hereditarily
(1976)
35, 140-143 (1980)
en Ringen.

semi-algebraic
)
deals.
quadratic forms.

du spectre réel.

t topos étale

ischer Räume.

topology over a
algebraic spaces.

tures and reduced
burg 51 149-195

igen Körpern.

real algebraic
metry and

TRANSVERSAL ZEROS AND POSITIVE SEMIDEFINITE FORMS

by

Man-Duen Choi*), Manfred Knebusch**),

Tsit-Yuen Lam***), Bruce Reznick***).

Introduction

For any natural number $n \geq 2$ and any even natural number $d \geq 2$ we consider the convex cone $P(n,d)$ consisting of the positive semidefinite (= psd) forms over $\mathbb{R}$ in n variables $x_1,\ldots,x_n$ of degree d, and the convex subcone $\Sigma(n,d)$ consisting of the finite sums of squares of forms of degree $d/2$ in the variables $x_1,\ldots,x_n$. As is well known $\Sigma(n,d) \neq P(n,d)$ except for very special pairs (n,d), namely the pairs with $n = 2$ or $d = 2$ or $(n,d) = (3,4)$ (Hilbert, cf. [CL] for an elementary proof).

In this paper we ask for relations between the sets $EP(n,d)$ and $E\Sigma(n,d)$ of extremal elements of the cones $P(n,d)$ and $\Sigma(n,d)$. Notice that, since our cones are closed (after adding the origin), every element in $P(n,d)$ resp. $\Sigma(n,d)$ is a finite sum of elements in $EP(n,d)$ resp. $E\Sigma(n,d)$. Thus the sets $EP(n,d)$ and $E\Sigma(n,d)$ deserve special attention.

Our main result, Theorem 6.1 in §6, is the determination of all pairs (n,d) such that $E\Sigma(n,d)$ is contained in $EP(n,d)$, which means $E\Sigma(n,d) = EP(n,d) \cap \Sigma(n,d)$. This answers Problem B in the survey article $[CL_1]$.

In order to obtain the result a general observation turns out to be helpful:

a) Let H be an irreducible indefinite form in $\mathbb{R}[x_1,\ldots,x_n]$ of degree r. Then for any $F \in P(n,d)$

$$F \in EP(n,d) \Longleftrightarrow FH^2 \in EP(n,d+2r),$$

$$F \in E\Sigma(n,d) \Longleftrightarrow FH^2 \in E\Sigma(n,d+2r),$$

*) supported by NSERC of Canada
**) supported by DFG during a stay at Berkeley 1980
***) supported by NSF

cf. Theorem 5.1. We also feel that the following observation sheds light on the problem:

b) If $F \in EP(n,d)$ then $F^2 \in E\Sigma(n,2d)$,

cf. Theorem 5.2.

Our "counterexamples" $G \in E\Sigma(n,d)$, $G \notin EP(n,d)$ are of the form $G = H^2F^2$ with H a product of irreducible indefinite forms and F an irreducible psd form of some degree e which is not extremal in $P(n,e)$. Basic counterexamples will be explicitly constructed in §6 for $(n,d) = (3,12)$ and $(n,d) = (4,8)$.

The observations a) and b) rely on the presence of "transversal zeros" for some forms coming up in the proofs. A transversal zero of a polynomial $F(x_1,\ldots,x_n)$ over $\mathbb{R}$ is a point $c \in \mathbb{R}^n$ such that F changes sign in every neighbourhood of c. If F has no multiple irreducible factors then a point c of the zero set $Z(F) \subset \mathbb{R}^n$ turns out to be a transversal zero if and only if $Z(F)$ has local dimension n-1 at c, cf. Theorem 3.4.

The first half of our paper is devoted to a geometric study of transversal zeros and to the question how far a polynomial is determined by its transversal zeros. We try to do all this on a natural level of generality. This leads us to study the set $|D|_{\mathbb{R}}$ of real points of an effective Weil divisor D on a normal algebraic variety X over $\mathbb{R}$. But for the applications of the theory of transversal zeros made in §5 and §6 it suffices to consider the case when X is a projective space $\mathbb{P}^{n-1}_{\mathbb{R}}$, or - if one wants to study also multiforms - a direct product of projective spaces.

We suspect that many of our considerations on transversal zeros are more or less "folklore", well known to the experts. However, to our knowledge, no coherent account of this useful theory seems to exist in the literature. Thus we feel that these Proceedings are a good place to explicate the basic facts.

In the whole paper we admit any real closed field R as ground field instead of the field $\mathbb{R}$ of real numbers. Using some standard results from semialgebraic topology, all contained in [DK] and §1 of the present paper, this does not cause additional difficulties. Thus we never need Tarski's principle to transfer elementary statements from $\mathbb{R}$ to other real closed fields.

§ 1 The pure dimensional parts of a semialgebraic set

We start with a variety X over a real closed field R, i.e. a reduced algebraic scheme over R. The set X(R) of rational points of X is a semialgebraic space in the sense of [DK], and we use freely the language of "semialgebraic topology" developed in that paper. In particular we make use of the dimension theory in [DK, §8].

Let N be a semialgebraic subset of X(R). For any point x of N the local dimension $\dim_x N$ of N at x is defined as the infimum of the dimensions of all semialgebraic neighbourhoods of x in N [DK, §13]. We introduce the sets (k = 0,1,2,...)

$$\Sigma_k(N) := \{x \in N \mid \dim_x N \geq k\}.$$

Of course $\Sigma_k(N)$ is empty if k exceeds the dimension d of N. It is clear from [DK, §8] that every $\Sigma_k(N)$ is a closed subset of N (in the strong topology, as always). We shall need some elementary facts about the sets $\Sigma_k(N)$ (actually only about $\Sigma_d(N)$), not covered by the paper [DK].

Proposition 1.1. $\Sigma_k(N)$ is semialgebraic for every $k \geq 0$.

It is trivial to verify this lemma using the theorem that every affine semialgebraic space can be triangulated [DK_1]. A more elementary proof, which also gives additional insight, runs as follows. Let d = dim(N). For k > d there is nothing to prove. We now deal with the case k = d. We may assume that X is affine. Let Y denote the Zariski closure of N in X, and let S denote the singular locus of Y. Then

$$N' := (Y(R) \setminus S(R)) \cap N$$

is an open semialgebraic subset of N and the complement in N, i.e. $N \cap S(R)$, has dimension at most d-1. Suppose we know already that $\Sigma_d(N')$ is semialgebraic. Let L be the closure of $\Sigma_d(N')$ in N. This is again a semialgebraic set. $N \setminus L$ is open in N and has dimension at most d-1. Thus $N \setminus L$ is disjoint from $\Sigma_d(N)$. On the other hand L is contained in $\Sigma_d(N)$, since $\Sigma_d(N)$ is closed and contains $\Sigma_d(N')$. Thus $\Sigma_d(N)$ coincides with the semialgebraic set L.

Replacing N by N' and X by $X \setminus S$ we assume now that Y is smooth. Let $Y_1,\dots,Y_t$ denote the connected components of Y. The set $\Sigma_d(N)$ is the union of the sets $\Sigma_d(N \cap Y_i(R))$, and it suffices to prove that these sets are semialgebraic. $N \cap Y_i(R)$ is Zariski dense in Y_i. Re-

placing N by anyone of the sets $N \cap Y_i(R)$ we assume that in addition Y is connected, hence irreducible.

We have $N = N_1 \cup \ldots \cup N_r$ with non empty sets

$$N_i = \{x \in Y(R) \mid g_i(x) = 0,\ f_{ij}(x) > 0,\ j = 1,\ldots,s_i\},$$

where g_i, f_{ij} are functions in the affine ring $R[Y]$. If g_i is not zero then $\dim N_i \leq n-1$. But if g_i is zero then N_i is open in $Y(R)$, hence $N_i \subset \Sigma_d(N)$, since Y is smooth and thus $Y(R)$ has local dimension d at every point [DK, §8]. It is now clear that $\Sigma_d(N)$ is the closure of the union of all N_i with $g_i = 0$ in the set N. Thus $\Sigma_d(N)$ is indeed semialgebraic.

Consider now the open semialgebraic subset $N_1 := N \setminus \Sigma_d(N)$ of N. Clearly

$$\Sigma_{d-1}(N) = \Sigma_d(N) \cup \Sigma_{d-1}(N_1).$$

We know from the proof already given that $\Sigma_{d-1}(N_1)$ and $\Sigma_d(N)$ are semialgebraic. Thus $\Sigma_{d-1}(N)$ is semialgebraic. Repeating this argument we see that all $\Sigma_\kappa(N)$ are semialgebraic, and our lemma is proved.

Proposition 1.2. For every $k \geq 0$ the semialgebraic set

$$\Sigma_k^o(N) := \Sigma_k(N) \setminus \Sigma_{k+1}(N),$$

consisting of all points $x \in N$ with $\dim_x N = k$, is pure of dimension k, i.e. $\dim_x \Sigma_k^o(N) = k$ for every $x \in \Sigma_k^o(N)$.

Proof. Let x be a point of $\Sigma_k^o(N)$ and let U_o be an open semialgebraic neighbourhood of x in N with $\dim U_o = k$. For any open semialgebraic neighbourhood $U \subset U_o$ of x in N we then also have $\dim U = k$. Moreover for every such U there exists an open semialgebraic subset V of U which is semialgebraically isomorphic to an open non empty subset of R^k [DK, §8]. Clearly V is contained in $\Sigma_k^o(N) \cap U$. Thus $\dim(\Sigma_k^o(N) \cap U) = k$. Q.E.D.

We call $\Sigma_k^o(N)$ the pure k-dimensional part of N. More specifically, if $\dim N = d$ we call $\Sigma_d^o(N) = \Sigma_d(N)$ the pure part of N.

Example 2.3. If X is irreducible of dimension n, and if the set $X(R)_{reg}$ of regular points of X in X(R) is not empty, then the pure part $\Sigma_n(X(R))$ of X(R) is the closure of $X(R)_{reg}$ in X(R).

Indeed, $X(R)_{reg}$ is pure of dimension n, and X(R) has local dimension at most n-1 at every singular point which is not contained in the closure of $X(R)_{reg}$.

§ 2 Transversal zeros of algebraic functions

We assume in this section that the variety X over R is irreducible, that the set X(R) of real points is not empty, and that X is regular at every point of X(R). Then X(R) is an n-dimensional semialgebraic manifold [DK, §13] with n = dim X. We also assume that X is affine, and we denote the ring R[X] of regular functions on X by A. On the space X(R) every $f \in A$ takes values in R. We are interested in the zeros and the sign behaviour of the functions $f : X(R) \to R$.

Definition 2.1. Let L be a subset of X(R) on which f does not vanish everywhere. We say that f is positive semidefinite (resp. positive definite) on L if $f(x) \geq 0$ (resp. $f(x) > 0$) for all $x \in L$. In the same way we use the words "negative semidefinite" and "negative definite". If there exist points $x \in L$ and $y \in L$ with $f(x) > 0$ and $f(y) < 0$, then we call f indefinite on L.

Definition 2.2. Let f be a non zero element of A. A transversal zero of f is a point $x \in X(R)$ such that f is indefinite on every semialgebraic neighbourhood V of x in X(R). Notice that f cannot vanish everywhere on V since dim V = n.

We denote by Z(f) the set of zeros of f on X(R) and by $Z_t(f)$ the set of transversal zeros of f. We finally denote by N(f) the closed reduced subscheme of all zeros of f on X. Thus Z(f) is the set of real points of N(f) and $Z_t(f)$ is a subset of Z(f). The set Z(f) is closed and semialgebraic in X(R). The set $Z_t(f)$ is the intersection of the closure of the set of points of X(R) where f is positive with the closure of the set where f is negative. Thus $Z_t(f)$ is also closed and semialgebraic in X(R).

Proposition 2.3. For every non zero regular function f on X the set $Z_t(f)$ of transversal zeros is either empty or pure of dimension n-1.

Proof. Let a be a given point of $Z_t(f)$. We choose an open neighbourhood V of a in X(R) with a semialgebraic isomorphism $\varphi : V \xrightarrow{\sim} V'$ onto an open semialgebraic convex subset V' of R^n. (Recall that X(R) is a semialgebraic manifold.) We then choose a point $x_o \in V$ with $f(x_o) > 0$

and an open semialgebraic subset $U \subset V$ such that $f(y) < 0$ for every $y \in U$ and such that $U' := \varphi(U)$ is convex in R^n. We finally choose a hyperplane H of R^n with $H \cap U' \neq \emptyset$ and not containing the point $x_o' := \varphi(x_o)$. Now consider the central projection

$$\pi : R^n \setminus \{x_o'\} \to H$$

onto H with center x_o'. We claim that

$$(*) \qquad \pi \circ \varphi(Z_t(f) \cap V) \supset H \cap U'.$$

Indeed, let $y' \in H \cap U'$ be given and let $\gamma' : [0,1] \to V'$ be the straight path from x_o' to y',

$$\gamma'(t) = (1-t)x_o' + ty'.$$

Then $\gamma := \varphi^{-1} \circ \gamma'$ is a semialgebraic path in V running from the point x_o to the preimage y of y'. Since $f(x_o) > 0$ and $f(y) < 0$ there exists some point $\tau \in]0,1[$ where the semialgebraic function $f \circ \gamma$ on $[0,1]$ changes sign. $\gamma(\tau)$ is clearly a transversal zero of f. The point $\gamma'(\tau)$ lies in $\varphi(Z_t(f) \cap V)$ and maps under π to the point y'. Thus the inclusion $(*)$ holds true. This implies that

$$\dim Z_t(f) \cap V \geq n-1,$$

since $\dim (H \cap U') = n-1$. But $Z(f) \cap V$ has dimension at most $n-1$ since this set is contained in $N(f)$. Thus $Z_t(f) \cap V$ has dimension $n-1$ for every open semialgebraic neighbourhood V of a.

Q.E.D.

Corollary 2.4. Let f and g be non zero regular functions on X. Let $a \in X(R)$ be a transversal zero of f and assume that $Z_t(f) \cap U$ is contained in $Z(g)$ for some neighbourhood U of a. Then f and g have a non trivial common factor in the regular local ring $\mathcal{O}_{X,a}$. {Recall that $\mathcal{O}_{X,a}$ is a unique factorization domain.}

Proof. For every affine Zariski neighbourhood W of a in X the semialgebraic set $W \cap U \cap Z_t(f)$ has dimension $n-1$ by Proposition 2.3 above. Our hypothesis implies that this set is contained in the intersection $N(f) \cap N(g) \cap W$ of the hypersurfaces $f = 0$ and $g = 0$ on W. Thus the (algebraic!) dimension of $N(f) \cap N(g) \cap W$ cannot be smaller than $n-1$ for any Zariski neighbourhood W of a. This implies that there exists some $h \in A$ which is a prime element in $\mathcal{O}_{X,a}$ and has the property that

$$N(h) \cap W \subset N(f) \cap N(g) \cap W$$

for small Zariski neighbourhoods W of a. By the local Nullstellensatz h divides both f and g in $\mathcal{O}_{X,a}$.

In the same vein we obtain

Corollary 2.5. Let again f and g be non zero functions on X. Suppose that for some open semialgebraic subset U of X(R) the set $Z_t(f) \cap U$ is not empty and contained in $Z(g)$. Then the complex hypersurfaces N(f) and N(g) have a common irreducible component. In particular, if A is factorial then f and g have a non trivial common factor in A.

Proposition 2.6. Let f and g be non zero regular functions on X, and assume that the hypersurfaces N(f) and N(g) have no irreducible component in common. Then

$$Z_t(fg) = Z_t(f) \cup Z_t(g).$$

Proof. a) Let $\underline{a}$ be a point of X(R) which is not contained in $Z_t(f) \cup Z_t(g)$. Then there exists a neighbourhood U of $\underline{a}$ in X(R) such that both f and g are semidefinite on U (positive or negative). Then also the product fg is semidefinite on U, and $\underline{a}$ is not a transversal zero of fg. This proves that $Z_t(fg)$ is contained in $Z_t(f) \cup Z_t(g)$. (Our hypothesis, that N(f) and N(g) have no common component, is not yet needed for that.)

b) We show that the set $M := Z_t(f)$ is contained in $Z_t(fg)$, which will finish the proof. We may assume that M is not empty. By Proposition 2.3 M is pure of dimension n-1. On the other hand the set $N := Z_t(f) \cap Z_t(g)$ has dimension at most n-2, since N is contained in the intersection of the hypersurfaces N(f) and N(g) which have no common irreducible component. Thus the set $M \setminus N$ is dense in M (a trivial argument, cf. [DK, §13]). Since $Z_t(fg)$ is closed it suffices to verify that $M \setminus N$ is contained in $Z_t(fg)$.

Let x be a point of $M \setminus N$, which means that $x \in Z_t(f)$, $x \notin Z_t(g)$. We choose a neighbourhood U_o of x on which g is semidefinite. Now f is indefinite on every neighbourhood $U \subset U_o$ of x. Thus also fg is indefinite on every such U. This implies that $x \in Z_t(fg)$.

Q.e.d.

Corollary 2.7. Assume that A is factorial. Let f be a non zero element of A and let

$$f = u\, p_1^{e_1} \cdots p_t^{e_t}$$

be the decomposition of f into powers of pairwise non associated prime elements $p_1, \ldots, p_t$, with u a unit of A. Then $Z_t(f)$ is the union of the sets $Z_t(p_i)$ with e_i odd.

Proof. Apply Proposition 2.6 and observe that $Z_t(p_i^{e_i})$ is empty if e_i is even and $Z_t(p_i^{e_i}) = Z_t(p_i)$ if e_i is odd.

In the same vein we obtain for the semialgebraic set germ $Z_t(f)_a$ of a non zero function $f \in A$ at any point $a \in X(R)$:

Corollary 2.8. Let

$$f = u\, p_1^{e_1} \cdots p_t^{e_t}$$

be the decomposition of f into prime elements in the factorial ring $\mathcal{O}_{X,a}$. Then $Z_t(f)_a$ is the union of the set germs $Z_t(p_i)_a$ with e_i odd.

§ 3 Purely indefinite divisors

We still assume that X is an irreducible n-dimensional variety over R and that the set X(R) is not empty and contains no singular points of X. But we no longer assume that X is affine. Our terminology from §2 then takes over from functions to effective divisors $D \geq 0$ on X, by which we always mean effective Weil divisors.

Definition 3.1. Let D be an effective divisor on X and let a be a point of X(R). Let f be the local equation of D on some affine Zariski open neighbourhood V of a. We call D indefinite at a, if f is indefinite on every neighbourhood of a in V(R). Similarly we call D semidefinite (resp. definite) at a, if f is positive or negative semidefinite (resp. definite) on some neighbourhood of a in V(R). The points of X(R) where D is indefinite are called the transversal points of D and the set of these points is denoted by $|D|_t$. This set is a closed semialgebraic subset of the set of real points $|D|_R := |D| \cap X(R)$ of the support $|D|$ of D.

Let $D = e_1 D_1 + \ldots + e_t D_t$ be the decomposition of D into irreducible components.

Proposition 3.2. $|D|_t$ is the union of all sets $|D_i|_t$ with e_i odd.

This is clear from Proposition 2.6 in §2, or its corollary 2.8.

Definition 3.3. We call an effective divisor D indefinite, if $|D|_t$ is not empty, i.e. if D is indefinite at some point of X(R). We call D semidefinite, if $|D|_t$ is empty, and we call D definite if $|D|_R$ is empty. Finally, we call D purely indefinite, if $D \neq 0$ and there does not exist a semidefinite effective divisor $E \neq 0$ with $E \leq D$. This means that D is non zero, has no multiple components, and that all irreducible components of D are indefinite.

It is clear from Proposition 2.3 in §2 that for every effective divisor D on X the set $|D|_t$ is either empty or pure of dimension n-1. This result can be improved.

Theorem 3.4. Assume that D has no multiple components. Then the semialgebraic set $|D|_t$ of transversal points of D coincides with the pure (n-1)-dimensional part $\Sigma_{n-1}(|D|_R)$ of the set $|D|_R$ of real points on $|D|$.

Proof. It remains to verify that D is indefinite at any given point a of $|D|_R$ with $\dim_a |D|_R = n-1$. We choose a local equation f of D on some affine Zariski open neighbourhood W of a in X. Let U be any semialgebraic open neighbourhood of a in W(R). The set $U \cap |D|_R$ has dimension n-1, but the set of points in $|D|_R$ which are singular on $|D|$ has dimension at most n-2. Thus $U \cap |D|_R$ contains some regular point b of $|D|$. There exists a regular system of parameters $f_1, f_2, \ldots, f_n$ of the regular local ring $\mathcal{O}_{X,b}$ such that f_1 defines the germ of the variety $|D|$ at b. The functions f_1 and f differ in $\mathcal{O}_{X,b}$ only by a unit, hence we may assume that $f = f_1$. By the implicit function theorem the system $(f_1, \ldots, f_n)$ yields a semialgebraic isomorphism of some open semialgebraic neighbourhood $U' \subset U$ of b in X(R) onto some open semialgebraic subset of R^n. Since $f_1(b) = 0$ certainly $f = f_1$ changes sign on U'. A fortiori f is indefinite on U.

Q.e.d.

We mention that the theorem now proved implies a generalization of the "Sign-Changing Criterion" of Dubois and Efroymson for extending an ordering P of a field k to a given function field over k ([DE, Th.2.7], cf. also [ELW, §4 bis]

Corollary 3.5. (Dubois - Efroymson for $V = A_k^n$). Let k be an ordered field, R be a real closure of k with respect to the given ordering. Let V be an absolutely irreducible variety without singular points over k and D a prime divisor on V. Let V_R denote the variety over R obtained from V by base extension and let $\tilde{D}$ denote the effective divisor on V_R obtained from D by base extension. Then the ordering of k can be extended to the function field k(D) of D if and only if $\tilde{D}$ is indefinite.

Proof. The ordering of k extends to k(D) if and only if there exists a field composite of k(D) and R over k, which is formally real. These field composites are the function fields $R(D_1),\ldots,R(D_s)$ of the irreducible components $D_1,\ldots,D_s$ of the divisor $\tilde{D}$. The prime divisors D_i all occur with multiplicity one in $\tilde{D}$. Thus $\tilde{D}$ is indefinite if at least one D_i is indefinite. By Theorem 3.4 a given D_i is indefinite if and only if the set of real points $D_i(R)$ of D_i has dimension n-1 with $n := \dim V = \dim V_R$. But $\dim D_i(R) = n-1$ means that the variety D_i has nonsingular real points, cf. §1. Now it is a well known fact, due to Artin, that D_i has nonsingular real points if and only if the field $R(D_i)$ is formally real ([A, §4], cf. also [E]).

We return to our irreducible variety X over R.

Proposition 3.6. Let D be an effective divisor $\neq 0$ without multiple components. Then $|D|_R$ is Zariski dense in $|D|$ if and only if D is purely indefinite. In this case even $|D|_t$ is Zariski dense in $|D|$.

Proof. Let $D_1,\ldots,D_r$ denote the irreducible components of D. Clearly

$$|D|_R = D_1(R) \cup \ldots \cup D_r(R)$$

is Zariski dense in D if and only if every $D_i(R)$ is Zariski dense in D_i. This means that $D_i(R)$ has the semialgebraic dimension n-1, i.e. that $\Sigma_{n-1}(D_i(R))$ is not empty, and in that case of course already $\Sigma_{n-1}(D_i(R))$ is Zariski dense in D_i. The proposition now follows from the preceding Theorem 3.4.

This is perhaps the appropriate place to indicate a relation between our investigations and the real Nullstellensatz of Dubois-Risler-Stengle [S, Theorem 2]. Assume that X is an affine variety over R and that W is a closed subvariety of X. Let A denote the affine ring of X and $\mathfrak{a}$ the ideal of functions in A vanishing on W. Then the real Nullstellensatz says in particular that W(R) is Zariski dense in W if and

only if the ideal $\mathfrak{a}$ is "real", i.e.

$$h_1^2 + \ldots + h_r^2 \in \mathfrak{a} \Rightarrow h_1 \in \mathfrak{a}, \ldots, h_r \in \mathfrak{a}$$

for arbitrary elements $h_1, \ldots, h_r$ of A. (This is essentially Risler's version of the real Nullstellensatz [Ri], [Ri_1].) Thus if X is irreducible and has no singular real points then the proposition we just proved says the following:

Corollary 3.7. Let X be affine and I(D) denote the ideal of functions in R[X] vanishing on $|D|$ for D an effective divisor $\neq 0$ without multiple components. Then I(D) is real if and only if D is purely indefinite.

If D is a prime divisor then clearly I(D) is real if and only if the function field R(D) is formally real, and we are back to the arguments which led to the Sign-Changing Criterion above (Corollary 3.5).

Definition 3.8. We call a semialgebraic subset M of X(R) pure and full of dimension k in X, if dim M = k (hence the Zariski closure Z of M in X has dimension k) and M is the pure part $\Sigma_k(Z(R))$ of Z(R).

In this terminology we can say according to Theorem 3.4 and Proposition 3.6 that for every non zero purely indefinite divisor on X the set $|D|_t$ is pure and full of dimension n-1 in X. We now prove a converse of this statement.

Theorem 3.9. Let M be a pure and full (n-1)-dimensional semialgebraic subset of X(R). Then there exists a unique purely indefinite divisor D on X such that M coincides with the set $|D|_t$ of transversal points of D. The variety $|D|$ is the Zariski closure of M in X.

Proof. Let Z denote the Zariski closure of M in X and let $Z_1, \ldots, Z_r$ denote the irreducible components of Z. The set M is the union of the closed semialgebraic subsets $M_i := M \cap Z_i(R)$, $i = 1, \ldots, r$. Denoting by Z_i' the Zariski closure of M_i in X we have $Z_i' \subset Z_i$ and

$$Z_1' \cup \ldots \cup Z_r' = Z_1 \cup \ldots \cup Z_r,$$

and we conclude that $Z_i' = Z_i$ for $i = 1, \ldots, r$. This means that every M_i is Zariski dense in Z_i. Since Z_i is not contained in the union of the Z_j with $j \neq i$, also M_i is not contained in the union of the M_j with $j \neq i$. Thus

$$M_i' := M \setminus \bigcup_{j \neq i} M_j$$

is a non empty open subset of M, which is therefore pure of dimension n-1. This implies dim M_i = n-1 and dim Z_i = n-1 for every i = 1,...,n. The set $Z_i(R)$ contains M_i, hence has again dimension n-1. We now conclude from Theorem 3.4 that for every i = 1, ...,r the prime divisor Z_i is indefinite. We introduce the purely indefinite divisor

$$D := Z_1 + \ldots + Z_r.$$

By construction |D| is the Zariski closure Z of M. Since M is pure and full, M coincides with $\Sigma_{n-1}(|D|_R)$. By Theorem 3.4 this last set is $|D|_t$. It is now also clear that D is the only purely indefinite divisor with $|D|_t$ = M, since by Proposition 3.6 for any such divisor D' the variety |D'| is the Zariski closure of M in X.

Q.e.d.

A mild generalization of these results is possible. Assume only that X is an irreducible n-dimensional variety which is normal at every real point, and that X(R) has dimension n. Let X' denote the open subvariety of all regular points of X. Then $X(R) \setminus X'(R)$ has dimension at most n-2. In particular X'(R) is not empty. Let D be an effective divisor on X and let D' denote the restriction of D to X'.

Definition 3.10. We call D indefinite (resp. semidefinite, resp. purely indefinite) if D' is indefinite (resp. semidefinite, resp. purely indefinite). We denote by $|D|_t$ the closure of the semialgebraic set $|D'|_t$ in X(R).

It is evident that all the theorems, propositions and corollaries in this section, except Corollary 3.5, remain true word by word in the present more general situation. Corollary 3.5 remains true for a normal irreducible variety V over k instead of a regular variety.

§ 4 A remark on semidefinite prime divisors

As before let X be an irreducible n-dimensional variety over R such that X(R) is also n-dimensional and contains only normal points. We regard on X(R) beside the strong topology also the coarser Zariski topology. This is the topology on X(R) induced by the Zariski topology of X. Every Zariski closed subset M of X(R) is a finite union of irreducible closed subsets $M_1,\ldots,M_r$ with $M_i \not\subset M_j$ for $i \neq j$. We call these subsets M_i the irreducible components of M. They are uniquely determined by M.

Every irreducible Zariski closed subset M of X(R) which has dimension n-1 is clearly the set of real points of an indefinite prime divisor D on X uniquely determined by M (cf. Theorem 3.9, which says much more than this.) We now prove a weak analogue of this statement for lower dimensional irreducible Zariski closed subsets of X(R). Uniqueness of the prime divisor D can no longer be expected. Thus the following theorem is less valuable than Theorem 3.9.

Theorem 4.1. Suppose that X is also quasiprojective, i.e. a locally closed subscheme of some projective space $\mathbb{P}^N_R$. Let M be an irreducible Zariski closed subset of X(R) of dimension at most n-2. Then there exists some semidefinite prime divisor D on X such that M = D(R).

For the proof we replace X by its normalization, which does not change anything for the space X(R). Now the zero divisor $\mathrm{div}(f)_+$ and the pole divisor $\mathrm{div}(f)_-$ of any non zero rational function f on X are honestly defined as Weil divisors.

The set X(R) is contained in the affine open subscheme V of $\mathbb{P}^N_R$ which is the complement of the hypersurface $x_o^2 + \ldots + x_N^2 = 0$. We introduce the Zariski closure X_1 of $X \cap V$ in V. Then $X(R) = X_1(R)$ and X_1 is an affine variety. Let W denote the Zariski closure of M in X_1. We choose regular functions $g_1,\ldots,g_r$ on X_1 such that W is the reduced subscheme $N_{X_1}(g_1) \cap \ldots \cap N_{X_1}(g_r)$ of all common zeros of $g_1,\ldots,g_r$ on X_1. For the regular function

$$g := g_1^2 + \ldots + g_r^2$$

on X_1 we have

$$M = \{x \in X_1(R) \mid g(x) = 0\}.$$

We now extend the regular function g : $X \cap V$ to a rational function f on X in the unique possible way. The domain of definition of f contains $X \cap V$, hence X(R). Thus the pole divisor $E := \mathrm{div}(f)_-$ has in its support no real points, i.e. E is definite. On the other hand we have for the zero divisor $D := \mathrm{div}(f)_+$

$$|D|_R = \{x \in X(R) \mid f(x) = 0\} = \{x \in X_1(R) \mid g(x) = 0\} = M.$$

Let $D = e_1D_1 + \ldots + e_sD_s$ be the decomposition of D into prime divisors. M is the union of the Zariski closed subsets $D_1(R),\ldots,D_s(R)$. Since M is irreducible, M coincides with one of these sets, say $M = D_1(R)$. The prime divisor D_1 is semidefinite according to Theorem 3.4, or already Proposition 2.3, and our theorem is proved.

§ 5 Extremal positive semidefinite forms and extremal squares

Let X be the (n-1)-dimensional projective space $\mathbb{P}_R^{n-1}$ ($n \geq 2$). Every effective divisor D on X is the divisor div(F) of a form $F(x_1,\ldots,x_n)$ with coefficients in R uniquely determined by D up to a multiplicative constant. In this way the prime divisors correspond with the irreducible forms, the indefinite divisors correspond with the indefinite forms in the usual sense - notice that X(R) is connected -, and the semidefinite (resp. definite) divisors correspond with the positive semidefinite (resp. definite) forms, of course also with the negative semidefinite (resp. definite) forms.

We call a form $F \in R[x_1,\ldots,x_n]$ purely indefinite, if the divisor div(F) is purely indefinite. This means that F is not constant, all irreducible factors of F are indefinite, and no irreducible factors occur with multiplicity > 1.

For any integral number $r \geq 0$ we denote by $F(r)$ the set all non zero forms of degree r in $R[x_1,\ldots,x_n]$ and by F the union of all $F(r)$. For any even number $d \geq 0$ we denote by $P(d)$ the convex cone in $F(d)$ consisting of all psd (= postive semidefinite) forms of degree d in $R[x_1,\ldots,x_n]$, and by P the union of all $P(d)$. Similarly we denote by $\Sigma(d)$ the convex subcone of $P(d)$ consisting of all finite sums of squares of non zero forms in $R[x_1,\ldots,x_n]$ of degree $\frac{d}{2}$, and by Σ the union of the sets $\Sigma(d)$.

The cones $P(d) \cup \{0\}$ and $\Sigma(d) \cup \{0\}$ are well known to be closed semialgebraic subsets of the vector space $F(d) \cup \{0\}$. Our theory in §2 has some applications to the theory of the sets $E(P(d))$ and $E(\Sigma(d))$ of extremal points of the cones $P(d)$ and $\Sigma(d)$. We refer the reader to the paper [CL] for the background, some results, and concrete examples in this theory. Let again $E(P)$ denote the union of sets $E(P(d))$ and $E(\Sigma)$ the union of the sets $E(\Sigma(d))$.

If nothing else is said all forms in the sequel are understood to be forms in $x_1,\ldots,x_n$ over R. For any two such forms we mean by "$F \geq G$" that $F - G$ lies in $P \cup \{0\}$. In particular then F and G must have the same degree. Similarly we mean by "$F >> G$" that $F - G$ lies in $\Sigma \cup \{0\}$. Clearly an element F of P lies in $E(P)$ if and only if $F \geq G \geq 0$ implies $G = \lambda F$ with some constant λ. Similarly an element F of Σ lies in $E(\Sigma)$ if and only if $F >> G >> 0$ implies $G = \lambda F$ with some constant λ. Of course in both cases the constant λ lies in the interval $[0,1]$.

Theorem 5.1. i) Let F and G be psd forms. Assume that $F \in E(P)$ and G divides F. Then $G \in E(P)$.

ii) Assume that $F \in E(\Sigma)$ and $F = G \cdot H^2$ with some forms G and H. Then $G \in E(\Sigma)$.

iii) Let G be a psd form and H a purely indefinite form. Then G lies in E(P) if and only if GH^2 lies in E(P).

iv) Let again G be a psd form and H a purely indefinite form. Then G lies in $E(\Sigma)$ if and only if GH^2 lies in $E(\Sigma)$.

Proof. i) We have $F = GH$ with some psd form H. Suppose that $G \geq G' \geq 0$. We have to verify that $G' = \lambda G$ with some constant λ. Since $H \geq 0$ we have $GH \geq G'H \geq 0$. Since F is extremal this implies $G'H = \lambda GH$ with some constant λ and then $G' = \lambda G$.

ii) We may induct on the number of irreducible factors of H and thus assume that H is irreducible. Since F is an extremal sum of squares F is actually a square L^2. Now H divides L. We have $L = HS$ with some form S and then $F = H^2S^2$. From this we obtain $G = S^2$. In particular $G \in \Sigma$. We see now by the same argument as in i) that G is extremal in Σ.

iii) If GH^2 is extremal then also G is extremal as has been proved above. Assume now that G is extremal. It suffices to consider the case that H is indefinite and irreducible, since we then obtain the full result by iteration. Let L be a non zero form with $GH^2 \geq L \geq 0$. The set of real zeros $Z(H)$ is contained in $Z(L)$. By a mild application of Corol. 2.5 we see that H divides L. (Restrict H and L to the n-standard open affine subvarieties of $\mathbb{P}_R^{n-1}$.) Since H is indefinite then also H^2 divides L, cf. Proposition 3.2. We have $L = H^2L'$ with some psd form L' and obtain from $GH^2 \geq L'H^2 \geq 0$ that $G \geq L' \geq 0$. Since G is extremal this implies $L' = \lambda G$ with some constant λ and then $L = \lambda GH^2$.

iv) We again retreat to the case that H is irreducible and indefinite. If GH^2 lies in $E(\Sigma)$ then by ii) also G lies in $E(\Sigma)$. Assume now that $G \in E(\Sigma)$. Suppose that $GH^2 \gg L \gg 0$. We have

$$L = M_1^2 + \ldots + M_r^2$$

with some forms $M_1, \ldots, M_r$ of same degree. The set $Z(H)$ is contained in every zero set $Z(M_i)$. Thus by Corollary 2.5 we have $M_i = HN_i$ with some forms N_i and $L = H^2L_1$, where

$$L_1 = N_1^2 + \ldots + N_r^2 \in \Sigma.$$

We can apply the same argument to the sum of squares $GH^2 - L$ and have $GH^2 - L = H^2 S_1$ with some $S_1 \in \Sigma$. We obtain $G = L_1 + S_1$. Since G is extremal in Σ this implies $L_1 = \lambda G$ with some constant $\lambda \in [0,1]$ and then $L = \lambda G H^2$. Thus GH^2 is indeed extremal in Σ. Theorem 5.1 is now completely proved.

We may ask for which forms F the square F^2 is extremal in Σ or even in P. By part iii) of Theorem 5.1 the latter is true for any product F of irreducible indefinite forms. We also know from parts i) and ii) of the theorem that

$$(F_1 F_2)^2 \in E(\Sigma) \Rightarrow F_1^2 \in E(\Sigma),\ F_2^2 \in E(\Sigma);$$

$$(F_1 F_2)^2 \in E(P) \Rightarrow F_1^2 \in E(P),\ F_2^2 \in E(P).$$

To pursue this question further we may omit in a given form F all irreducible indefinite factors, according to Theorem 5.1, and assume that F is psd. We have the following partial result.

Theorem 5.2. Let F be a form in $E(P)$. Then F^2 has the following property: If $F^2 = G^2 + H$ with some psd form H and some form G then $G^2 = \varepsilon F^2$ with some constant ε. (Of course ε lies in the interval $[0,1]$.) In particular $F^2 \in E(\Sigma)$.

Proof. We may assume that $F \neq \pm G$. We distinguish two cases.

Case 1: $F - G$ is semidefinite. If $F - G$ would be negative semidefinite then also $F + G$ would be negative semidefinite, since $F^2 - G^2 \geq 0$. Thus the sum $2F$ of $F - G$ and $F + G$ would be negative semidefinite, which is not true. Thus $F - G \geq 0$. Since $F^2 - G^2 = (F - G)(F + G)$ is psd, also $F + G \geq 0$. From the relation

$$F = (F + G)/2 + (F - G)/2$$

we obtain, since F is extremal,

$$(F - G)/2 = \lambda F,\ (F + G)/2 = \mu F$$

with constants $\lambda > 0$, $\mu > 0$ such that $\lambda + \mu = 1$. This implies $G = (\mu - \lambda)F$ and then $G^2 = (\mu - \lambda)^2 F^2$, as desired.

Case 2. $F - G$ is indefinite. According to Proposition 3.2 there exists an irreducible indefinite form P which divides $F - G$ with an odd multiplicity. Since $F^2 - G^2 \geq 0$ the form P occurs in $F^2 - G^2$ with even multiplicity, again by Propostion 3.2. Thus P divides also $F + G$, hence P divides both F and G. Since F is psd even P^2 divides F. We have $F = P^2 F_1$ with a form $F_1 \in E(P)$ by Theorem 5.1.i. We also have $G = P G'$ with some form G' and the equation

$$P^4 F_1^2 = P^2 G'^2 + H.$$

Thus $H = P^2 H'$ with a form $H' \in P$, and

$$P^2 F_1^2 = G'^2 + H'.$$

The zero set $Z(P)$ is contained in $Z(G')$ and also in $Z(H')$. Thus by §2 the irreducible indefinite from P divides both G' and H', the latter one with an even multiplicity. We obtain $G' = P G_1$, $H' = P^2 H_1$ with $H_1 \in P$, and

$$F_1^2 = G_1^2 + H_1.$$

The proof can now be completed by induction on the degree of F, since F_1 has smaller degree than F.

Q.e.d.

Remark. In all these considerations we could have replaced our projective space $\mathbb{P}_R^{n-1}$ by a product $\mathbb{P}_R^{n_1} \times \ldots \times \mathbb{P}_R^{n_r}$, i.e. work with multiforms instead of forms. Thus Theorems 5.1 and 5.2 remain true for multiforms instead of forms.

§ 6 Comparison of the sets $EP(n,d)$ and $E\Sigma(n,d)$.

Looking again for forms F such that F^2 is extremal in Σ or even in P it is natural to ask whether every $F^2 \in E(\Sigma)$ actually lies in $E(P)$. In case of a positive answer we would know from Theorems 5.1 and 5.2 for any psd form F that F^2 lies in $E(\Sigma)$ if and only if F lies in $E(P)$, and the relation between the sets $E(\Sigma)$ and $E(P)$ would be well understood.

Unfortunately things turn out to be not that simple. Let us write more precisely $P(n,d)$ instead of $P(d)$ and $\Sigma(n,d)$ instead of $\Sigma(d)$ to indicate the number n of variables of the forms under consideration. We ask for which pairs (n,d) with $n \geq 2$, $d \geq 2$ and even, the set $E\Sigma(n,d)$ of extremal points of the cone $\Sigma(n,d)$ is contained in the set $EP(n,d)$ of extremal points of the cone $P(n,d)$. The following theorem gives a complete answer to this question.

Theorem 6.1. Let $n \geq 2$ be a natural number and d be an even natural number. Then $E\Sigma(n,d) \subset EP(n,d)$ precisely in the following cases.

i) $n = 2$; ii) $d \leq 6$; iii) $(n,d) = (3,8)$; iv) $(n,d) = (3,10)$.

Thus the question, whether $E\Sigma(n,d)$ is contained in $EP(n,d)$ is answered by the following chart:

n \ d	2	4	6	8	10	12	14
2	✓	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	x	x
4	✓	✓	✓	x	x	x	x
5	✓	✓	✓	x	x	x	x
6	✓	✓	✓	x	x	x	x

Legend: ✓ = positive answer
x = negative answer

The rest of the section is devoted to a proof of this theorem. If $n = 2$ or $d = 2$ then $\Sigma(n,d) = P(n,d)$ and there is nothing to be proved. Thus we assume henceforth that $n \geq 3$ and $d \geq 4$.

Consider now the case that $d = 4$ or $d = 6$. Let F be a form with

$F^2 \in E\Sigma(n,d)$. Suppose that F^2 does not lie in $EP(n,d)$. Cancelling out in F all indefinite irreducible factors we obtain a form with the same properties, as follows from Theorem 5.1. Thus we may assume that F has only psd factors. Then F cannot have degree 3. Thus F is a psd quadratic form. After a linear change of coordinates we have

$$F = x_1^2 + \ldots + x_r^2$$

with $1 < r \leq n$. Now

$$F^2 = x_1^4 + 2x_1^2(x_2^2 + \ldots + x_r^2) + (x_2^2 + \ldots + x_r^2)^2.$$

We see that F^2 is not extremal in $\Sigma(n,4)$. This contradiction proves that $E\Sigma(n,d)$ is contained in $EP(n,d)$ for $d \leq 6$.

Suppose now that F is a form of degree 4 in n variables such that F^2 lies in $E\Sigma$ but not in EP. If F would contain an indefinite irreducible factor then taking out this factor we would obtain a form G with $G^2 \in E\Sigma(n,d)$ but $G^2 \notin EP(n,d)$ for some $d \leq 6$ (Theorem 5.1). This has been proved to be impossible. Thus F does not contain an indefinite factor and we may assume in particular that F is psd. If F would be reducible then $F = Q_1Q_2$ with psd quadratic forms Q_1 and Q_2. But then also the factors Q_1^2 and Q_2^2 of $Q_1^2Q_2^2$ would lie in $E\Sigma$ (Theorem 5.1), which means that Q_1 and Q_2 would be squares of linear forms. This contradicts the fact that F has no indefinite factors. Thus F must be an irreducible positive semidefinite quartic.

It is known since Hilbert that $P(3,4) = \Sigma(3,4)$, cf. [CL, §6] for an elementary proof in the case $R = \mathbb{R}$ *). Thus in the case $n = 3$ our form F has to be a sum of squares, but not a square, and we obtain as above a contradiction to the assumption that F^2 is extremal in $\Sigma(3,8)$. We have proved that $E\Sigma(3,8)$ is contained in $EP(3,8)$.

Assume now that F is a form in 3 variables of degree 5 such that F^2 is extremal in $\Sigma(3,10)$. F contains an irreducible factor H of odd degree, $F = HG$. By Theorem 5.1 the form G^2 is extremal in Σ. Since $\deg G^2 \leq 8$ we know that G^2 is extremal in P. Thus, again by Theorem 5.1, the form F^2 is extremal in P. We have proved that $E\Sigma(3,10)$ is contained in $EP(3,10)$.

---

*) This proof works equally well over all real closed fields R, taking into account the rudiments of [DK, §9]. No appeal to Tarski's principle is necessary.

We now have verified all the affirmative answers in the chart above. To get all negative answers it suffices to check that $E\Sigma(3,12)$ is not contained in $EP(3,12)$ and $E\Sigma(4,8)$ is not contained in $EP(4,8)$. Indeed, regarding a form F in the variables $x_1,\ldots,x_n$ also as a form in the variables $x_1,\ldots,x_{n+1}$, it is an easy exercise to prove that

$$F^2 \in E\Sigma(n,d) \Rightarrow F^2 \in E\Sigma(n+1,d),$$

and it is trivial that

$$F^2 \notin EP(n,d) \Rightarrow F^2 \notin EP(n+1,d).$$

Furthermore choosing some linear form L in the variables $x_1,\ldots,x_n$, it is evident from Theorem 5.1 that

$$F^2 \in E\Sigma(n,d) \Rightarrow F^2L^2 \in E\Sigma(n,d+2)$$

and

$$F^2 \notin EP(n,d) \Rightarrow F^2L^2 \notin EP(n,d+2).$$

We shall now exhibit a form in $E\Sigma(3,12)$ which is not extremal in $P(3,12)$. Fortunately a counterexample for $(n,d) = (4,8)$ can be constructed by similar principles. Thus it will be sufficient to devote our main efforts to the case $(n,d) = (3,12)$.

We start with the ternary sextic

$$S(x,y,z) = x^4y^2 + y^4z^2 + z^4x^2 - 3x^2y^2z^2$$

in [CL]. This form has seven zeros: (1,0,0), (0,1,0), (0,0,1), (1,1,1), (-1,1,1), (1,-1,1) and (1,1,-1). We shall look at an auxiliary form

$$T(x,y,z) = (x^2y + y^2z - z^2x - xyz)^2$$

which is chosen in such a way that it vanishes on all zeros of S, except (-1,1,1).

Theorem 6.2. Let $f(x,y,z) = S(x,y,z) + T(x,y,z)$. Then $p := f^2$ lies in $E\Sigma(3,12)$ but not in $EP(3,12)$.

The fact that p is not extremal in $P(3,12)$ will be deduced from an easy lemma (Lemma 1), and follows by the way also from Theorem 5.1. i, while the fact that p is extremal in $\Sigma(3,12)$ will be deduced from a difficult lemma (Lemma 2).

Lemma 1. The forms S^2, ST, T^2 are linearly independent over R.

Proof. Suppose $\alpha S^2 + \beta ST + \gamma T^2 = 0$, where $\alpha,\beta,\gamma \in R$. Evaluating at $(-1,1,1) \in Z(S) \setminus Z(T)$, we get $\gamma = 0$. Dividing by S, we get $\alpha S + \beta T = 0$, so clearly $\alpha = \beta = 0$.

Q.e.d.

Since $p = f^2 = S^2 + 2ST + T^2$, this lemma clearly implies that p cannot be extremal in $P(3,12)$. It remains to be shown that p is extremal in $\Sigma(3,12)$.

Lemma 2. Let f be as in the theorem. If $f^2 = h_1^2 + \ldots + h_r^2$ in $R[x,y,z]$ then each h_i is an R-linear combination of S and T.

Using this lemma we can show that $p = f^2$ is extremal in $\Sigma(3,12)$ as follows. If $f^2 = h_1^2 + \ldots + h_r^2$, we write $h_i = a_i S + b_i T$ with $a_i, b_i \in R$. Then

$$f^2 = S^2 + 2ST + T^2 = (\sum_1^r a_i^2)\, S^2 + 2(\sum_1^r a_i b_i)\, ST + (\sum_1^r b_i)\, T^2,$$

so by Lemma 1,

$$\sum_1^r a_i^2 = \sum_1^r b_i^2 = \sum_1^r a_i b_i = 1.$$

This implies that $a_i = b_i$ for $1 \le i \le r$, so $h_i^2 = a_i^2 (S+T)^2 = a_i^2 p$, as desired.

Our job is now to prove Lemma 2. For this we need a third lemma which is true for arbitrary polynomials instead of just ternary forms.

Lemma 3. Suppose $f \in R[x_1,\ldots,x_n]$ is positive semidefinite and $f^2 = h_1^2 + \ldots + h_r^2$ with polynomials $h_i \in R[x_1,\ldots,x_n]$. Let $a \in R^n$ be a zero of f. Then a is also a zero of h_i and of every partial derivative $\partial h_i / \partial x_j$ $(1 \le i \le r,\ 1 \le j \le n)$.

Proof. Since f is psd clearly a is a zero of every $\partial f / \partial x_j$, $1 \le j \le n$. Computing the partial derivatives of f^2, we have

$$\frac{\partial}{\partial x_j} f^2 = 2f \frac{\partial f}{\partial x_j},$$

$$\frac{\partial^2}{\partial x_j \partial x_k} f^2 = 2f \frac{\partial^2 f}{\partial x_j \partial x_k} + 2 \frac{\partial f}{\partial x_j} \frac{\partial f}{\partial x_k},$$

so these partial derivatives all vanish at a. (In fact even the third order partial derivatives of f^2 vanish at a. We do not need this in the following.) From $0 = h_1(a)^2 + \ldots + h_r(a)^2$ we have of course $h_1(a) = \ldots = h_r(a) = 0$. Computing $(\partial^2/\partial x_j^2)(f^2)$ from the expression $f^2 = h_1^2 + \ldots + h_r^2$, we get

$$0 = \sum_{i=1}^{r} [2h_i(a) \frac{\partial^2 h_i}{\partial x_j}(a) + 2(\frac{\partial h_i}{\partial x_j}(a))^2] = 2 \sum_{i=1}^{r} \frac{\partial h_i}{\partial x_j}(a)^2,$$

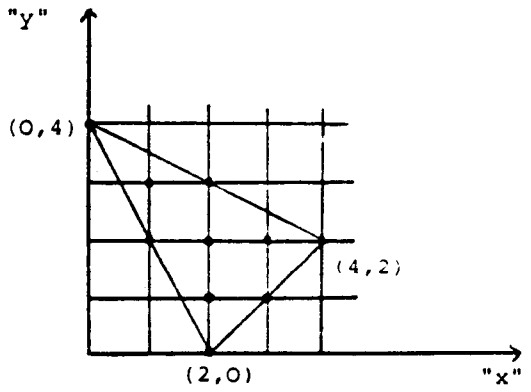
so $\frac{\partial h_i}{\partial x_j}(a) = 0$ for all i,j.

Q.e.d.

We now enter the proof of Lemma 2. Thus $f = S + T$, and a decomposition $f^2 = h_1^2 + \ldots + h_r^2$ with forms $h_i \in R[x,y,z]$ of degree 6 is given. Let h be any of the forms h_i. The first step in the proof is to determine which are the sextic monomials which may occur in h. This can be done by inspection - but it is easier to invoke the general method of "cages", cf. [R].*) Denoting the cage of a form g by C(g) we have by the latter method

$$C(h) \subset \tfrac{1}{2}C(f^2) = C(f),$$

and C(f) contains the lattice points (4,2,0), (0,4,2), (2,0,4), (2,2,2), (3,2,1), (3,1,2), (2,3,1), (2,1,3), (1,2,3), (1,3,2). If we represent the points of C(f) by their first two coordinates, we have the following picture of a "projection" of C(f).



(Actually all lattice points of C(f) occur as monomials in f.) Thus we may express the sextic form h in the following way:

*) A more detailed account of this method will be given in [CLR].

$$h(x,y,z) = ax^4y^2 + by^4z^2 + cx^2z^4 + dx^2y^2z^2 + ex^3y^2z + gx^3yz^2 + ix^2y^3z + jx^2yz^3 + kxy^3z^2 + lxy^2z^3.$$

By Lemma 3 the partial derivatives $\partial h/\partial x$, $\partial h/\partial y$, $\partial h/\partial z$ must vanish at the points (1,1,1), (1,1,-1) and (1,-1,1) of $Z(f)$. This leads to the following system of nine linear homogeneous equations in the ten "unknowns" a,b,...,k,l.

$$\begin{cases}
(1) & 4a & +2c+2d+3e+3g+2i+2j+ k+ l=0 & (\frac{\partial}{\partial x} \text{ at } (1,1,1)) \\
(2) & 4a & +2c+2d-3e+3g-2i-2j+ k- l=0 & (.....(1,1,-1)) \\
(3) & 4a & +2c+2d+3e-3g-2i-2j- k+ l=0 & (.....(1,-1,1)) \\
(4) & 2a+4b & +2d+2e+ g+3i+ j+3k+2l=0 & (\frac{\partial}{\partial y} \text{ at } (1,1,1)) \\
(5) & 2a+4b & +2d-2e+ g-3i- j+3k-2l=0 & (.....(1,1,-1)) \\
(6) & 2a+4b & +2d+2e- g-3i- j-3k+2l=0 & (.....(1,-1,1)) \\
(7) & & 2b+4c+2d+ e+2g+ i+3j+2k+3l=0 & (\frac{\partial}{\partial z} \text{ at } (1,1,1)) \\
(8) & & 2b+4c+2d- e+2g- i-3j+2k-3l=0 & (.....(1,1,-1)) \\
(9) & & 2b+4c+2d+ e-2g- i-3j-2k+3l=0 & (.....(1,-1,1))
\end{cases}$$

By explicit computation we shall show that this linear system of equations has a solution space of dimension 2 (with a basis corresponding, of course, to S and T). We proceed as follows:

$$\begin{cases}
(1') = \frac{(1)-(2)}{2}: & 3e + 2i + 2j + l = 0 \\
(2') = \frac{(1)-(3)}{2}: & 3g + 2i + 2j + k = 0 \\
(3') = \frac{(1)+(2)}{2}: & 4a + 2c + 2d + 3g + k = 0 \\
(4') = \frac{(4)-(5)}{2}: & 2e + 3i + j + 2l = 0 \\
(5') = \frac{(4)-(6)}{2}: & g + 3i + j + 3k = 0 \\
(6') = \frac{(4)+(5)}{2}: & 2a + 4b + 2d + g + 3k = 0 \\
(7') = \frac{(7)-(8)}{2}: & e + i + 3j + 3l = 0 \\
(8') = \frac{(7)-(9)}{2}: & 2g + i + 3j + 2k = 0 \\
(9') = \frac{(7)+(8)}{2}: & 2b + 4c + 2d + 2g + 2k = 0
\end{cases}$$

Note that $\frac{(1')+(4')+(7')}{6}$ gives (1") $e + i + j + l = 0$

$\frac{(2')+(5')+(8')}{6}$ gives (2") $g + i + j + k = 0$

From (1"), (4') and (7'), we get $i = j = -e = -l$.
From (2"), (5') and (8'), we get $i = j = -g = -k$.

Eliminating g from (3'), (6') and (9') and dividing by 2, we get

$$\begin{aligned} &(3") \quad 2a + c + d + 2k = 0,\\ &(6") \quad a + 2b + d + 2k = 0,\\ &(9") \quad b + 2c + d + 2k = 0, \end{aligned}$$

which leads easily to $a = b = c$ and $d = -3a - 2k$. Thus, a and k are the free parameters, and the solution space to our linear system of equations has dimension 2. Since S and T do give rise to independent solutions in the solution space, we can conclude that $h = \alpha S + \beta T$ $(\alpha,\beta \in R)$. More explicitely, the general solution to the linear system is given by

$$\begin{aligned} (a,b,c,d,e,g,i,j,k,l) &= (a,a,a,-3a-2k,k,k,-k,-k,k,k)\\ &= a(1,1,1,-3,0,\ldots,0) + k(0,0,0,-2,1,1,-1,-1,1,1)\\ &= (a+\tfrac{k}{2})(1,1,1,-3,0,\ldots,0) - \tfrac{k}{2}(1,1,1,1,-2,-2,2,2,-2,-2) \end{aligned}$$

So we are finished by noting that $(1,1,1,-3,0,\ldots,0)$ corresponds to S and $(1,1,1,1,-2,-2,2,2,-2,-2)$ corresponds to T. We now have proved Lemma 2 and Theorem 6.2.

The counterexample needed to show that $E\Sigma(4,8)$ is not contained in $EP(4,8)$ is entirely analogous. We use $p := (Q+U)^2$ where

$$Q(w,x,y,z) = w^4 + x^2y^2 + y^2z^2 + z^2x^2 - 4xyzw,$$

$$U(w,x,y,z) = (w^2 + xy - yz - zx)^2.$$

The form Q has seven zeros: $(0,1,0,0)$, $(0,0,1,0)$, $(0,0,0,1)$, $(1,1,1,1)$, $(1,1,-1,-1)$, $(1,-1,1,-1)$, $(1,-1,-1,1)$, all of which are zeros of U except the last one. By a cage consideration similar to the one used before we can see that, if $p = h_1^2 + \ldots + h_r^2$, then any of the h_i's has the form

$$\begin{aligned} h(w,x,y,z) = {} & aw^4 + bx^2y^2 + cy^2z^2 + dz^2x^2 + exyzw\\ & + gw^2xy + iw^2yz + jw^2zx\\ & + kx^2xy + lx^2yz + my^2zx, \end{aligned}$$

with eleven possible terms. By Lemma 3 the four first partial derivatives of h must vanish on $(1,1,1,1)$, $(1,1,-1,-1)$ and $(1,-1,1,-1)$. This gives us 12 linear homogeneous equations in the 11 unknowns

a,b,...,l,m. A calcu
solution space has d
corresponding to Q a

There remains on
of ideas of this pap

Question. For which
that $F^2 \notin EP(n,2d)$?

Notice that by
question is for a "

a,b,...,l,m. A calculation similar to the one we did shows that the solution space has dimension 2, hence is spanned by the 11-tuples corresponding to Q and U.

There remains one problem open which fits naturally into the circle of ideas of this paper:

Question. For which (n,d) does there exist a form $F \in EP(n,d)$ such that $F^2 \notin EP(n,2d)$?

Notice that by Theorem 5.2 the form F^2 lies in $E\Sigma(n,2d)$. Thus the question is for a "stronger" counterexample to the inclusion $E\Sigma \subset EP$.

References

[A] E. Artin, Über die Zerlegung definiter Funktionen in Quadrate, Abh. Math. Seminar, Universität Hamburg 5, 100-115 (1927).

[CL] M.D. Choi, T.Y. Lam, Extremal positive semidefinite forms, Math. Ann. 231, 1-18 (1977).

[CL_1] M.D. Choi, T.Y. Lam, An old question of Hilbert, Proceedings Quadratic Forms Conference 1976 (ed. G. Orzech), Queen's Papers in Pure and Appl. Math. 46, 385-405.

[DK] H. Delfs, M. Knebusch, Semialgebraic topology over a real closed field II: Basic theory of semialgebraic spaces, Math. Z. 178, 175-213 (1981).

[DK_1] H. Delfs, M. Knebusch, On the homology of algebraic varieties over real closed fields, to appear, preprint Univ. Regensburg.

[DE] D.W. Dubois, G. Efroymson, Algebraic theory of real varieties I, Studies and Essays presented to Yu-Why Chen on his sixtieth birthday (1970), 107-135.

[E] G. Efroymson, Henselian fields and solid k-varieties II, Proc. Amer. Math. Soc. 35, 362-366 (1972).

[ELW] R. Elman, T.Y. Lam, A. Wadsworth, Orderings under field extensions, J. reine angew. Math. 306, 7-27 (1979).

[R] B. Reznick, Extremal psd forms with few terms, Duke Math. J. 45, 363-374 (1978).

[R_1] J.J. Risler, Une caractérisation des idéaux des variétés algébriques réelles, C.R. Acad. Sc. Paris 271, 1171-1173 (1970).

[R_{-1}] J.J. Risler, Le théorème des zéros en géometrie algébrique et analytique réelles, Bull. Soc. math. France 104, 113-127 (1976).

[S] C. Stengle, A Nullstellensatz and a Positivstellensatz in semialgebraic geometry, Math. Ann. 207, 87-97 (1974).

[CLR] M.D. Choi, T.Y. Lam, B. Reznick, A combinatorial theory for sums of squares of polynomials, in preparation.

Man-Duen Choi
Department of Mathematics, University of Toronto,
Toronto, M5S 1A1, Canada.

Manfred Knebusch
Fakultät für Mathematik der Universität,
D-8400 Regensburg, Universitätsstr.31, F.R.G.

Tsit-Yuen Lam
Department of Mathematics, University of California,
Berkeley CA 94720, U.S.A.

Bruce Reznick
Department of Mathematics, University of Illinois,
Urbana, Ill. 61801, U.S.A.

Zur Theori

Über

Hans I

Bei unserem
reell abgeschlos
benutzen wir bei
dem zweiten Auto
Menge X(R) der r
R. Durch diesen
und erzielten Re
Formen (z.B. der
Formen) und Jacc
Geyer für abelsc
disch gesehen sc
gehören. Semialc
ten, nämlich de
reellen Punkte

Nun lassen
hängig von den
kann man, wie s
durch Übertrage
Lemma 9.3 im Be
ser Note ist es
zur Theorie der
schen Standard-
braische Geomet
von der Arbeit
diejenigen Sätz
nicht auf alge
§1-§5 geleiste