), 446–456.

Received October 20, 2003, *and in revised form August* 30, 2004, *and January* 26, 2005.
Online publication August 3, 2005.