

Linear Matrix Inequality Representation of Sets

J. WILLIAM HELTON

University of California, San Diego

AND

VICTOR VINNIKOV

Ben Gurion University of the Negev

Abstract

This article concerns the question, Which subsets of \mathbb{R}^m can be represented with linear matrix inequalities (LMIs)? This gives some perspective on the scope and limitations of one of the most powerful techniques commonly used in control theory. Also, before having much hope of representing engineering problems as LMIs by automatic methods, one needs a good idea of which problems can and cannot be represented by LMIs. Little is currently known about such problems. In this article we give a necessary condition that we call “rigid convexity,” which must hold for a set $C \subseteq \mathbb{R}^m$ in order for C to have an LMI representation. Rigid convexity is proved to be necessary and sufficient when $m = 2$. This settles a question formally stated by Pablo Parrilo and Berndt Sturmfels in [15]. As shown by Lewis, Parillo, and Ramana [11], our main result also establishes (in the case of three variables) a 1958 conjecture by Peter Lax on hyperbolic polynomials. © 2006 Wiley Periodicals, Inc.

1 The Problem of Representing Sets with LMIs

Let C be a closed convex set in \mathbb{R}^m bounded by algebraic hypersurfaces

$$S_{p_j} := \{x \in \mathbb{R}^m : p_j(x) = 0\},$$

with all polynomials $p_j(x) > 0$ for x in the interior of C . Which C can be represented in terms of some linear pencil $L = \{L_0, L_1, \dots, L_m\}$ in the form

$$C^L := \{x = (x_1, \dots, x_m) : L_0 + L_1x_1 + \dots + L_mx_m \text{ is positive semidefinite}\}?$$

We shall require by our use of the term *linear pencil* that L_0, L_1, \dots, L_m be symmetric, real-entried matrices. Also, we frequently abbreviate positive definite to PD and positive semidefinite to PSD. A *monic pencil* is one with $L_0 = I$. We call such a representation of a set C a *linear matrix inequality* (LMI) representation of C .

The question is formally stated by Pablo Parrilo and Berndt Sturmfels in [15] for $m = 2$, and we resolve the $m = 2$ situation in this paper. Near the end of the paper we speculate about generalizations and other possible representations (Section 6).

In this paper a polynomial always stands for a polynomial with real coefficients. Since the paper might be of interest to several audiences—operator theory, real semialgebraic geometry, systems engineering, and possibly partial differential equations—it is written less tersely than is common. Our result here was announced in a survey talk at the conference Mathematical Theory of Networks and Systems in 2002; see [10].

1.1 Motivation

A technique that has had remarkable success in many types of optimization problems is to convert them to linear matrix inequalities. These can be solved using semidefinite programming algorithms provided the number of variables x_j is not extremely large (a few thousand). The technique was introduced first in [12] and has spread to various areas. In particular, control is an area within the authors' experience and there LMIs have had a great impact; cf. [9, 16]. For a collection of applications in many other areas, see the theses [14, 15].

In the many applications in which LMIs are found there is no systematic way to produce LMIs for general classes of problems. Each area has a few special tricks that convert “lucky problems” to LMIs. Before there is any hope of producing LMIs systematically, one must have a good idea about which types of constraint sets convert to LMIs and which do not. That is what this paper addresses.

After the preliminary version of this paper was circulated as a preprint, Adrian Lewis, Pablo Parillo, and M. V. Ramana informed us that our main result, Theorem 2.2, was critical to their solution of a 1958 conjecture by Peter Lax on hyperbolic polynomials arising from the study of linear hyperbolic partial differential equations; see [11]. We make some brief historical remarks in Section 8.

2 Solution

In this section we give an algebraic statement of the solution to the Parrilo-Sturmfels problem; see Theorem 2.2. In Section 3 we give an equivalent geometric statement of the solution; see Theorem 3.1. In Section 5 we give an equivalent topological statement (see Corollary 5.1 and Theorem 5.2), which while intuitively appealing requires projective spaces to state in full generality. The solution depends on results of Victor Vinnikov [17] (see also Dubrovin [8]) and variations on it.

2.1 A Key Class of Polynomials

We shall soon see that it is a special class of polynomials we call real zero polynomials that bound sets amenable to LMI representation. So now we define a *real zero polynomial* (RZ polynomial) to be a polynomial in m variables satisfying for each $x \in \mathbb{R}^m$,

$$(RZ) \quad p(\mu x) = 0 \quad \text{implies} \quad \mu \text{ is real.}$$

A few properties of RZ polynomials are as follows:

- (1) The product $p_1 \cdots p_k$ of RZ polynomials p_1, \dots, p_k is an RZ polynomial.
- (2) If an RZ polynomial p factors as $p = p_1 \cdots p_k$, then all factors p_1, \dots, p_k are RZ polynomials.
- (3) The determinant

$$(2.1) \quad \check{p}(x) := \det[L_0 + L_1 x_1 + \cdots + L_m x_m]$$

of a pencil L with a PD matrix L_0 is an RZ polynomial. Call this a *positive determinantal representation* for \check{p} .

- (4) The product of \check{p}_j defined as determinants of pencils L^j on \mathbb{R}^{N_j} , to wit

$$(2.2) \quad \check{p}_j(x) := \det[L_0^j + L_1^j x_1 + \cdots + L_m^j x_m],$$

is also the determinant of a pencil L , namely, of the direct sum $L_i := L_i^1 \oplus \cdots \oplus L_i^k$ on a single \mathbb{R}^N with $N = N_1 + \cdots + N_k$. We have

$$\begin{aligned} \det L &= (\det L^1)(\det L^2) \cdots (\det L^k) \\ &= \check{p}_1 \cdots \check{p}_k. \end{aligned}$$

- (5) See Section 3 and Section 5 for a characterization of RZ polynomials in terms of geometric and topological properties of their zero sets. The topological characterization and related topological properties of RZ polynomials (see Section 5.3) are essentially based on connectivity results of Nuij [13].

PROOF: The proofs of properties (1) and (2) about products and factors follows from the fact that the product $w = w_1 \cdots w_k$ of a set of numbers is 0 if and only if one of the factors is 0.

To show property (3) we note that for monic pencils

$$\check{p}\left(\frac{x}{\lambda}\right) = \frac{1}{\lambda^N} \det[\lambda I + L_1 x_1 + \cdots + L_m x_m],$$

and all of the eigenvalues of the symmetric matrix $L_1 x_1 + \cdots + L_m x_m$ are real, so we have that \check{p} is RZ. Replacing the monic condition by the requirement that L_0 be PD is obvious. A partial extension to the case when L_0 is merely PSD is provided by Lemma 2.3 below.

For property (4), the proof is in the statement. \square

We start the presentation of our main result with a key definition.

2.2 Algebraic Interiors and Their Degree

A closed set \mathcal{C} in \mathbb{R}^m is an *algebraic interior* if there is a polynomial p in m variables such that \mathcal{C} equals the closure of a connected component of

$$\{x \in \mathbb{R}^m : p(x) > 0\}.$$

In other words, there is a polynomial p in m variables that vanishes on the boundary of \mathcal{C} and such that $\{x \in \mathcal{C} : p(x) > 0\}$ is connected with closure equal to \mathcal{C} . (Notice that in general p may vanish also at some points in the interior of \mathcal{C} ; for

example, look at $p(x_1, x_2) = x_2^2 - x_1^2(x_1 - 1)$.) We denote the closure of the connected component of x^0 in $\{x \in \mathbb{R}^m : p(x) > 0\}$ by $\mathcal{C}_p(x^0)$. We often assume for normalization that 0 is in the interior of \mathcal{C} and set $\mathcal{C}_p = \mathcal{C}_p(0)$.

LEMMA 2.1 *A polynomial p of the lowest degree for which $\mathcal{C} = \mathcal{C}_p(x^0)$ is unique (up to multiplication by a positive constant), and any other polynomial q such that $\mathcal{C} = \mathcal{C}_q(x^0)$ is given by $q = ph$ where h is an arbitrary polynomial that is strictly positive on a dense connected subset of \mathcal{C} .*

PROOF: We shall be using some properties of algebraic and semialgebraic sets in \mathbb{R}^m , so many readers may want to skip over it and go to our main results, which are much more widely understandable; our reference is [4]. We notice first that \mathcal{C} is a semialgebraic set (since it is the closure of a connected component of a semialgebraic set; see proposition 2.2.2 and theorem 2.4.5 in [4]). Therefore the interior $\text{int } \mathcal{C}$ of \mathcal{C} is also semialgebraic, and so is the boundary $\partial \mathcal{C} = \mathcal{C} \setminus \text{int } \mathcal{C}$. Notice also that \mathcal{C} equals the closure of its interior.

We claim next that *for each $x \in \partial \mathcal{C}$, the local dimension $\dim \partial \mathcal{C}_x$ equals $m - 1$.* On the one hand, we have

$$\dim \partial \mathcal{C}_x \leq \dim \partial \mathcal{C} < \dim \text{int } \mathcal{C} = m;$$

here we have used propositions 2.8.13 and 2.8.4 in [4] and the fact that $\partial \mathcal{C} = \text{clos int } \mathcal{C} \setminus \text{int } \mathcal{C}$, since \mathcal{C} equals the closure of its interior. On the other hand, let B be an open ball in \mathbb{R}^m around x ; then

$$B \cap \partial \mathcal{C} = B \setminus [(B \cap (\mathbb{R}^m \setminus \mathcal{C})) \cup (B \cap \text{int } \mathcal{C})].$$

Since \mathcal{C} equals the closure of its interior, every point of $\partial \mathcal{C}$ is an accumulation point of both $\mathbb{R}^m \setminus \mathcal{C}$ and $\text{int } \mathcal{C}$; therefore $B \cap (\mathbb{R}^m \setminus \mathcal{C})$ and $B \cap \text{int } \mathcal{C}$ are disjoint, open, nonempty semialgebraic subsets of B . Using [4, lemma 4.5.2], we conclude that $\dim B \cap \partial \mathcal{C} \geq m - 1$; hence $\dim \partial \mathcal{C}_x \geq m - 1$.

Let now V be the Zariski closure of $\partial \mathcal{C}$, and let $V = V_1 \cup \dots \cup V_k$ be the decomposition of V into irreducible components. We claim that $\dim V_i = m - 1$ for each i . Assume by contradiction that V_1, \dots, V_l have dimension $m - 1$ while V_{l+1}, \dots, V_k have smaller dimension. Then there exists $x \in \partial \mathcal{C}$ such that $x \notin V_1, \dots, V_l$, and consequently there exists an open ball B in \mathbb{R}^m around x such that

$$B \cap \partial \mathcal{C} = (B \cap \partial \mathcal{C} \cap V_{l+1}) \cup \dots \cup (B \cap \partial \mathcal{C} \cap V_k).$$

By assumption each set in the union on the right-hand side has dimension smaller than $m - 1$; hence it follows (by [4, proposition 2.8.5, I]) that $\dim B \cap \partial \mathcal{C} < m - 1$, a contradiction with $\dim \partial \mathcal{C}_x = m - 1$.

Suppose now that p is a polynomial of the lowest degree with $\mathcal{C} = \mathcal{C}_p(x^0)$. Lowest degree implies that p can have no multiple irreducible factors, i.e., $p = p_1 \cdots p_s$, where p_1, \dots, p_s are distinct, irreducible polynomials; we may assume without loss of generality that every p_i is nonnegative on \mathcal{C} . Since p vanishes on $\partial \mathcal{C}$, it also vanishes on $V = V_1 \cup \dots \cup V_k$. We claim that for every V_i there exists a

p_j so that p_j vanishes on V_i ; otherwise, $\mathcal{Z}(p_j) \cap V_i$ is a proper algebraic subset of V_i for every $j = 1, \dots, s$; therefore (since V_i is irreducible) $\dim \mathcal{Z}(p_j) \cap V_i < \dim V_i$ for every j and thus also $\dim \mathcal{Z}(p) \cap V_i < \dim V_i$, a contradiction since p vanishes on V_i . If p_j vanishes on V_i , it follows (since p_j is irreducible and $\dim V_i = m - 1$) that $\mathcal{Z}(p_j) = V_i$. The fact that p is a polynomial of the lowest degree with $\mathcal{C} = \mathcal{C}_p(x^0)$ implies now (after possibly renumbering the irreducible factors of p) that $p = p_1 \cdots p_k$ where $\mathcal{Z}(p_i) = V_i$ for every i . Since $\dim V_i = m - 1$ it follows from the real Nullstellensatz for principal ideals [4, theorem 4.5.1] that the irreducible polynomials p_i are uniquely determined (up to multiplication by a positive constant); hence so is their product p . This proves the uniqueness of p .

The rest of the lemma now follows easily. If $\mathcal{C} = \mathcal{C}_q(x^0)$, then the polynomial q vanishes on $\partial\mathcal{C}$, hence also on $V = V_1 \cup \cdots \cup V_k$. Since q vanishes on $\mathcal{Z}(p_i) = V_i$ and $\dim V_i = m - 1$, the real Nullstellensatz for principal ideals implies that q is divisible by p_i . This holds for every i ; hence q is divisible by $p = p_1 \cdots p_k$, i.e., $q = ph$. It is obvious that h must be strictly positive on a dense connected subset of \mathcal{C} . \square

We refer to p as a *minimal defining polynomial for \mathcal{C}* and say that \mathcal{C} is an *algebraic interior of degree d* , where $d = \deg p$.

We note in passing (compare to the proof of Lemma 2.1 above) that a connected semialgebraic set that equals the closure of its interior is not necessarily an algebraic interior; for example, look at the set of all points in \mathbb{R}^2 to the left of the unbounded branch of the curve $x_2^2 - x_1(x_1 + 1)(x_1 + 2) = 0$.

2.3 Main Theorem on LMI Representations: Algebraic Version

THEOREM 2.2 *If a set \mathcal{C} in \mathbb{R}^m , with 0 in the interior of \mathcal{C} , has an LMI representation, then it is a convex algebraic interior, and the minimal defining polynomial p for \mathcal{C} satisfies the RZ condition (with $p(0) > 0$).*

Conversely, when $m = 2$, if p is an RZ polynomial of degree d and $p(0) > 0$, then \mathcal{C}_p has a monic LMI representation with $d \times d$ matrices.

The idea behind our organization of the proof is to do the informative elementary arguments first, then a geometric interpretation in Section 3 (readers less interested in proofs might go there next), and then some examples. Finally, in a separate section we do the converse side, which refers heavily to [17] and shows that *an RZ polynomial p in two variables has a positive determinantal representation.*

Proof That RZ Is Necessary

We are given a pencil L that represents \mathcal{C} , $\mathcal{C} = \mathcal{C}^L$; since $0 \in \mathcal{C}$ we have L_0 is PSD. Lemma 2.3 below reduces general pencils with L_0 a PSD matrix to monic pencils, so we assume that L is a monic pencil. Define a polynomial \check{p} by

$$(2.3) \quad \check{p}(x) := \det[I + L_1 x_1 + \cdots + L_m x_m]$$

and note that property (3) of RZ polynomials says that \check{p} satisfies RZ. Now we show that $\mathcal{C}_{\check{p}}$ equals \mathcal{C}^L (equals \mathcal{C}). If $x(t)$, $t \in [0, 1]$, is any continuous path in \mathbb{R}^m with $x(0) = 0$ and $\check{p}(x(t)) > 0$ for all t , then $L(x(t))$ is a continuous path in real symmetric matrices with $L(x(0)) = I$ and $\det L(x(t)) > 0$ for all t ; hence $L(x(1))$ is necessarily PD. It follows that the connected component of 0 in $\{x \in \mathbb{R}^m : \check{p}(x) > 0\}$ is contained in \mathcal{C}^L and therefore $\mathcal{C}_{\check{p}}$ is contained in \mathcal{C}^L . Conversely, $x \in \mathcal{C}^L$ implies $I + \mu[L_1x_1 + \cdots + L_mx_n]$ is PD for $0 \leq \mu < 1$, so $\check{p}(\mu x) > 0$ for $0 \leq \mu < 1$ and therefore necessarily $x \in \mathcal{C}_{\check{p}}$.

We conclude that $\mathcal{C} = \mathcal{C}_{\check{p}}$ is a convex algebraic interior. If p is the minimal defining polynomial for \mathcal{C} , then p is a factor of \check{p} . By property (2) for RZ polynomials the polynomial p is itself RZ.

Notice that the same argument as above shows that for a monic pencil L the interior of $\mathcal{C} = \mathcal{C}^L$ is exactly $\{x : I + x_1L_1 + \cdots + x_mL_m \text{ is PD}\}$, and therefore the polynomial \check{p} —and thus also the minimal defining polynomial p —is strictly positive on the interior of \mathcal{C} .

Reduction to Monic Pencils

We reduce the pencil representation problem to one involving monic pencils as promised.

LEMMA 2.3 *If \mathcal{C} contains 0 in its interior and if there is a pencil L such that*

$$\mathcal{C} = \mathcal{C}^L,$$

then there is a monic pencil \widehat{L} such that

$$\mathcal{C} = \mathcal{C}^{\widehat{L}}.$$

PROOF: Given \mathcal{C} and L , since $0 \in \mathcal{C}$ we have $0 \in \mathcal{C}^L$, or equivalently, L_0 is PSD. We need, however, to take L_0 invertible. Since 0 is in the interior of $\mathcal{C} \subset \mathbb{R}^m$, it is also in the interior of \mathcal{C}^L , so we have

$$L_0 \geq \pm \varepsilon_1 L_1, \dots, L_0 \geq \pm \varepsilon_m L_m$$

for small enough ε_1 through ε_m . Set $Y = \text{range } L_0 \subset \mathbb{R}^N$, and set

$$\widetilde{L}_j := L_j|_Y, \quad j = 0, 1, \dots, m.$$

Clearly $\widetilde{L}_0 : Y \rightarrow Y$ is invertible and indeed PD. We next show that $\text{range } L_0$ contains $\text{range } L_1$. If $x \perp \text{range } L_0$, i.e., $L_0x = 0$, then $0 = x^\top L_0x \geq \pm \varepsilon_1 x^\top L_1x$ and hence $x^\top L_1x = 0$. Since $L_0 + \varepsilon_1 L_1 \geq 0$ and $x^\top (L_0 + \varepsilon_1 L_1)x = 0$, it follows that $(L_0 + \varepsilon_1 L_1)x = 0$, and since $L_0x = 0$, we finally conclude that $L_1x = 0$, i.e., $x \perp \text{range } L_1$.

Likewise, $\text{range } L_0$ contains the ranges of L_2, \dots, L_m . Consequently, $\widetilde{L}_j : Y \rightarrow Y$ are all symmetric and $\mathcal{C}^L = \mathcal{C}^{\widetilde{L}}$.

To build \widehat{L} , factor $\widetilde{L}_0 = B^\top B$ with B invertible and set

$$\widehat{L}_j := B^{-\top} \widetilde{L}_j B^{-1}, \quad j = 0, \dots, m.$$

The resulting pencil \widehat{L} is monic and $\mathcal{C}^{\widehat{L}} = \mathcal{C}^{\widetilde{L}} = \mathcal{C}^L$. □

2.4 Shifted Real Zero Condition

The assumption that $0i$ is in \mathcal{C} is just a normalization, which is justified since we can shift \mathbb{R}^m by x^0 and obtain a shifted polynomial

$$\tilde{p}(x) := p(x + x^0)$$

with $\mathcal{C}_p(x^0) - x^0 = \mathcal{C}_{\tilde{p}}(0)$. Then the RZ condition on \tilde{p} shifts back to the RZ_{x^0} condition on p at x^0 . Namely, for each $x \in \mathbb{R}^m$

$$(\text{RZ}_{x^0}) \quad p(\mu x + x^0) = 0 \quad \text{implies} \quad \mu \text{ is real.}$$

An obvious consequence of our main theorem is the following:

COROLLARY 2.4 *If a set \mathcal{C} in \mathbb{R}^m , with x^0 in the interior of \mathcal{C} , has an LMI representation, then it is a convex algebraic interior and the minimal defining polynomial p for \mathcal{C} satisfies the RZ_{x^0} condition.*

Conversely, when $m = 2$, if p is polynomial of degree d satisfying RZ_{x^0} and $p(x^0) > 0$, then $\mathcal{C}_p(x^0)$ has an LMI representation with $d \times d$ matrices.

Remark 2.5. For $m = 2$, “going through” LMI representations, i.e., invoking the second assertion of Corollary 2.4 followed by the first assertion, we conclude that

If the polynomial p satisfies $\text{RZ}_{\tilde{x}^0}$ for some point \tilde{x}^0 and $p(\tilde{x}^0) > 0$,
then p satisfies RZ_{x^0} for every interior point x^0 of $\mathcal{C}_p(\tilde{x}^0)$.

This statement holds in any dimension m and can be proved, without any reference to LMI representations, using the alternative topological characterization of RZ polynomials; see Section 5.3.

PROOF OF COROLLARY 2.4: Assume that a set \mathcal{C} in \mathbb{R}^m , with x^0 in the interior of \mathcal{C} , has a pencil representation L , $\mathcal{C} = \mathcal{C}^L$. Shift L by x^0 to obtain a pencil \widetilde{L} , that is,

$$\widetilde{L}_0 + \widetilde{L}_1 x_1 + \cdots + \widetilde{L}_m x_m = L_0 + L_1(x_1 + x_1^0) + \cdots + L_m(x_m + x_m^0).$$

Then $\mathcal{C}^L - x^0 = \mathcal{C}^{\widetilde{L}}$, i.e., \widetilde{L} is a pencil representation of the set $\widetilde{\mathcal{C}} = \mathcal{C} - x^0$, with 0 in the interior of $\widetilde{\mathcal{C}}$. By Lemma 2.3 we may assume that \widetilde{L} is a monic pencil. The determinant $\check{\tilde{p}}$ of \widetilde{L} satisfies the RZ_0 property. Thus the minimal defining polynomial \tilde{p} of $\widetilde{\mathcal{C}}$, which is the shift of the minimal defining polynomial p of \mathcal{C} and is contained in the RZ_0 polynomial $\check{\tilde{p}}$ as a factor, satisfies RZ_0 . Shift 0 back to x^0 to get that p satisfies RZ_{x^0} .

The proof of the second assertion, for $m = 2$, proceeds analogously. □

Notice that rather than saying that the shifted pencil \widetilde{L} is monic, we can say a little more generally that \widetilde{L}_0 is PD. This is equivalent to saying that $L_0 + x_1^0 L_1 + \cdots + x_m^0 L_m$ is PD, and then, by the remark at the end of Section 2.3, the interior of $\mathcal{C} = \mathcal{C}^L$ is exactly $\{x : L_0 + x_1 L_1 + \cdots + x_m L_m \text{ is PD}\}$.

3 Geometrical Viewpoint and Examples of Sets with No LMI Representation

In this section we give a geometric characterization of sets \mathcal{C} with an LMI representation, which is equivalent to Theorem 2.2.

3.1 Rigid Convexity

An algebraic interior \mathcal{C} of degree d in \mathbb{R}^m with minimal defining polynomial p will be called *rigidly convex* provided for every point x^0 in the interior of \mathcal{C} and a generic line ℓ through x^0 , ℓ intersects the (affine) real algebraic hypersurface $p(x) = 0$ in exactly d points.¹ We soon show that rigid convexity of \mathcal{C}_p is the same as p having the (shifted) RZ property, and as a consequence we obtain the following *geometric version of our main theorem*.

THEOREM 3.1 *If \mathcal{C} is a closed, convex set in \mathbb{R}^m with an LMI representation, then \mathcal{C} is rigidly convex. When $m = 2$, the converse is true; namely, a rigidly convex degree d set has an LMI representation with symmetric matrices $L_j \in \mathbb{R}^{d \times d}$.*

PROOF: The theorem follows from Theorem 2.2 (more precisely, its shifted version Corollary 2.4) provided we can show that *an algebraic interior \mathcal{C} with minimal defining polynomial p is rigidly convex if and only if p satisfies the shifted RZ condition*. We now set about to prove this equivalence.

Let $\ell := \{x \in \mathbb{R}^m : x = \mu v + x^0, \mu \in \mathbb{R}\}$, $v \in \mathbb{R}^m$, be a parametrization of a line through x^0 in the interior of \mathcal{C} . The points of intersection of the real algebraic hypersurface $p(x) = 0$ with the line ℓ are parametrized by exactly those $\mu \in \mathbb{R}$ at which

$$f(\mu) := p(\mu v + x^0) = 0,$$

and rigid convexity of \mathcal{C}_p says that (for a generic ℓ) there are exactly d such distinct μ . However, the degree of p is d , so for a generic direction $v \in \mathbb{R}^m$, the degree of the polynomial f equals d (the exceptional directions are given by $p_d(v) = 0$, where p_d is the sum of degree d terms in p). Thus these f , by the fundamental theorem of algebra, have exactly d zeroes $\mu \in \mathbb{C}$, counting multiplicities. Furthermore, for a generic direction v , all the zeroes are simple: multiple zeroes correspond to common zeroes of f and $f' = \sum_{i=1}^m v_i \frac{\partial p}{\partial x_i}$, that is, to lines ℓ which pass through a point of intersection of algebraic hypersurfaces $p(x) = 0$ and $\sum_{i=1}^m v_i \frac{\partial p}{\partial x_i}(x) = 0$. However, since p is irreducible and $\deg \sum_{i=1}^m v_i \frac{\partial p}{\partial x_i} < \deg p$, this intersection is a proper algebraic subvariety of $p(x) = 0$ and hence of dimension at most $m - 2$. (This argument applies to both real and complex zeroes by

¹*Generic* means that the exceptional lines are contained in a proper algebraic subvariety (in $\mathbb{P}^{m-1}(\mathbb{R})$); when $m = 2$ this simply means all but a finite number of lines. One can replace here “generic line” by “every line” if one takes multiplicities into account when counting the number of intersections, and also counts the intersections at infinity, i.e., replaces the affine real algebraic hypersurface $p(x_1, \dots, x_m) = 0$ by the projective real algebraic hypersurface $X_0^d p(X_1/X_0, \dots, X_m/X_0) = 0$.

considering real algebraic hypersurfaces and their complexifications, respectively.) Thus rigid convexity says precisely that all zeroes of f are at real μ for f arising from a generic v .

To obtain that this is equivalent to the RZ_{x^0} property for p , we give an elementary continuity argument that implies that all f have only real zeroes. Pick an exceptional direction v (which might make f have degree less than d): if f_n is a sequence of polynomials of degree d in one variable with real coefficients such that every f_n has only real zeroes, and if the polynomials f_n converge coefficientwise to a polynomial f (of degree less than or equal to d), then f has only real zeroes. \square

Notice that it follows from Theorem 3.1 that for $m = 2$ a rigidly convex set is in fact convex, and (as in the remark following Corollary 2.4), it is enough to require the defining property of rigid convexity to hold for lines through a single point of \mathcal{C} . Again, this holds for any dimension m and can be proved directly using the alternative topological characterization of RZ polynomials—see Section 5.3. We also indicate there a strong stability property of rigidly convex algebraic interiors and speculate about this stability being characteristic of rigid convexity.

3.2 Examples

Example 1

The polynomial

$$p(x_1, x_2) = x_1^3 - 3x_2^2x_1 - (x_1^2 + x_2^2)^2$$

has the zero set shown in Figure 3.1.

The complement to $p = 0$ in \mathbb{R}^2 consists of four components, three bounded convex components where $p > 0$ and an unbounded component where $p < 0$. Let us analyze one of the bounded components, say the one in the right half-plane; \mathcal{C} is the closure of

$$\{x : p(x) > 0, x_1 > 0\}.$$

Does \mathcal{C} have an LMI representation? To check this, fix a point O inside \mathcal{C} , e.g., $O = (0.7, 0)$.

By Theorem 3.1 almost every line l thru $(0, 0.7)$ (as in Figure 3.1) must intersect $p = 0$ in four real points or the RZ_O condition is violated. We can see from the picture in \mathbb{R}^2 that there is a continuum of real lines ℓ through $(0, 0.7)$ intersecting $p = 0$ in exactly two real points. Thus by Theorem 3.1 the set \mathcal{C} does not have an LMI representation. (Since p is irreducible, it is the minimal defining polynomial for \mathcal{C} .)

Example 2

Consider the polynomial

$$p(x_1, x_2) = 1 - x_1^4 - x_2^4$$

Clearly, $\mathcal{C}_p := \{x : p(x) \geq 0\}$ has degree 4 but all lines in \mathbb{R}^2 through it intersect the set $p = 0$ in exactly two places. Thus \mathcal{C}_p is not rigidly convex.

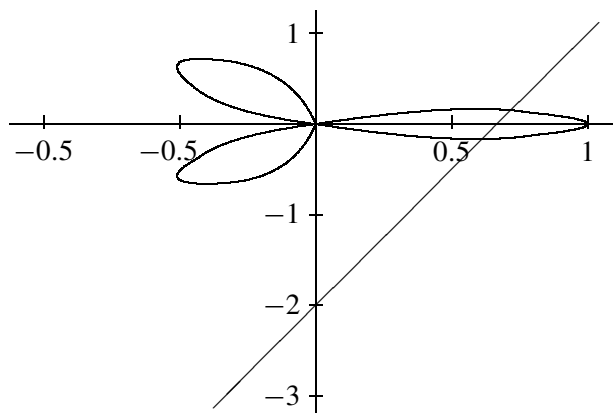


FIGURE 3.1. $p = x_1^3 - 3x_2^2x_1 - (x_1^2 + x_2^2)^2$ with a line through $O = (0.7, 0)$ hitting \mathcal{Z}_p in only two points.

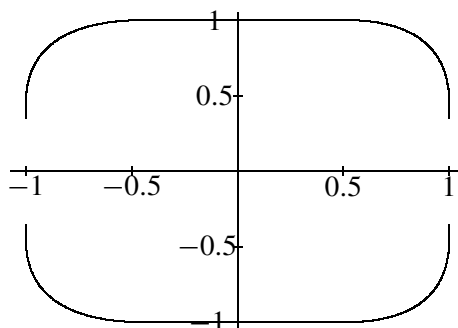


FIGURE 3.2. $p(x_1, x_2) = 1 - x_1^4 - x_2^4$.

4 Proof of the Converse Side of Theorem 2.2

The proof is an adaptation of the results of [17].

Theorem 6.1 of [17] (see also [8]) says that any irreducible polynomial p on \mathbb{R}^2 such that the projective closure of $p(x_1, x_2) = 0$ is a smooth projective plane curve satisfying (RZ) (with $p(0) > 0$) admits a representation (2.1) with $L_0, L_1,$ and L_2 complex hermitian matrices on $\mathbb{C}^d, d = \deg p,$ where L_0 is PD. To obtain the positive determinantal representation underlying Theorem 2.2, we need this to hold generally (not just for smooth irreducible curves), and we must also obtain $L_0, L_1,$ and $L_2,$ which are real symmetric.

The possibility to take $L_0, L_1,$ and L_2 real symmetric follows from the arguments given in the last four paragraphs of the introduction of [17, p. 456]. The irreducibility assumption on p is irrelevant here, since we may consider one irreducible factor of p at a time (and then form a direct sum of the corresponding

determinantal representations (2.1)). The smoothness assumption can be removed by considering *maximal* determinantal representations as in [2, 3].

This proof might lend itself to construction of the pencil L via a construction of a basis for a linear system (or a Riemann-Roch space) on a plane algebraic curve.

Since the proof above is actually an outline that leaves much for the reader to fill in, we present an *alternative direct proof* based on the results of [3, sec. 5]. As before, we may assume without loss of generality that p is an irreducible polynomial; we consider the complex affine algebraic curve $\{(x_1, x_2) \in \mathbb{C}^2 : p(x_1, x_2) = 0\}$, take its projective closure in the complex projective plane $\mathbb{P}^2(\mathbb{C})$, and let M be the desingularizing Riemann surface. M is a compact real Riemann surface, i.e., a compact Riemann surface equipped with an antiholomorphic involution induced by $(x_1, x_2) \mapsto (\bar{x}_1, \bar{x}_2)$. The set $M_{\mathbb{R}}$ of real points of M is by definition the set of fixed points of the antiholomorphic involution, and x_1 and x_2 are real meromorphic functions on M , i.e., meromorphic functions taking real values at real points. Furthermore, it follows from the RZ condition that the real meromorphic function $\frac{x_2}{x_1}$ on M takes real values at real points only. It follows that M is a compact real Riemann surface of dividing type; i.e., $M \setminus M_{\mathbb{R}}$ consists of two connected components. We denote by M_+ and M_- the components where $\Im \frac{x_2}{x_1} > 0$ and $\Im \frac{x_2}{x_1} < 0$, respectively. We orient $M_{\mathbb{R}}$ so that $M_{\mathbb{R}} = \partial M_+$; then the real meromorphic differential

$$-d \frac{x_1}{x_2} = \frac{x_1 dx_2 - x_2 dx_1}{x_2^2}$$

is everywhere nonnegative on $M_{\mathbb{R}}$.

Since both the assumptions and the conclusions of the theorem are invariant under a linear change of coordinates x_1 and x_2 , we may assume without loss of generality that $p(0, \mu) = 0$ has d distinct real roots c_1, \dots, c_d ; they correspond to d nonsingular points $(0, c_1), \dots, (0, c_d)$ on the affine algebraic curve $p(x_1, x_2) = 0$, and we denote by q_1, \dots, q_d the corresponding points on M . We may now write down explicit formulae for the $d \times d$ matrices L_0, L_1 , and L_2 :

$$(4.1) \quad L_0 = I_d,$$

$$(4.2) \quad [L_1]_{ij} = \begin{cases} i = j : \frac{dx_2}{dx_1}(q_i) \\ i \neq j : \left(\frac{1}{c_i} - \frac{1}{c_j}\right) \frac{\theta_{[b]}^{[a]}(\phi(q_j) - \phi(q_i))}{\theta_{[b]}^{[a]}(0)E(q_j, q_i)} \frac{1}{\sqrt{d \frac{x_1}{x_2}(q_i)} \sqrt{d \frac{x_1}{x_2}(q_j)}} \end{cases}$$

$$(4.3) \quad L_2 = \text{diag}\left(-\frac{1}{c_1}, \dots, -\frac{1}{c_d}\right).$$

Here $\theta_{[b]}^{[a]}(\cdot)$ is the theta function with characteristics associated to the Jacobian variety of M , ϕ is the Abel-Jacobi map from M into the Jacobian variety, and $E(\cdot, \cdot)$ is the prime form on M ; for details see [17, sec. 2] or [3, sec. 4]. Furthermore, we choose a symmetric homology basis on M as in [17, sec. 2] with a

resulting period matrix Ω , and $b + \Omega a \in \mathbb{C}^g$ is a half period in the Jacobian with $a_1, \dots, a_{g+1-k} = 1$ and $b_{g+2-k}, \dots, b_g = 0$ (g is the genus of M and k is the number of connected components of $M_{\mathbb{R}}$).

The fact that $\det(L_0 + x_1 L_1 + x_2 L_2) = p(x_1, x_2)$ (up to a constant positive factor) is exactly [3, theorem 5.1] (with a pair of meromorphic functions $\lambda_1 = 1/x_1$ and $\lambda_2 = x_2/x_1$ on M and $-x_1/x_2$ as a local parameter at the poles of λ_1 and λ_2 [zeroes of x_1]). The fact that L_1 is a real symmetric matrix follows immediately from the standard properties of theta functions and [17, prop. 2.3]. Notice that $b + \Omega a \in T_0$ hence $\theta \begin{bmatrix} a \\ b \end{bmatrix} (0) \neq 0$; see [17, theorem 3.1 and cor. 4.3].

5 Topological Characterization of RZ Polynomials

This section gives an equivalent topological characterization of zero sets of RZ polynomials. First we show that RZ polynomials of degree d for which the sets of real zeros are compact are exactly those whose zero sets consist of a nest of $d/2$ “ovaloids” (having compact zero set implies d even). A result of similar definitiveness holds when real zero sets are not compact, but it requires projective space to neatly account for the effects of the zero set passing through infinity. Thus we describe the general case only after describing p with compact real zero set despite generating a little redundancy. Then we give some related properties of RZ polynomials.

5.1 RZ Polynomials Whose Sets of Real Zeroes Are Compact

We first define carefully what is meant by an “ovaloid” in \mathbb{R}^m , and later give an analogous definition in projective space; see, for example, [18]. We call $W \subset \mathbb{R}^m$ an *ovaloid* in \mathbb{R}^m if W is isotopic in \mathbb{R}^m to a sphere $S \subset \mathbb{R}^m$; i.e., there is a homeomorphism F of \mathbb{R}^m with $F(S) = W$, and furthermore F is homotopic to the identity (i.e., there is a homeomorphism H of $[0, 1] \times \mathbb{R}^m$ such that $H_t = H|_{\{t\} \times \mathbb{R}^m}$ is a homeomorphism of \mathbb{R}^m for every t , $H_0 = \text{Id}_{\mathbb{R}^m}$, and $H_1 = F$). Notice that $\mathbb{R}^m \setminus S$ consists of two connected components only one of which is contractible; hence the same is true of $\mathbb{R}^m \setminus W$; we call the contractible component the *interior* of the ovaloid W , and the noncontractible component the *exterior*.

The following is a consequence of Theorem 5.2 below:

COROLLARY 5.1 *Let p be a polynomial whose set V of real zeros is compact (then the degree d of p must be even, $d = 2k$) and which defines a smooth (affine) real algebraic hypersurface; i.e., p has no critical points on V . Then p satisfies RZ_{x^0} with $p(x^0) \neq 0$ if and only if V is a disjoint union of k ovaloids W_1, \dots, W_k in \mathbb{R}^m , with W_i contained in the interior of W_{i+1} , $i = 1, \dots, k - 1$, and x^0 lying in the interior of W_1 . Furthermore, in this case every line L through x^0 intersects V in d distinct points.*

The corollary misses being as strong as possible and in fact being sharp because of the requirement that ovaloids be disjoint. Intuitively, if the real zero set of an

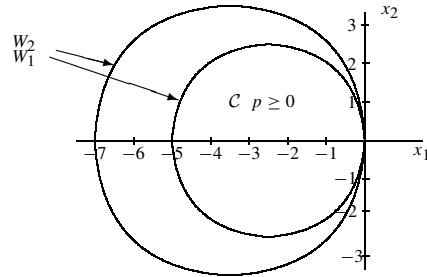


FIGURE 5.1. $p(x_1, x_2) = (x_1^2 + x_2^2)(x_1^2 + x_2^2 + 12x_1 - 1) + 36x_1^2 \geq 0$.

RZ polynomial is a compact singular (nonsmooth) hypersurface, then it consists of nested ovaloids that touch at the singular points as in Figure 5.1. Notice that it follows from property (1) in Section 5.3 below that any RZ polynomial of degree $d = 2k$ with a compact real zero set is the limit of RZ polynomials of degree d with compact real zero sets and which define smooth real algebraic hypersurfaces (that is, the zero sets are disjoint unions of k nested ovaloids).

5.2 RZ Polynomials with a Noncompact Set of Real Zeroes

We shall need to consider the behavior of the polynomial at infinity, so we introduce the homogenization P of p ,

$$P(X_0, X_1, \dots, X_m) = X_0^d p(X_1/X_0, \dots, X_m/X_0),$$

where $d = \deg p$. Notice that P is not divisible by X_0 , and that conversely given any homogeneous polynomial P of degree d in $m + 1$ variables that is not divisible by X_0 , we may recover a polynomial p of degree d in m variables so that P is the homogenization of p by

$$p(x_1, \dots, x_m) = P(1, x_1, \dots, x_m).$$

We identify as usual the m -dimensional real projective space $\mathbb{P}^m(\mathbb{R})$ with the union of \mathbb{R}^m and the hyperplane at infinity $X_0 = 0$, so that the affine coordinates x and the projective coordinates X are related by $x_1 = X_1/X_0, \dots, x_m = X_m/X_0$; the *projective real algebraic hypersurface* $P(X) = 0$ in $\mathbb{P}^m(\mathbb{R})$ is called the *projective closure* of the affine real algebraic hypersurface $p(x) = 0$ in \mathbb{R}^m . We shall say that p defines a *smooth projective real algebraic hypersurface* if P has no critical points on $P(X) = 0$. We shall often abuse the notation and denote a point in $\mathbb{P}^m(\mathbb{R})$ and a $(m + 1)$ -tuple of its projective coordinates by the same letter.

We call $W \subset \mathbb{P}^m(\mathbb{R})$ an *ovaloid* if W is isotopic in $\mathbb{P}^m(\mathbb{R})$ to a sphere $S \subset \mathbb{R}^m \subset \mathbb{P}^m(\mathbb{R})$; i.e., there is a homeomorphism F of $\mathbb{P}^m(\mathbb{R})$ with $F(S) = W$, and furthermore F is homotopic to the identity (i.e., there is a homeomorphism H of $[0, 1] \times \mathbb{P}^m(\mathbb{R})$ such that $H_t = H|_{\{t\} \times \mathbb{P}^m(\mathbb{R})}$ is a homeomorphism of $\mathbb{P}^m(\mathbb{R})$ for every t , $H_0 = \text{Id}_{\mathbb{P}^m(\mathbb{R})}$, and $H_1 = F$). Notice that $\mathbb{P}^m(\mathbb{R}) \setminus S$ consists of two connected components, only one of which is contractible; hence the same is true

of $\mathbb{P}^m(\mathbb{R}) \setminus W$; we call the contractible component the *interior* of the ovaloid W , and the noncontractible component the *exterior*.

We call $W \subset \mathbb{P}^m(\mathbb{R})$ a *pseudohyperplane* if W is isotopic in $\mathbb{P}^m(\mathbb{R})$ to a (projective) hyperplane $H \subset \mathbb{P}^m(\mathbb{R})$.

THEOREM 5.2 *Let p be a polynomial of degree d in m variables that defines a smooth projective real algebraic hypersurface V in $\mathbb{P}^m(\mathbb{R})$. Then p satisfies RZ_{x^0} with $p(x^0) \neq 0$ if and only if*

- (i) *if $d = 2k$ is even, V is a disjoint union of k ovaloids W_1, \dots, W_k , with W_i contained in the interior of W_{i+1} , $i = 1, \dots, k - 1$, and x^0 lying in the interior of W_1 ;*
- (ii) *if $d = 2k + 1$ is odd, V is a disjoint union of k ovaloids W_1, \dots, W_k , with W_i contained in the interior of W_{i+1} , $i = 1, \dots, k - 1$, and x^0 lying in the interior of W_1 , and a pseudohyperplane W_{k+1} contained in the exterior of W_k .*

Furthermore, in this case every (projective) line L through x^0 intersects V in d distinct points.

The proof appears in Section 7.

5.3 Topological Properties of RZ Polynomials

We list now some properties of RZ polynomials, mostly of a topological nature, which are closely related to Theorem 5.2.

(1) If two polynomials p_0 and p_1 of degree d in m variables satisfy RZ and $p_0(0), p_1(0) > 0$, then there is a continuous path through RZ polynomials of degree d in m variables from p_0 to p_1 with each polynomial p in the path satisfying $p(0) > 0$. Furthermore, we may assume that except for possibly the endpoints p_0 and p_1 , the polynomial p defines a smooth projective real algebraic hypersurface.

(2) If the polynomial p satisfies $RZ_{\tilde{x}^0}$ for some point \tilde{x}^0 and $p(\tilde{x}^0) > 0$, then p satisfies RZ_{x^0} for every interior point x^0 of $\mathcal{C}_p(\tilde{x}^0)$.

(3) If p is a RZ polynomial with $p(0) > 0$, then the algebraic interior \mathcal{C}_p is convex (so a rigidly convex algebraic interior is in fact convex).

(4) Suppose that there is a continuous path through polynomials of degree d in m variables from a polynomial p_0 to a polynomial p_1 such that each polynomial p in the path (including p_0 and p_1) defines a smooth projective real algebraic hypersurface and $p(0) > 0$. Then if p_0 is a RZ polynomial, so is p_1 .

(5) Suppose that there is a continuous path through polynomials of degree d in m variables from a polynomial p_0 to a polynomial p_1 such that each polynomial p in the path (including p_0 and p_1) defines a smooth projective real algebraic hypersurface. Assume that p_0 satisfies RZ_{x^0} for some point x^0 and $p(x^0) > 0$; then p_1 satisfies RZ_{x^1} for some point x^1 and $p(x^1) > 0$.

The proof appears in Section 7.

6 Generalizations

We now discuss several conjectures that extend the results proved in this paper.

6.1 $m > 2$

It is a natural question to which extent the converse side of our main results, Theorem 2.2 and Theorem 3.1, extend to dimension $m > 2$. A simple count of parameters shows that given a polynomial $p(x_1, \dots, x_m)$ in $m > 2$ variables, in general there do not exist $d \times d$ matrices L_0, \dots, L_m such that $p(x) = \det(L_0 + x_1 L_1 + \dots + x_m L_m)$ (here the polynomial has complex coefficients and the matrices are likewise complex). Thus we cannot expect a rigidly convex degree d set in \mathbb{R}^m , $m > 2$, to admit an LMI representation with $d \times d$ symmetric matrices. However, the count of parameters does not preclude the existence of a pencil representation with matrices of a size larger than the degree. In fact, we conjecture the following:

A convex set \mathcal{C} in \mathbb{R}^m has an LMI representation if and only if it is rigidly convex.

This follows if

every RZ polynomial has a positive determinantal representation,

which we conjecture to be true.

6.2 LMI Lifts

Another practically important question (in the spirit of Nesterov and Nimmerovski [12]) is the following:

Find a test which insures that a convex set \mathcal{C} in \mathbb{R}^m lifts to some LMI representable set $\tilde{\mathcal{C}}$ in a possibly bigger space \mathbb{R}^{m+k} . That is, find necessary and sufficient properties on a given set \mathcal{C} in \mathbb{R}^m which insure that there exists an LMI representable set $\tilde{\mathcal{C}}$ in \mathbb{R}^{m+k} whose projection onto \mathbb{R}^m equals the set \mathcal{C} .

No a priori restriction on such a \mathcal{C} is apparent, either from theoretical considerations or from numerical experiments (done informally by Pablo Parrilo). Of course, \mathcal{C} has to be a semialgebraic set with a connected interior, which is equal to the closure of its interior. It is not clear whether \mathcal{C} should be an algebraic interior as we defined it.

6.3 Convexity under Deformation Is Ephemeral

Properties (4) and (5) in Section 5.3 imply that the convexity of a rigidly convex algebraic interior \mathcal{C} is stable under deformation of the minimal defining polynomial p as long as the projective real algebraic hypersurface defined by p (the projective closure in $\mathbb{P}^m(\mathbb{R})$ of the Zariski closure of $\partial\mathcal{C}$ in \mathbb{R}^m) remains smooth. We conjecture that this property characterizes rigidly convex algebraic interiors; more precisely, we make the following conjecture:

Let $\mathcal{C}_0 = \mathcal{C}_{p_0}$ be a convex algebraic interior of degree d in \mathbb{R}^m that is not rigidly convex and such that the polynomial p_0 defines a smooth projective real algebraic hypersurface. Then there exists a continuous path through polynomials of degree d in m variables from a polynomial p_0 to a polynomial p_1 such that each polynomial p in the path defines a smooth projective real algebraic hypersurface and $p(0) > 0$, and the algebraic interior \mathcal{C}_{p_1} is not convex.

For more information about various connected components of the set of polynomials of degree d in m variables that define a smooth projective real algebraic hypersurface, see [18].

7 Appendix: Proofs of Topological Properties

We first establish property (1) of Section 5.3 and use it to prove Theorem 5.2. Then we establish properties (2) through (5).

PROOF OF PROPERTY (1): This crucial arcwise connectivity property follows from the results of Nuij [13] on hyperbolic polynomials. Recall that a homogeneous polynomial P in $m + 1$ variables is said to be hyperbolic with respect to $X^0 = (X_0^0, X_1^0, \dots, X_m^0) \in \mathbb{R}^{m+1}$ if $P(X^0) \neq 0$ and for each $X \in \mathbb{R}^{m+1}$ all the zeroes λ of $P(X + \lambda X^0)$ are real. It is called strictly hyperbolic if all the zeroes are furthermore simple for each $X \in \mathbb{R}^{m+1}$ (not a multiple of X^0), and it is called normalized if $P(X^0) = 1$.

It is immediate that a polynomial p in m variables satisfies RZ_{x^0} and $p(x^0) \neq 0$ for $x^0 = (x_1^0, \dots, x_m^0)$ if and only if its homogenization P is hyperbolic with respect to $X^0 = (1, x_1^0, \dots, x_m^0)$. Notice that if P is strictly hyperbolic, then necessarily p defines a smooth projective real algebraic hypersurface V .² For (as in the proof of Theorem 3.1) the zeroes of $P(X + \lambda X^0)$ correspond to intersections of the line through X and X^0 with V ; if X were a singular point of V , the intersection multiplicity at X of any line through X with V would be greater than 1 and the corresponding zero of $P(X + \lambda X^0)$ could not be simple.

Property (1) now follows from the following results of [13] (where the polynomials are of degree d and are hyperbolic with respect to a fixed X^0):

- (i) the set of strictly hyperbolic homogeneous polynomials is open,
- (ii) every hyperbolic homogeneous polynomial is the endpoint of an open continuous path of strictly hyperbolic homogeneous polynomials, and
- (iii) the set of normalized strictly hyperbolic homogeneous polynomials is arcwise connected.

Notice that since $X^0 = (1, x_1^0, \dots, x_m^0)$, homogeneous polynomials that are divisible by X_0 form an affine subspace of codimension at least 2 in the affine space of homogeneous polynomials P with $P(X^0) = 1$; so the set of normalized strictly

²It will follow from Theorem 5.2 that the converse is also true—if p defines a smooth projective real algebraic hypersurface V , then P is strictly hyperbolic.

hyperbolic homogeneous polynomials that are not divisible by X_0 is still arcwise connected. \square

PROOF OF THEOREM 5.2, FIRST DIRECTION: Assume that p_0 is a RZ polynomial of degree d in m variables with $p_0(0) > 0$ that defines a smooth projective real algebraic hypersurface V_0 in $\mathbb{P}^m(\mathbb{R})$. Let

$$p_1(x_1, \dots, x_m) = (r_1^2 - x_1^2 - \dots - x_m^2) \cdots (r_k^2 - x_1^2 - \dots - x_m^2)$$

if $d = 2k$ is even and

$$p_1(x_1, \dots, x_m) = (r_1^2 - x_1^2 - \dots - x_m^2) \cdots (r_k^2 - x_1^2 - \dots - x_m^2)(r_{k+1} - x_1)$$

if $d = 2k + 1$ is odd for some $0 < r_1 < \dots < r_k < r_{k+1}$. By property (1), there exists a continuous path $\{p_t\}_{0 \leq t \leq 1}$ from p_0 to p_1 through RZ polynomials of degree d in m variables such that $p_t(0) > 0$ and p_t defines a smooth projective real algebraic hypersurface V_t in $\mathbb{P}^m(\mathbb{R})$ for each t . It follows from the Thom isotopy theorem (see [4, theorem 14.1.1] or [1, theorem 20.2]) that there exists a homeomorphism H of $[0, 1] \times \mathbb{P}^m(\mathbb{R})$ such that $H_t = H|_{\{t\} \times \mathbb{P}^m(\mathbb{R})}$ is a homeomorphism of $\mathbb{P}^m(\mathbb{R})$ for every t , $H_0 = \text{Id}_{\mathbb{P}^m(\mathbb{R})}$, and $H_t(V_0) = V_t$. Therefore V_0 is isotopic to V_1 , and V_1 by construction is a disjoint union of k nested ovaloids if $d = 2k$ and a disjoint union of k nested ovaloids and a pseudohyperplane if $d = 2k + 1$. Thus so is V_0 , as is required.

It remains only to show that 0 is contained in the interior of the innermost ovaloid. Let $x^t = H_t^{-1}(0)$; since $p_t(0) > 0$ for all t , $0 \notin V_t$ and hence $x^t \notin V_0$. Thus x^t is a continuous path in $\mathbb{P}^m(\mathbb{R}) \setminus V_0$. Since 0 obviously belongs to the interior of the innermost sphere in V_1 , it follows that x^1 belongs to the interior of the innermost ovaloid in V_0 . Therefore so does $x^0 = 0$. \square

PROOF OF THEOREM 5.2, SECOND DIRECTION: Now let p be a polynomial of degree d in m variables that defines a smooth projective real algebraic hypersurface V in $\mathbb{P}^m(\mathbb{R})$ that is a disjoint union of k nested ovaloids W_1, \dots, W_k if $d = 2k$ and a disjoint union of k nested ovaloids W_1, \dots, W_k and a pseudohyperplane W_{k+1} if $d = 2k + 1$. Assume that X^0 is in the interior of W_1 , and let L be a (projective) line in $\mathbb{P}^m(\mathbb{R})$ through X^0 .

It is obvious that there exists a pseudohyperplane W contained in the exterior of the outermost ovaloid W_k (just take the image of a hyperplane contained in the exterior of a sphere S in $\mathbb{P}^m(\mathbb{R})$). Since the $\mathbb{Z}/2$ fundamental class of W coincides with that of a hyperplane, it follows that the $\mathbb{Z}/2$ intersection number of L and W equals 1; hence $L \cap W \neq \emptyset$. Let $X \in L \cap W$; in case $d = 2k + 1$ is odd, we take $W = W_{k+1}$ and denote $X^{k+1} = X$.

The points X^0 and X decompose the projective line L into two closed intervals L_+ and L_- . Each of L_\pm is a continuous path between X^0 , which lies in the interior of W_1 , and X , which lies in the exterior of W_k ; therefore L_\pm intersects each of W_1, \dots, W_k in at least one point X_\pm^1, \dots, X_\pm^k . Adding to these X^{k+1} in case $d = 2k + 1$ is odd, we obtain d distinct points of intersection of L with V ; since the

defining polynomial p of V has degree d , $L \cap V$ can contain at most d points, so we conclude that $L \cap V$ contains exactly d distinct points $X_{\pm}^1, \dots, X_{\pm}^k$ and (in case $d = 2k + 1$ is odd) X^{k+1} . This completes the proof. Notice that it follows that on L_{\pm} we have the following picture: the interval $[X^0, X_{\pm}^1]$ lies in the interior of W_1 , the interval (X_{\pm}^1, X_{\pm}^2) lies in the interior of W_2 intersected with the exterior of W_1 , \dots , the interval $(X_{\pm}^{k-1}, X_{\pm}^k)$ lies in the interior of W_k intersected with the exterior of W_{k-1} , and the interval $(X_{\pm}^k, X]$ lies in the exterior of W_k . \square

PROOF OF PROPERTY (2): By property (1), we may assume, using approximation, that the polynomial p defines a smooth projective real algebraic hypersurface. The conclusion now follows immediately from Theorem 5.2. \square

PROOF OF PROPERTY (3): We may assume, again, that the polynomial p defines a smooth projective real algebraic hypersurface V in $\mathbb{P}^m(\mathbb{R})$. By Theorem 5.2, V is a disjoint union of k nested ovaloids W_1, \dots, W_k if $d = 2k$ and a disjoint union of k nested ovaloids W_1, \dots, W_k and a pseudohyperplane W_{k+1} if $d = 2k + 1$; $X^0 = (1, 0, \dots, 0)$ (we are using here projective coordinates) lies in the interior \mathcal{I} of W_1 .

It is obvious that \mathcal{C}_p is the closure of the connected component of X^0 in $\mathcal{I} \setminus \mathcal{I} \cap H_{\infty}$ where $H_{\infty} = \{X_0 = 0\}$ is the hyperplane at infinity. We shall show that either $\mathcal{I} \cap H_{\infty} = \emptyset$, in which case $\mathcal{I} \subset \mathbb{R}^m$ is convex, or $\mathcal{I} \setminus \mathcal{I} \cap H_{\infty}$ consists of two connected convex components in \mathbb{R}^m .

Before proceeding, let us notice that there exists a hyperplane H in $\mathbb{P}^m(\mathbb{R})$ such that $H \cap \mathcal{I} = \emptyset$. We may take H to be the (projectivized) tangent hyperplane to V at a point $Y^0 \in W_1$. If H contained a point Y in \mathcal{I} , then by Theorem 5.2 the line through Y and Y^0 would intersect V in d distinct points; but since this line is contained in H , it has intersection multiplicity at least 2 with V at Y^0 , so it cannot intersect V in more than $d - 1$ distinct points.

Let now $Y^1, Y^2 \in \mathcal{I} \setminus \mathcal{I} \cap H_{\infty}$. Let L be the projective line through Y^1 and Y^2 , and let $X^{\infty} = L \cap H_{\infty}$. The two points Y^1 and Y^2 decompose the projective line L into two closed intervals; the line segment between Y^1 and Y^2 in \mathbb{R}^m is the interval that does not contain the point X^{∞} . Notice that by the last sentence in the proof of Theorem 5.2, exactly one of the two intervals is contained in \mathcal{I} ; let us denote it by $L_{\mathcal{I}}$.

Assume that $\mathcal{I} \cap H_{\infty} = \emptyset$. Then $X^{\infty} \notin \mathcal{I}$ and therefore $X^{\infty} \notin L_{\mathcal{I}}$. It follows that $L_{\mathcal{I}}$ is the line segment between Y^1 and Y^2 in \mathbb{R}^m , and $\mathcal{I} \subset \mathbb{R}^m$ is convex.

If $\mathcal{I} \cap H_{\infty} \neq \emptyset$, let $a_0X_0 + a_1X_1 + \dots + a_mX_m = 0$ be a hyperplane in $\mathbb{P}^m(\mathbb{R})$ that does not intersect \mathcal{I} . Let \mathcal{I}_+ and \mathcal{I}_- be the nonempty open subsets of $\mathcal{I} \setminus \mathcal{I} \cap H_{\infty}$ where $X_0/(a_0X_0 + a_1X_1 + \dots + a_mX_m) > 0$ and $X_0/(a_0X_0 + a_1X_1 + \dots + a_mX_m) < 0$, respectively. We shall show that \mathcal{I}_+ and \mathcal{I}_- are convex subsets of \mathbb{R}^m , thereby completing the proof. It is enough to show that if the line segment between Y^1 and Y^2 in \mathbb{R}^m is not contained in \mathcal{I} and say $Y^1 \in \mathcal{I}_+$, then necessarily $Y^2 \in \mathcal{I}_-$.

Assume then that the line segment between Y^1 and Y^2 in \mathbb{R}^m is not contained in \mathcal{I} . This means that this line segment is not $L_{\mathcal{I}}$, so $X^\infty \in L_{\mathcal{I}}$. It is now immediate that the function $X_0/(a_0X_0+a_1X_1+\cdots+a_mX_m)$ on $L_{\mathcal{I}}$ changes its sign at X^∞ . \square

PROOF OF PROPERTY (4): This follows immediately from the facts that a limit of RZ polynomials is a RZ polynomial (see the last sentence in the proof of Theorem 3.1), that a RZ polynomial p , $p(0) > 0$, defines a smooth projective real algebraic hypersurface if and only if its homogenization P is strictly hyperbolic (with respect to $X^0 = (1, 0, \dots, 0)$), and that the set of strictly hyperbolic homogeneous polynomials is open [13]. \square

(We could also prove property (4) by using Theorem 5.2 and an isotopy argument, like the proof of property (5) below.)

PROOF OF PROPERTY (5): By the Thom isotopy theorem, smooth projective real algebraic hypersurfaces V_0 and V_1 in $\mathbb{P}^m(\mathbb{R})$ defined by the polynomials p_0 and p_1 , respectively, are isotopic; the result now follows from Theorem 5.2. \square

8 Remarks on the History of the Proof of the Lax Conjecture

In 1958 Peter Lax conjectured that a homogeneous polynomial P with real coefficients of degree d in three variables x_0 , x_1 , and x_2 is (normalized) hyperbolic (see Section 7) with respect to $(c_0, c_1, c_2) \in \mathbb{R}^3$ if and only if it can be written as a determinant

$$P(x_0, x_1, x_2) = \det(x_0A_0 + x_1A_1 + x_2A_2)$$

where A_0 , A_1 , and A_2 are $d \times d$ real symmetric matrices with $c_0A_0 + c_1A_1 + c_2A_2 = I$. This was recently established by Lewis, Parrilo, and Ramana [11] based on the positive determinantal representation underlying Theorem 2.2.

The study of representations of a polynomial as a determinant of a matrix of linear forms has a long history; see, e.g., Dixon [7], Wall [19], and Cook and Thomas [6]. Dubrovin in [8, theorems 6.3 and 6.4] proved under an additional smoothness assumption the existence of a determinantal representation of the above form with complex hermitian (rather than real symmetric) matrices. He treated an (irreducible) homogeneous polynomial p for which the corresponding (irreducible) complex projective plane algebraic curve is smooth, i.e., has no singular points. This result also follows from the later work of Vinnikov [17, theorem 6.1], which gave a complete description of these determinantal representations in terms of a certain real torus in the Jacobian variety of the curve. [17] also contains some results on determinantal representations with real symmetric A_j .

Theorem 2.2 removes the smoothness assumptions and passes from complex hermitian matrices to real symmetric ones. This is essential for proving the Lax conjecture as well as for the description of convex sets with an LMI representation. As already noted, our conjecture in Section 6.1 is a natural modification of the Lax conjecture to homogeneous hyperbolic polynomials in more than three variables (see also [11]).

Acknowledgment. We are grateful to Bernd Sturmfels for suggesting this problem and to Pablo Parrilo for discussions bearing on possible extensions of these results.

Thanks for Helton's support are due to the National Science Foundation, the Defense Advanced Research Projects Agency, and the Ford Motor Company.

Thanks for Vinnikov's support are due to the Israel Science Foundation.

Bibliography

- [1] Abraham, R.; Robbin, J. *Transversal mappings and flows*. W. A. Benjamin, New York-Amsterdam, 1967.
- [2] Ball, J. A.; Vinnikov, V. Zero-pole interpolation for meromorphic matrix functions on an algebraic curve and transfer functions of 2D systems. *Acta Appl. Math.* **45** (1996), no. 3, 239–316.
- [3] Ball, J. A.; Vinnikov, V. Zero-pole interpolation for matrix meromorphic functions on a compact Riemann surface and a matrix Fay trisecant identity. *Amer. J. Math.* **121** (1999), no. 4, 841–888.
- [4] Bochnak, J.; Coste, M.; Roy, M.-F. *Real algebraic geometry*. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), 36. Springer, Berlin, 1998.
- [6] Cook, R. J.; Thomas, A. D. Line bundles and homogeneous matrices. *Quart. J. Math. Oxford Ser. (2)* **30** (1979), no. 120, 423–429.
- [7] Dixon, A. C. Note on the reduction of ternary quartic to a symmetrical determinant. *Proc. Cambridge Phil. Soc.* **2** (1900-1902), 350–351.
- [8] Dubrovin, B. A. Matrix finite-gap operators. Current Problems in Mathematics, vol. 23, 33–78. *Itogi Nauki i Tekhniki*, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1983.
- [9] El Ghaoui, L.; Niculescu, S.-I., eds. *Advances in linear matrix inequality methods in control*. Advances in Design and Control. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2000.
- [10] Helton, J. W. Manipulating matrix inequalities automatically. *Mathematical systems theory in biology, communications, computation, and finance (Notre Dame, IN, 2002)*, 237–256. The IMA Volumes in Mathematics and Its Applications, 134. Springer, New York, 2003.
- [11] Lewis, A. S.; Parrilo, P. A.; Ramana, M. V. The Lax conjecture is true. *Proc. Amer. Math. Soc.* **133** (2005), no. 9, 2495–2499.
- [12] Nesterov, Y.; Nimmerovski, A. *Interior-point polynomial algorithms in convex programming*. SIAM Studies in Applied Mathematics, 13. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1994.
- [13] Nuij, W. A note on hyperbolic polynomials. *Math. Scand.* **23** (1968), 69–72.
- [14] Parrilo, P. Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Doctoral dissertation, California Institute of Technology, May 2000. Available at: <http://www.mit.edu/~parrilo/pubs/index.html>
- [15] Parrilo, P. A.; Sturmfels, B. Minimizing polynomial functions. *Algorithmic and quantitative real algebraic geometry (Piscataway, NJ, 2001)*, 83–99. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 60. American Mathematical Society, Providence, R.I., 2003.
- [16] Skelton, R. E.; Iwasaki, T.; Grigoriadis, K. M. *A unified algebraic approach to linear control design*. The Taylor & Francis Systems and Control Book Series. Taylor & Francis, London, 1998.
- [17] Vinnikov, V. Selfadjoint determinantal representations of real plane curves. *Math. Ann.* **296** (1993), no. 3, 453–479.

- [18] Viro, O. Ya. Achievements in the topology of real algebraic varieties in the last six years. *Russian Math. Surveys* **41** (1986), 55–82.
- [19] Wall, C. T. C. Nets of quadrics, and theta-characteristics of singular curves. *Philos. Trans. Roy. Soc. London Ser. A* **289** (1978), no. 1357, 229–269.

J. WILLIAM HELTON
University of California, San Diego
Department of Mathematics
9500 Gilman Drive
La Jolla, CA 92093-0112
E-mail: helton@math.ucsd.edu

VICTOR VINNIKOV
Ben Gurion University of the Negev
Department of Mathematics
P.O.B. 653
Be'er Sheva 84105
ISRAEL
E-mail: vinnikov@
math.bgu.ac.il

Received May 2005.