TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 352, Number 3, Pages 1039–1069
S 0002-9947(99)02522-2
Article electronically published on September 8, 1999

SUMS OF SQUARES OF REGULAR FUNCTIONS ON REAL ALGEBRAIC VARIETIES

CLAUS SCHEIDERER

Dedicated to Manfred Knebusch on the occasion of his 60th birthday

ABSTRACT. Let V be an affine algebraic variety over $\mathbb{R}$ (or any other real closed field R). We ask when it is true that every positive semidefinite (psd) polynomial function on V is a sum of squares (sos). We show that for $\dim V \geq 3$ the answer is always negative if V has a real point. Also, if V is a smooth non-rational curve all of whose points at infinity are real, the answer is again negative. The same holds if V is a smooth surface with only real divisors at infinity. The "compact" case is harder. We completely settle the case of smooth curves of genus ≤ 1: If such a curve has a complex point at infinity, then every psd function is sos, provided the field R is archimedean. If R is not archimedean, there are counter-examples of genus 1.

INTRODUCTION

The basic problem addressed in this paper goes back to Minkowski and Hilbert. Consider homogeneous polynomials (*alias* forms) $f(x_1, \ldots, x_n)$ with real coefficients which are positive semidefinite (*psd*, for short), in the sense that they take only non-negative values on $\mathbb{R}^n$. The question of whether every such f can be written as a sum of squares (*sos*, for short) of forms seems to have originated with Minkowski. In the oral defense of his doctoral dissertation, held in Königsberg (Prussia) on July 30th, 1885, one of the two theses he proposed was: "Es ist nicht wahrscheinlich, daß eine jede positive Form sich als eine Summe von Formenquadraten darstellen läßt." ([23], p. 202) Hilbert happened to be one of the two officially appointed "opponents" in the disputation. At the end of the discussion he remarked that Minkowski's arguments had convinced him that there should exist ternary psd forms which are not sos. Working these ideas out, Hilbert proved in 1888 that for any $n \geq 3$ and any even $d \geq 4$ there exists a psd n-ary form of degree d which is not sos, the only exception being the case $(n, d) = (3, 4)$ [15]. Hilbert later acknowledged that Minkowski's proposition had been the original motivation for him to study the question of representing psd forms as sums of squares. (See [18] for this and for the detail mentioned before.)

Having proved this negative result, Hilbert turned to the question of whether each psd form might be the quotient of two sums of squares of forms. He was able

Received by the editors October 5, 1997.
1991 *Mathematics Subject Classification.* Primary 14P99; Secondary 11E25, 12D15, 13H05, 14G30, 14H99, 14J99.
Key words and phrases. Sums of squares, positive semidefinite functions, preorders, real algebraic curves, Jacobians, real algebraic surfaces, real spectrum.

to prove this for ternary forms [16], but couldn't decide the general case. As is well known, he included the question as number 17 in his famous list of unsolved problems [17]. This problem was later solved by E. Artin in the affirmative [1].

For the purpose of this paper, it is preferable to use a non-homogeneous setting. Thus, the original question of Minkowski and Hilbert asked whether there are psd polynomials over $\mathbb{R}$ which are not sums of squares of polynomials. There is an extensive literature on various aspects of this question and of Hilbert's 17th problem; see, e.g., the surveys by Gondard [12], Reznick [27], Powers [26] and Delzell [11], as well as references given there. But although the same question makes perfect sense for finitely generated $\mathbb{R}$-algebras other than polynomial rings, or even for arbitrary rings (see below), there have been surprisingly few authors dealing with such natural generalizations. (See however the paper [7], in which Choi, Lam, Reznick and Rosenberg study certain integral domains A with quotient field K and ask when $A \cap \Sigma K^2 = \Sigma A^2$ holds, where ΣA^2 denotes the set of sums of squares in A.)

There is a quantitative counterpart to this problem which has found more attention, namely the question of how many squares are needed in sums of squares representations. Here we will not touch on this aspect at all. For a guide to what has been done one may consult [5], [25] and references given there.

Thus, our principal goal will be the study of the following question: Let V be an affine variety over $\mathbb{R}$ and consider regular functions f on V which are psd, i.e., take non-negative values on the set $V(\mathbb{R})$ of real points of V. When is it true that every such f is a sum of squares of regular functions; in short, when do we have "psd = sos on V"?

It will be convenient to place our question in a more general context. If A is any (commutative) ring, let A_+ denote the set of psd elements of A, i.e., those $a \in A$ which are non-negative in every point of $\operatorname{Sper} A$, the real spectrum of A. (See the notations section at the end of this introduction for further explanation.) When is it true that $A_+ = \Sigma A^2$? For V an affine $\mathbb{R}$-variety and $A = \mathbb{R}[V]$, this is our original question.

For connected varieties of dimension ≥ 3 we can give a complete answer: If k is a field and A is a finitely generated connected k-algebra with $\operatorname{Sper} A \neq \emptyset$ and $\dim A \geq 3$, then there always exists a psd element in A which is not sos, Theorem 6.2. (If $\operatorname{Sper} A$ is empty, then every element of A is a sum of squares, at least if $\operatorname{char}(k) \neq 2$.) Essentially, the reason is of local nature; it is the fact that $A_+ \neq \Sigma A^2$ for every regular local domain A of dimension ≥ 3 whose residue field is formally real (Proposition 1.2).

We also study other regular local domains A (always with a formally real quotient field). If $\dim(A) \geq 4$ and the residue field is non-real, we have again $A_+ \neq \Sigma A^2$, at least for A of geometric origin (Proposition 1.5). As for lower dimensions, it is well known that psd = sos for $\dim(A) \leq 1$. The case $\dim(A) = 2$, however, seems much harder to understand. We can show psd = sos only under quite restrictive conditions, but on the other hand we do not know of any counter-example.

Coming back to affine $\mathbb{R}$-varieties, we are left with curves and surfaces. We restrict ourselves to the smooth case here. The above discussion shows that we do not know any local obstructions to psd = sos, so we have to find other ways of reasoning.

The main part of the paper is concerned with curves Y (smooth affine with $Y(\mathbb{R}) \neq \emptyset$). Since it is well known that psd = sos if Y is rational, we assume that Y has genus $g \geq 1$. For the purpose of this introduction, let us assume that Y has

only one (scheme-theoretic) point at infinity, call it ∞. Both the arguments and the results turn out to be basically different depending on whether the point ∞ is complex or real, or equivalently, whether $Y(\mathbb{R})$ is compact or not.

In the non-compact case (Sect. 3), there always exist psd functions f which are not sos, even with $f > 0$ on $Y(\mathbb{R})$ (Theorem 3.2). We actually prove stronger statements, which say that in a suitable sense there are many such functions (Theorems 3.4 and 3.5). On the other hand (Sect. 4), we don't know of any example of a psd, non-sos function f if $Y(\mathbb{R})$ is compact! As a consequence of a theorem of Schmüdgen, it is known that every *strictly* positive $f \in \mathbb{R}[Y]$ is a sum of squares. But it is not clear whether this fact can be used for settling the general psd case. We obtain partial results which allow a reduction from general psd f to more special cases. Essentially, we are trying to reduce to cases where most or all zeros of f are real. For elliptic curves these methods combine nicely to give a general proof that *every psd function is a sum of squares* (Theorem 4.10). After many fruitless attempts at finding a counterexample, this result came as quite a surprise to the author.

It is natural to study all these questions not only for varieties over $\mathbb{R}$, but also over other real closed base fields R. Most results mentioned so far extend to this more general case (although for Theorem 3.5 we only get a weakened version). But this is definitely not true for curves with $Y(R)$ "compact" (if $R \neq \mathbb{R}$, one has to understand compactness in the semi-algebraic sense). Rather we show that if R is non-archimedean, then among elliptic curves Y over R with $Y(R)$ semi-algebraically compact, there are examples for both psd = sos and psd $\neq$ sos (Theorem 4.11).

Finally, consider the case of (smooth affine) surfaces V over R. We try to prove that psd $\neq$ sos here by restricting the problem to suitable curves on V. This is suggested by an extension theorem for psd functions (Theorem 5.6), which says that a psd regular function on a smooth curve can always be extended to a psd regular function on any ambient smooth variety. Combined with the results on curves, this gives the existence of psd, non-sos functions on V as soon as we find suitable curves on V (6.4). However there are many surfaces left for which this technique does not help at all, in particular all surfaces for which $V(R)$ is semi-algebraically compact. It seems not an easy task to decide in a single such case whether psd = sos holds or not!

It should be pointed out that the study of varieties more general than affine space, e.g., curves, can be fruitful for the understanding of psd polynomials as well; see Remarks 6.7 and 6.8.

The results and techniques of this papers have applications to the study of preorders (see 3.3 for a short introduction to this concept). Preorders play a role in real algebraic geometry which is in some sense similar to the role of ideals in ordinary algebraic geometry. There is an interest in the understanding of preorders in particular from the computational point of view. However, preorders are much harder to study than ideals since they tend not to be finitely generated. We present some examples which show how our techniques allow a systematic approach to such questions. In particular, we show that, given a closed semi-algebraic set S, the preorder consisting of all polynomial functions psd on S is infinitely generated if S is 1-dimensional and "sufficiently non-compact" (Theorem 3.5), or if $\dim S \geq 3$ (Theorem 6.1). It seems clear that more can (and should) be done here.

Another main result should be mentioned. Although it is not directly related to the central question of this paper, it turns out to be extremely useful and has some

independent interest: If X is a connected smooth projective curve over a real closed field R, with $X(R) \neq \emptyset$, then every divisor on X is linearly equivalent to a divisor whose support consists only of real points (Theorem 2.7 and its corollaries). So this is a moving lemma which permits one to move the support of a divisor entirely into the real locus. The proof is easy for $R = \mathbb{R}$, but seems less obvious if the field R is not archimedean. We reduce this case to the archimedean case using properties of the real spectrum.

Acknowledgements. I want to acknowledge with thanks some helpful correspondence with Chip Delzell, Manuel Ojanguren and Bruce Reznick on matters related to this paper. In particular, Chip provided me with an amazingly detailed list of comments. My primary debt is however to Vicki Powers. It was her intriguing questions which started me thinking about the problems considered here. Numerous discussions with her helped me to clarify matters. I want to thank her warmly.

Support by the Deutsche Forschungsgemeinschaft (DFG) through the Heisenberg program is gratefully acknowledged.

General notations. Let A be a ring (we always assume that $\frac{1}{2} \in A$). By $\operatorname{Sper} A$ we denote the real spectrum of A, i.e., the topological space consisting of all pairs $\alpha = (\mathfrak{p}, \omega)$ with $\mathfrak{p} \in \operatorname{Spec} A$ and ω an ordering of the residue field $\operatorname{Quot}(A/\mathfrak{p})$ of $\mathfrak{p}$ [9], [3], [19]. The prime ideal $\mathfrak{p}$ is called the support of α, written $\mathfrak{p} = \operatorname{supp}(\alpha)$. A prime ideal is called real if it supports an element of $\operatorname{Sper} A$. Recall that for $f \in A$ the notation "$f(\alpha) \geq 0$" indicates that the residue class $f \bmod \mathfrak{p}$ is non-negative under ω. We put

$$A_+ = \{f \in A\colon\ f(\alpha) \geq 0 \text{ for every } \alpha \in \operatorname{Sper} A\};$$

this is the set of positive semidefinite (psd) elements in A. On the other hand, ΣA^2 denotes the set of sums of squares (*alias* sos elements) in A. By Stengle's Nichtnegativstellensatz (e.g. [19], p. 143), A_+ is the set of all $f \in A$ which satisfy an identity $fs = f^{2n} + t$ with $s,\ t \in \Sigma A^2$ and $n \geq 0$. Our main concern is the question whether the inclusion $\Sigma A^2 \subset A_+$ is an equality, for various rings of geometric nature. If $\operatorname{Sper} A = \emptyset$, then $-1 \in \Sigma A^2$, and thus $\Sigma A^2 = A = A_+$ ([19] III §2).

The concept of preordering of a ring is explained in Sect. 3.3 below. The real spectrum is functorial: A ring homomorphism $\varphi\colon A \to B$ induces a continuous map $\varphi^*\colon \operatorname{Sper} B \to \operatorname{Sper} A$. The map φ^* is characterized by the fact that, for $a \in A$ and $\beta \in \operatorname{Sper} B$, the sign of a at $\varphi^*(\beta)$ is the sign of $\varphi(a)$ at β. Note that $\varphi(A_+) \subset B_+$.

The following fact is well known, but for lack of a handy reference we include its proof:

0.1. Lemma. *If A is a regular noetherian domain, then the set of α in* $\operatorname{Sper} A$ *with* $\operatorname{supp}(\alpha) = (0)$ *is dense in* $\operatorname{Sper} A$.

Proof. Given $\alpha,\ \beta \in \operatorname{Sper} A$, write $\alpha \succ \beta$ iff $\beta \in \overline{\{\alpha\}}$. Assume that A is local, and let $\beta \in \operatorname{Sper} A$ with $\operatorname{supp}(\beta) = \mathfrak{m}$, the maximal ideal. We show by induction on $\dim A$ that there is $\alpha \in \operatorname{Sper} A$ with $\operatorname{supp}(\alpha) = (0)$ and $\alpha \succ \beta$. Let x be a regular parameter of A, and let $\mathfrak{p} = (x)$. Since $A/\mathfrak{p}$ is regular, we find by induction $\gamma \in \operatorname{Sper} A$ with $\operatorname{supp}(\gamma) = \mathfrak{p}$ and $\gamma \succ \beta$. Since $A_{\mathfrak{p}}$ is a discrete valuation ring, there is α with $\operatorname{supp}(\alpha) = (0)$ and $\alpha \succ \gamma$, by the Baer-Krull theorem ([19] II §7). So $\alpha \succ \beta$. □

As a consequence, we have $A_+ = A \cap \Sigma K^2$, where K is the quotient field of A.

Let k be a field. By a k-variety we mean a reduced separated scheme V of finite type over $\operatorname{Spec} k$. If V is a k-variety and $E \supset k$ is an extension, then $V(E)$ is the set of E-valued points of V. If V is irreducible, then $k(V)$ is the function field of V; if V is affine, then $k[V]$ denotes the ring of regular functions on V.

Let K be a field and v a (Krull) valuation on K. If $\alpha \in \operatorname{Sper} K$, then α and v are said to be compatible if $v(a) > 0$ implies $(1+a)(\alpha) > 0$ for every $a \in K^*$. A well known useful fact is the following: If v has a formally real residue field, then $v(a_1^2 + \cdots + a_n^2) = 2 \min_i v(a_i)$ for any $a_1, \ldots, a_n \in K$. We will very often apply this, and occasionally also need the following generalization (whose proof is obvious):

0.2. Lemma. *Let v be a valuation of a field K, and let $a_1, \ldots, a_n \in K^*$. Suppose that K has an ordering which is compatible with v and with respect to which all the a_i have the same sign. Then $v(a_1 + \cdots + a_n) = \min_i v(a_i)$.* □

1. Regular local rings

Let A be a regular local ring ([21], §14) with maximal ideal $\mathfrak{m}$, quotient field K and residue field $\kappa = A/\mathfrak{m}$, and assume $\operatorname{char}(\kappa) \neq 2$. We try to understand when $A_+ = A \cap \Sigma K^2$ is equal to ΣA^2. We will always assume that K is formally real, since otherwise every element of A is sos (Lemma 0.1).

If $\dim A = 0$, then $A = K$ is a field, and the fact that $K_+ = \Sigma K^2$ is classical and due to E. Artin. If $\dim A = 1$, then A is a discrete valuation ring. Again it is known that $A_+ = \Sigma A^2$, even for arbitrary valuation rings A. In this generality it was probably first proved (independently) by Kneser and by Colliot-Thélène; see [7], p. 250 for Kneser's proof.

We will first consider the case $\dim A \geq 3$. Here one expects that $A_+ \neq \Sigma A^2$, and indeed we will prove this in many cases. We start with some general considerations. Let $\operatorname{Gr}(A) = \bigoplus_{n \geq 0} \operatorname{Gr}_n(A)$ be the graded ring associated to A, where $\operatorname{Gr}_n(A) = \mathfrak{m}^n/\mathfrak{m}^{n+1}$. Let μ be the discrete valuation on K coming from the exceptional divisor in the blowing-up of $\operatorname{Spec} A$ at the closed point; so $\mu(f) = \sup\{n \geq 0\colon\ f \in \mathfrak{m}^n\}$ is the multiplicity of $f \in A$. Given $0 \neq f \in A$ with $n := \mu(f)$, define the *leading form* $\ell(f) \in \operatorname{Gr}_n(A)$ of f to be $\ell(f) := f \bmod \mathfrak{m}^{n+1}$ (note $\ell(f) \neq 0$). If we fix a regular system of parameters $x_1, \ldots, x_d$ for A and write $\xi_i = \ell(x_i)$, then $\operatorname{Gr}(A)$ is the polynomial ring $\kappa[\xi_1, \ldots, \xi_d]$ with the standard grading, and $\ell(f)$ is a non-zero form of degree n in $\kappa[\xi_1, \ldots, \xi_d]$.

1.1. Lemma. *Assume that the residue field κ is formally real. If $0 \neq f \in A$ is a sum of r squares in A, then $\mu(f)$ is even, say $\mu(f) = 2s$, and $\ell(f) \in \operatorname{Gr}_{2s}(A)$ is a sum of r squares of elements in $\operatorname{Gr}_s(A)$.*

Proof. Assume $f = \sum_i f_i^2$, and let $s = \min_i \mu(f_i)$. The residue field of the valuation μ is a purely transcendental extension of κ (of dimension $\dim(A) - 1$), and in particular, it is formally real. Therefore $\mu(f) = 2s$ (Lemma 0.2). In addition it follows that $\ell(f)$ is the sum of the $\ell(f_j)^2$ for those indices j for which $\mu(f_j) = s$. □

It is well known that for $d \geq 3$ there are homogeneous polynomials $h(t_1, \ldots, t_d)$ with integer coefficients, which are psd but not sums of squares in $R[t_1, \ldots, t_d]$, for some (equivalently: any) real closed field R. For example, one may take the homogeneous Motzkin polynomial $h = t_1^2 t_2^2 (t_1^2 + t_2^2 - 3t_3^2) + t_3^6$ [24]. We conclude:

1.2. Proposition. *Let A be a regular local ring with $\dim A \geq 3$ for which $\kappa = A/\mathfrak{m}$ is formally real. Then there is an element $f \in A_+$ which is not a sum of squares in the completion $\widehat{A}$ of A (and, a fortiori, is not sos in A itself).*

Proof. Let $x_1, \ldots, x_d$ be a regular system of parameters for A, where $d = \dim A$. Choose a form $h \in \mathbb{Z}[t_1, \ldots, t_d]$ which is psd but not sos in $\mathbb{R}[t_1, \ldots, t_d]$, and put $f := h(x_1, \ldots, x_d)$. Then $\ell(f) = h(\xi_1, \ldots, \xi_d)$ where $\xi_i := \ell(x_i)$, so $\ell(f)$ is not a sum of squares in $\mathrm{Gr}(A)$. Therefore f is not sos (Lemma 1.1), not even in $\widehat{A}$. On the other hand, $f \in A_+$, since f is the image of the psd element h under the ring homomorphism $t_i \mapsto x_i$ from $\mathbb{Z}[t_1, \ldots, t_d]$ to A. □

1.3. Corollary. *Let A be a noetherian ring. Suppose that there is a real prime ideal $\mathfrak{p}$ in A such that $A_\mathfrak{p}$ is regular of dimension ≥ 3. Then $A_+ \neq \Sigma A^2$.*

Proof. By Proposition 1.2 we find $f \in A$ which is psd in $A_\mathfrak{p}$ but is not a sum of squares in this ring. Let $I = \bigcap_\alpha \mathrm{supp}(\alpha)$, the intersection over the $\alpha \in \mathrm{Sper}\, A$ with $f(\alpha) < 0$. Then $I \not\subset \mathfrak{p}$. Pick any $s \in I \setminus \mathfrak{p}$. Then $s^2 f$ is psd in A, but is not sos in A since it is not sos in $A_\mathfrak{p}$. □

1.4. Remarks.

1. Proposition 1.2 was proved before in the special case where A is the localization of $\mathbb{R}[x, y, z]$ at the maximal ideal (x, y, z); see [7], Prop. 3.5. The proof given above follows the same idea. The result accounts for the existence of what has been called the "bad points" of a psd polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$: These are those points in $\mathbb{R}^n$ where every denominator h in a sums of squares representation $h^2 f = \sum_i f_i^2$ (by polynomials f_i, h) vanishes. Clearly, $P \in \mathbb{R}^n$ is a bad point for f if and only if f is not a sum of squares in the local ring at P. See also [6], Thm. 4.3, where the existence of bad points was probably noticed for the first time in the published literature. Delzell's thesis and a recent abstract ([10], especially the introduction and pp. 57-62) discusses this question further and contains some historical remarks. There one also finds examples (p. 60) of psd elements $f \in \mathbb{R}[x, y, z]_{(x,y,z)}$ which are not sos although their leading forms are.

2. If $\dim A = 3$ and the residue field κ is not formally real (but $K = \mathrm{Quot}(A)$ is), I do not know in general whether $A_+ \neq \Sigma A^2$ still holds. This is indeed so for $\dim A \geq 4$, at least if A is essentially of finite type over a field:

1.5. Proposition. *Let k be a field and A be a regular local k-algebra which is a localization of a finitely generated k-algebra. If $\dim A \geq 4$ and the quotient field of A is formally real, then there exists a psd element in A which is not a sum of squares.*

The proof relies on the following geometric fact:

1.6. Lemma. *Let V be a smooth irreducible quasi-projective k-variety of dimension ≥ 2 whose function field $k(V)$ is formally real. If P is a closed point of V, there exists a smooth prime divisor H on V passing through P, such that the function field $k(H)$ of H is formally real.*

Proof of the proposition. There are a field extension F of k, a regular F-algebra B of finite type and a maximal ideal $\mathfrak{q}$ of B such that $A \cong B_\mathfrak{q}$. Therefore we can assume $A = \mathcal{O}_{V,P}$ where V is a smooth irreducible k-variety, $\dim V \geq 4$, and P is a closed point of V. By a repeated application of Lemma 1.6, we find a curve C on V with formally real function field which passes through P. The local ring $\mathcal{O}_{V,C}$

(of V at the generic point of C) is regular of dimension ≥ 3 and has formally real residue field $k(C)$. Therefore this ring contains psd, non-sos elements (1.2). Since $\mathcal{O}_{V,C}$ is a localization of $\mathcal{O}_{V,P}$, the latter ring must also contain such elements. □

Proof of the lemma. By enlarging k we can assume that $V \otimes_k \overline{k}$ is irreducible. Let ξ be an ordering of k which extends to $k(V)$, and let k_ξ denote the corresponding real closure of k. It suffices to prove the lemma for the (smooth and irreducible) k_ξ-variety $V_{k_\xi} := V \otimes_k k_\xi$ and a lift of P to a closed point of V_{k_ξ}.

So we can and will assume that the field k is real closed. Let $V \hookrightarrow \mathbb{P}^n_k$ be a locally closed embedding. Let $\mathcal{H}$ be the linear system of hyperplanes in $\mathbb{P}^n_k$ which pass through P, and let B be the base locus of the system $\mathcal{H}$. If P is a real point, then $B = \{P\}$; if P is a complex point, then B is the unique line defined over k which contains P. In the latter case we can assume that B is not contained in V, by changing the embedding $V \hookrightarrow \mathbb{P}^n_k$ suitably. Now for generic $L \in \mathcal{H}$, the intersection $L \cap V$ is smooth (and non-empty): Away from $B \cap V$ this is Bertini's theorem, while on $B \cap V$ it is clear since $B \cap V$ is finite. Noting that $V(k)$ is Zariski dense in V, take any such (generic) L for which $(L \cap V)(k) \neq \emptyset$, and let H be a connected component of $L \cap V$ with $H(k) \neq \emptyset$. The function field $k(H)$ is formally real (Artin-Lang), so H does what we want. □

1.7. Remark. Finally, we add a few comments on the case of dimension two, which seems to be the hardest. Given a two-dimensional regular local ring A, is it true that every psd element in A is a sum of squares? Here are a few cases in which this holds by elementary reasons. Let k be a field, $\operatorname{char} k \neq 2$. In each of the following cases we have $A_+ = \Sigma A^2$:

a) $A = \mathcal{O}_{V,P}$, where C is a smooth curve over k, $V = C \times_k \mathbb{A}^1_k$ and P is a closed point of V such that the residue field $k\big(\mathrm{pr}_C(P)\big)$ is formally real (here pr_C is the projection $V \to C$);
b) $A = k[x_1, \ldots, x_n]_{\mathfrak{p}}$ where $\mathfrak{p}$ is a prime ideal of height two with formally real residue field $k(\mathfrak{p})$;
c) $A = k[x, y]_{\mathfrak{m}}$, where $\mathfrak{m}$ is a maximal ideal of $k[x, y]$ with $[k(\mathfrak{m}) : k] \leq 2$ (and $k(\mathfrak{m})$ not necessarily formally real);

For the proof we use the following lemma, which is a corollary to a classical theorem of Artin:

1.8. Lemma. *Let B be a discrete valuation ring with a formally real residue field. Then every psd element in $B[t]$ is a sum of squares.*

Proof. Let $\pi \in B$ be a uniformizer, and let L be the quotient field of B. Let $f \in B[t]$ be psd. Then f is a sum of squares in $L(t)$, and therefore also in $L[t]$ [1]. So there are $f_1, \ldots, f_r \in L[t]$ such that $f = \sum_i f_i^2$, and hence also $g_1, \ldots, g_r \in B[t]$ and $n \geq 0$ with $\pi^{2n} f = g_1^2 + \cdots + g_r^2$. If $n \geq 1$, then reduction of this identity modulo π shows that π divides each g_i. Inductively it follows that π^n divides each g_i. □

In particular, $A_+ = \Sigma A^2$ holds for every ring A which is a localization of $B[t]$. This accounts for a) above, and also for c) since after an affine change of coordinates one can assume $x \in \mathfrak{m}$, using $[k(\mathfrak{m}) : k] \leq 2$. By a result of Lindel ([20], Lemma 1), the ring in b) is isomorphic to $K[u, v]_{\mathfrak{m}}$ for $K = k(x_1, \ldots, x_{n-2})$ and a suitable maximal ideal $\mathfrak{m}$ of $K[u, v]$; so b) is a special case of a).

1.9. Remarks.

1. If R is real closed and $f \in R[x_1, \dots, x_n]$ is psd, then b) above implies that the set of bad points (1.4.1) of f in R^n has dimension $\leq n-3$. This was already proved in [10], Prop. 5.1 by an argument similar to ours.

2. In any local (or even semi-local) ring A, neither necessarily regular nor noetherian, the psd *units* are sums of squares: $A_+ \cap A^* \subset \Sigma A^2$. See [2], p. 153, for example. (We assumed $\frac{1}{2} \in A$, as always.)

3. It is easy to find singular one-dimensional (complete) local domains A for which $A_+ \neq \Sigma A^2$. Indeed, if A is any local ring whose residue field is formally real, the elements of $\mathfrak{m} \setminus \mathfrak{m}^2$ can never be sos. But if A is singular, there may be such elements which are psd, e.g., the element x in $A = \mathbb{R}[\![x,y]\!]/(x^3 - y^2)$ or in $A = \mathbb{R}[\![x,y]\!]/(x^2 + y^2)$.

2. Preliminaries on real curves

Let R be a real closed field, and write $C = R(\sqrt{-1})$ for its algebraic closure. Let X be an irreducible smooth projective curve over R. We will always assume $X(R) \neq \emptyset$. We denote by g the genus and by J the Jacobian of X. Moreover $K = R(X)$, the function field of X. By X_C we denote the base extension of X to C, and we write $K_C = K(i)$ for the function field of X_C, where $i = \sqrt{-1}$.

We use the following conventions about divisors. We consider X as a scheme over $\operatorname{Spec} R$. The group $\operatorname{Div}(X)$ of divisors on X is the free abelian group on the closed points of X. If $D = \sum_P n_P \,.\, P$ is a divisor, then the support of D, $\operatorname{supp}(D)$, is the set of all points P with $n_P \neq 0$. A closed point P is said to be *real* (resp. *complex*) if its residue field is R (resp. C). The discrete valuation on K associated with P is written v_P; we always assume v_P to be normalized, i.e., the value group is $\mathbb{Z}$. If $f \in K^*$, then $\operatorname{div}(f) = \sum_P v_P(f) \,.\, P$ is the (principal) divisor of f. The quotient of $\operatorname{Div}(X)$ by the subgroup of principal divisors is the Picard group $\operatorname{Pic}(X)$. If D is a divisor, then the class of D in $\operatorname{Pic}(X)$ is written $[D]$. Two divisors D, D' are linearly equivalent, denoted $D \sim D'$, if $[D] = [D']$.

Occasionally, we will also have to consider divisors on the complexification X_C. Therefore, we sometimes use notations like $\operatorname{div}_X(f)$ or $\operatorname{div}_{X_C}(f)$, if there is any danger of confusion.

The degree of a closed point P is the degree of its residue field over R. This definition extends to a homomorphism $\deg\colon \operatorname{Pic}(X) \to \mathbb{Z}$ whose kernel is written $\operatorname{Pic}_0(X)$.

The set $X(R)$ decomposes into finitely many connected components (in the semi-algebraic sense; if $R = \mathbb{R}$, this coincides with the usual topological notion). We will also call them *ovals*. Let always s denote their number. We have $1 \leq s \leq g+1$, the first inequality by assumption, the second by Harnack's theorem. A rational function $f \in K^*$ is said to be *locally semidefinite*, or *lsd* for short, if the restriction of f to every oval is (positive or negative) semidefinite. Note that f is lsd if and only if $v_P(f)$ is even for every real point P on X.

Let $G = \operatorname{Gal}(C/R)$; we indicate the action of complex conjugation on points, divisors, functions etc. by a bar. If A is an abelian variety defined over R, then $A(R)_0$ denotes the connected component of the identity in the semi-algebraic group $A(R)$. This subgroup coincides with the image of the norm map $a \mapsto a + \bar{a}$ from $A(C)$ to $A(R)$, and also with $2A(R)$ (compare [8]). We write $H^1(R, A)$ for the Galois cohomology group $H^1\big(G,\, A(C)\big)$.

The exact sequence $0 \to J(C) \to \mathrm{Pic}(X_C) \stackrel{\deg}{\to} \mathbb{Z} \to 0$ of G-modules splits since $X(R) \neq \emptyset$. Therefore it gives an isomorphism

$$H^1(R, J) \stackrel{\sim}{\to} H^1\big(G, \mathrm{Pic}(X_C)\big). \tag{1}$$

On the other hand, the Hochschild-Serre spectral sequence of étale cohomology,

$$E_2^{ij} \;=\; H^i\big(G,\, H^j_{\text{ét}}(X_C, \mathbb{G}_m)\big) \quad \Longrightarrow \quad H^{i+j}_{\text{ét}}(X, \mathbb{G}_m),$$

shows that the inclusion $\mathrm{Pic}_0(X) \subset J(R)$ is an equality, and it gives an exact sequence

$$0 \to \mathrm{Br}(R) \to \mathrm{Br}(X) \to H^1\big(G,\, \mathrm{Pic}(X_C)\big) \to 0, \tag{2}$$

where Br is the Brauer group. (See [28], p. 222, for example.) Recall that $\mathrm{Br}(X)$ can be described as the subgroup of $\mathrm{Br}(K)$ consisting of all classes which are unramified at every real point $P \in X(R)$.

One can make (2) explicit as follows. If $f \in K^*$ is an lsd function, then the quaternion algebra $(-1, f)$ over K is unramified on X, hence lies in $\mathrm{Br}(X)$. Conversely, every element $\alpha \in \mathrm{Br}(X)$ can be written $\alpha = (-1, f)$ with such f (Witt [32]). Fix $\alpha = (-1, f)$, and choose a divisor D on X_C with $\mathrm{div}_{X_C}(f) = D + \overline{D}$. The class $[D] \in \mathrm{Pic}_0(X_C) = J(C)$ of D is anti-invariant under the Galois involution, and hence defines an element in $H^1\big(G,\, J(C)\big) = H^1(R, J)$.

2.1. Lemma. *The map which sends $\alpha = (-1, f)$ to the class of $[D]$ in $H^1(R, J)$ is a well-defined surjective homomorphism $\psi\colon\ \mathrm{Br}(X) \to H^1(R, J)$, whose kernel is generated by $(-1, -1)$.*

Modulo the isomorphism (1), this map ψ coincides with the map in (2), but we won't prove this fact since it will not be used.

Proof. Fix $\alpha = (-1, f) \in \mathrm{Br}(X)$. If $\Theta \in \mathrm{Div}(X_C)$ satisfies $\Theta + \overline{\Theta} = 0$, then one can write $\Theta = \Theta_1 - \overline{\Theta}_1$, so the class of Θ in $H^1(R, J)$ is zero. Thus $\psi(\alpha)$ does not depend on the choice of D. Second, $\psi(\alpha)$ is independent of the choice of f: If $(-1, f) = (-1, f')$, then $f' = f\,(h_1^2 + h_2^2)$ with $h_1,\, h_2 \in K$, and therefore

$$\mathrm{div}_{X_C}(f') \;=\; D' + \overline{D'} \;\text{ with }\; D' = D + \mathrm{div}_{X_C}(h_1 + i h_2) \sim D.$$

Thus it is clear that ψ is a well-defined homomorphism. From its construction one sees easily that ψ is surjective. On the other hand, assume that $\alpha = (-1, f)$ lies in the kernel of ψ. Writing $\mathrm{div}_{X_C}(f) = D + \overline{D}$, there exist $\Theta \in \mathrm{Div}(X_C)$ and $h \in K_C^*$ with $D = \Theta - \overline{\Theta} + \mathrm{div}_{X_C}(h)$. This gives $\mathrm{div}_{X_C}(f) = \mathrm{div}_{X_C}(h\overline{h})$, i.e., $f = c\,.\,h\overline{h}$ with $0 \neq c \in R$. Since clearly $h\overline{h}$ is a sum of two squares in K, we have $\alpha = (-1, c)$. □

2.2. Remark. Let $\alpha = (-1, f) \in \mathrm{Br}(X)$, with $f \in K^*$ lsd. The sign distribution of f, considered as an element of $\{\pm 1\}^s$, depends only on α, so we can denote it by $\mathrm{sgn}(\alpha)$. A classical theorem of Witt [32] states that $\alpha \mapsto \mathrm{sgn}(\alpha)$ is an isomorphism $\mathrm{Br}(X) \stackrel{\sim}{\to} \{\pm 1\}^s$. The last lemma gives us therefore an explicit isomorphism $H^1(R, J) \stackrel{\sim}{\to} \{\pm 1\}^s / \pm 1$, where the right-hand group denotes $\{\pm 1\}^s$ modulo multiplication by -1.

Now let $H := \{fK^{*2}\colon\ \mathrm{div}_X(f) \in 2\,\mathrm{Div}(X)\}$, a subgroup of K^*/K^{*2}. The map $\sigma\colon\ fK^{*2} \mapsto \big[\frac{1}{2}\,\mathrm{div}_X(f)\big]$ is a homomorphism from H to the 2-torsion subgroup ${}_2\,\mathrm{Pic}(X) = {}_2J(R)$ of the Picard group. It is obviously surjective, and its kernel is generated by the square class of -1. On the other hand, we have an obvious

homomorphism $\phi\colon\ H \to \mathrm{Br}(X)$, given by $fK^{*2} \mapsto (-1, f)$. These maps fit into a commutative diagram with exact lines

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \{\pm 1\} & \longrightarrow & H & \xrightarrow{\sigma} & {}_2J(R) & \longrightarrow & 0 \\ & & \| & & \phi\downarrow & & \bar{\phi}\downarrow & & \\ 0 & \longrightarrow & \mathrm{Br}(R) & \longrightarrow & \mathrm{Br}(X) & \xrightarrow{\psi} & H^1(R,J) & \longrightarrow & 0. \end{array} \tag{3}$$

One checks immediately that $\bar{\phi}$ is the natural map of Galois cohomology.

2.3. Lemma.

a) *ϕ is surjective.*
b) *The restriction of $\bar{\phi}$ to ${}_2J(R)_0$ is surjective.*
c) *$L := \ker(\bar{\phi})$ has order 2^g, and $L_0 := L \cap J(R)_0$ has order 2^{g-s+1}.*
d) *L meets every connected component of $J(R)$.*

Proof. b) is a general fact: If A is any abelian variety over R, then the natural map ${}_2A(R)_0 \to H^1(R, A)$ is surjective. Indeed, if $a \in A(C)$ satisfies $a + \bar{a} = 0$, choose $b \in A(C)$ with $2b = a$; then $a' := a - b + \bar{b}$ represents the same class in $H^1(R, A)$ as a, and $a' = b + \bar{b}$ lies in $A(R)_0$, hence in ${}_2A(R)_0$.

This proves b). In particular, $\bar{\phi}$ is surjective, which implies a). On the other hand, the groups ${}_2J(R)$, ${}_2J(R)_0$ and $J(R)/J(R)_0$ have orders 2^{g+s-1}, 2^g and 2^{s-1}, respectively (the first, resp. the third, essentially by a theorem of Weichold, reproved later by Geyer; compare [28], p. 221 and [8], Appendix). Therefore c) and d) follow from b), together with Remark 2.2. □

2.4. Example. X is said to be an *M-curve* if $s = g + 1$, i.e., if the number of connected components is the maximal one allowed by Harnack's theorem. By the last lemma, if X is an M-curve, then ϕ restricts to an isomorphism from ${}_2J(R)_0$ onto $\{\pm 1\}^{g+1}/\pm 1$.

2.5. Remark. Write H_+ for the kernel of $\phi\colon\ H \to \mathrm{Br}(X)$. So H_+ consists of all square classes $fK^{*2} \in K^*/K^{*2}$ for which f is psd and $\mathrm{div}(f)$ is divisible by 2. By (3), σ induces an isomorphism $\sigma\colon\ H_+ \xrightarrow{\sim} L = \ker(\bar{\phi})$, and therefore $H_+ \cong \{\pm 1\}^g$. This last fact was also proved in [29], 4.4, using a different method.

2.6. Lemma. *Let $D = \sum_P n_P \,.\, P$ be a divisor on X of degree zero. The class $[D]$ in $J(R)$ lies in the connected component $J(R)_0$ if and only if for every oval O of $X(R)$ the sum $D(O) := \sum_{P \in O} n_P$ is even.*

Proof. The map which sends D to the s-tuple $\big(D(O) \bmod 2\big)_O \in (\mathbb{Z}/2)^s$ is a homomorphism from $\mathrm{Div}(X)$ to $(\mathbb{Z}/2)^s$, which induces an isomorphism $\mathrm{Pic}(X)/2 \xrightarrow{\sim} (\mathbb{Z}/2)^s$ (Witt, cf. [28], p. 221, for example). So the second condition in the lemma is equivalent to $[D] \in 2\,\mathrm{Pic}(X)$, i.e., to $[D] \in 2J(R) = J(R)_0$. □

For references on the following see, e.g., [22], §3. Let $n \geq 1$, and let $X^{(n)}$ denote the n-fold symmetric power of X, i.e., the quotient of the n-fold direct product $X^n = X \times \cdots \times X$ over R by the natural action of the symmetric group on the n factors. This is a smooth proper R-variety. A C-rational point on $X^{(n)}$ is an unordered n-tuple of C-rational points on X, or, in other words, an effective divisor on X_C of degree n. Fix a rational point $P_0 \in X(R)$. There is a well-defined morphism $\varphi^{(n)}\colon\ X^{(n)} \to J$ of R-varieties whose effect on C-rational points is that

it sends the divisor D to the divisor class $[D - nP_0]$. It is known that, for $n = g$, the morphism $\varphi^{(g)}$ is birational ([22], §5).

The following theorem is the main result of this section. It will be very useful in the sequel:

2.7. Theorem. *Let $P_0 \in X(R)$ be a fixed real point. There is an integer $n \geq 1$ such that for any $\alpha \in J(R)$ there are n points $P_1, \ldots, P_n \in X(R)$ with $\alpha = \sum_i [P_i - P_0]$.*

Here are a few corollaries. Recall that we always assume $X(R) \neq \emptyset$.

2.8. Corollary. *Fix a complex closed point Q in X. Then there is an integer $n \geq 1$ such that, given any even integer $2m \geq n$ and any $\alpha \in J(R)$, there are $2m$ points $P_1, \ldots, P_{2m} \in X(R)$ such that α is the class of the divisor $\sum_i P_i - mQ$.*

Proof. Choose a real point P_0 on X, and let n be as in Theorem 2.7. Given $2m \geq n$ and $\alpha \in J(R)$, there are $P_1, \ldots, P_{2m} \in X(R)$ with $\sum_{i=1}^{2m} [P_i - P_0] = \alpha + m\,[Q - 2P_0]$, by Theorem 2.7. From this the assertion follows. □

2.9. Corollary. *The abelian group $\mathrm{Pic}(X)$ is generated by the classes $[P]$ of all closed points P on X which are real.* □

If D is a divisor on X, then $|D|$ denotes the complete linear system of D, i.e., the set of all effective divisors D' linearly equivalent to D. In this language we may rephrase the theorem as follows:

2.10. Corollary. *There exists an integer $n \geq 1$ with the following property: For every complete linear system $|D|$ with $\deg(D) \geq n$ there is $D' \in |D|$ such that $\mathrm{supp}(D')$ consists of real points.*

Proof. Take n to be the integer from Theorem 2.7, with fixed $P_0 \in X(R)$. If $\deg(D) = d \geq n$, then $[D - dP_0] \in J(R)$, so by 2.7 we find $P_1, \ldots, P_d \in X(R)$ with $D - dP_0 \sim \sum_i (P_i - P_0)$. It suffices to put $D' = \sum_i P_i$. □

The content of Theorem 2.7 is that the composite map $X(R)^n \to X^{(n)}(R) \overset{\varphi^{(n)}}{\to} J(R)$ is surjective for some $n \geq 1$. The proof is easy in the case where $R = \mathbb{R}$, the field of classical real numbers. However, this proof does not work for a non-archimedean real closed base field. Also, it does not seem to produce a bound on the number n in terms of the curve (e.g., in terms of g), at least not without further work. Therefore it does not seem possible to derive from it the case of general R by an immediate application of Tarski's principle. Still we will see, using properties of the real spectrum, that we can deduce the general case from the case $R = \mathbb{R}$.

We start by proving Theorem 2.7 over $R = \mathbb{R}$. For this we need the following lemma:

2.11. Lemma. *Let G be a compact real Lie group, and let U be a non-empty open subset of G. There exists an integer $n \geq 1$ such that every element of G_0 is a product of at most n elements of U.*

Proof. As usual, G_0 denotes the connected component of the identity of G. Write $U(p)$ for the set of all p-fold products $u_1 \cdots u_p$ with $u_i \in U$ ($p \geq 1$). The sets $U(p)$ are open in G. Since elements of finite order are dense in G_0, there is $p \geq 1$ such that $1 \in U(p)$. We can now replace U by a symmetric open neighborhood of 1 contained in $U(p)$. Then the subgroup of G generated by U is G_0, and it is the union of the sets $U(n)$, $n \geq 1$. Since G_0 is compact, we have $U(n) = G_0$ for some n. □

Consider now Theorem 2.7 in the case $R = \mathbb{R}$. We first show that the assertion holds for every $\alpha \in J(R)_0$. Indeed, this case is included in the following more general version, which will be useful later:

2.12. Lemma. *($R = \mathbb{R}$) Let $P_0 \in X(\mathbb{R})$, and let U be any non-empty open subset of $X(\mathbb{R})$. Then there is $n \geq 1$ such that for every $\alpha \in J(\mathbb{R})_0$ there are $P_1, \ldots, P_r \in U$ with $r \leq n$ and $\alpha = \sum_{i=1}^{r} [P_i - P_0]$.*

Proof. Let $\varphi^{(g)} \colon X^{(g)} \to J$ be the morphism of $\mathbb{R}$-varieties considered above. Let V be the subset of $X^{(g)}(\mathbb{R})$ which consists of all effective divisors $D \in \mathrm{Div}(X)$ of degree g which have $\mathrm{supp}(D) \subset U$. This V contains a non-empty open subset of $X^{(g)}(\mathbb{R})$. Since $\varphi^{(g)}$ is birational and $X^{(g)}$ is smooth, the image of V under $\varphi^{(g)}$ contains a non-empty open subset of $J(\mathbb{R})$. Therefore the assertion follows from Lemma 2.11. □

We return to the proof of Theorem 2.7 in the case $R = \mathbb{R}$. Let $C_0, \ldots, C_{s-1}$ be the ovals of $X(\mathbb{R})$, with $P_0 \in C_0$, and choose points $P_i \in C_i$ $(i = 1, \ldots, s-1)$. From Lemma 2.6 it follows that the divisor classes $[P_i - P_0]$ $(1 \leq i \leq s-1)$ generate the group $J(\mathbb{R})/J(\mathbb{R})_0$, which is isomorphic to $(\mathbb{Z}/2)^{s-1}$. From this, together with Lemma 2.12, follows the case $R = \mathbb{R}$ of the theorem.

In order to prove Theorem 2.7 over an arbitrary real closed base field R, note first that the theorem is certainly true if R is archimedean. (Either justify that the proof given for $\mathbb{R}$ remains valid for R, or consider the extension $R \subset \mathbb{R}$ and apply Tarski's principle.)

Now let R be arbitrary. There is a finitely generated subfield k of R such that X and P_0 can be defined over k. Slightly abusing notation, we continue to use the letter X for this curve over k, and also use X^n, $X^{(n)}$, J for the corresponding k-varieties. It suffices to show that for every ordering ξ of k there is some integer $n = n_\xi \geq 1$ such that the composite map $X(k_\xi)^n \to X^{(n)}(k_\xi) \overset{\varphi^{(n)}}{\to} J(k_\xi)$ is surjective. Here k_ξ is the real closure of k at ξ.

For the proof we need the language of the real spectrum; see the notations section at the end of the introduction and references given there. In particular, we need the notion of constructible subsets of the real spectrum. If V is any scheme, let V_r be the real spectrum of V. (If $\{V_i\}$ is an open affine covering of V, then V_r is the topological space obtained by glueing the real spectra of the rings $\Gamma(V_i, \mathcal{O}_V)$.) For $n \geq 1$ let $W(n)$ be the image set of the map $(X^n)_r \to J_r$ between real spectra induced by the composite morphism $X^n \to X^{(n)} \to J$ of k-varieties. Then $W(1) \subset W(2) \subset \cdots$ is an ascending sequence of constructible subsets of J_r ([9], Prop. 2.3). Given an ordering ξ of k, let $W(n)_\xi$ denote the intersection of $W(n)$ with the set $J_{r,\xi}$ of points in J_r which lie over ξ. Then $J_{r,\xi}$ is identified with $\widetilde{J(k_\xi)}$ (:= the real spectrum of the base extension $J \otimes_k k_\xi$), and its subset $W(n)_\xi$ is the constructible subset associated with the (semi-algebraic) image set of the map $X(k_\xi)^n \to J(k_\xi)$. Therefore, if the ordering ξ is archimedean, there is $n \geq 1$ with $J_{r,\xi} \subset W(n)$.

Now if ξ is any ordering of k with $J_{r,\xi} \subset W(n)$, then also $J_{r,\eta} \subset W(n)$ holds for all $\eta \in \mathrm{Sper}\, k$ sufficiently close to ξ. The reason is that the map $J_r \to \mathrm{Sper}\, k$ between the real spectra sends constructible sets to constructible sets ([9], Prop. 2.3). Therefore, if we knew that the archimedean orderings of k form a dense subset

of $\operatorname{Sper} k$, we would be finished since $\operatorname{Sper} k$ is compact. But this is actually true, since the field k is finitely generated. We isolate this fact as a lemma:

2.13. Lemma. *Let k be a finitely generated field extension of $\mathbb{Q}$. Then the archimedean orderings of k form a dense subset of* $\operatorname{Sper} k$.

Proof. Choose a subfield F of k which is purely transcendental over $\mathbb{Q}$ and over which k is finite. It suffices to prove density of archimedean orderings for $F = \mathbb{Q}(t_1, \dots, t_n)$, since the restriction map $\operatorname{Sper} k \to \operatorname{Sper} F$ is open. Let $p_1, \dots, p_r$ be polynomials in $\mathbb{Q}[t_1, \dots, t_n]$ such that there exists an ordering of F which makes all p_j positive. This means that there are R, a real closed field, and $x \in R^n$, such that $p_j(x) > 0$ for all j. Then the same is true for $R = \mathbb{R}$, the classical reals, by Tarski. Now it is clear that one can choose $x \in \mathbb{R}^n$ such that $p_j(x) > 0$ and in addition $x_1, \dots, x_n$ are algebraically independent over $\mathbb{Q}$. The embedding $F \hookrightarrow \mathbb{R}$ given by $t_i \mapsto x_i$ defines an archimedean ordering of F under which the p_j are positive. □

The proof of Theorem 2.7 is now complete. □

2.14. Remark. In Theorem 2.7 we can fix a finite subset S of $X(R)$ and require that $P_1, \dots, P_n$ do not lie in S. By an iterated application of this remark, we can even reach in addition that in 2.7 the points $P_1, \dots, P_n$ are distinct from each other.

2.15. Remark. Note that Lemma 2.12 is definitely false in general if the base field R is not archimedean. Indeed, the topological group $J(R)_0$ has many open (though not semi-algebraic!) subgroups in this case, e.g., subgroups defined with the help of a non-trivial order-compatible valuation on R. Therefore, if the open subset U is too small, the subgroup of $J(R)$ generated by classes $[P - P_0]$ with $P \in U$ can be a small proper subgroup of $J(R)_0$.

We conclude this section with two preparatory results about sums of squares on curves.

2.16. Lemma. *Let Y be a smooth irreducible affine curve over R, and let S be a finite set of closed points of Y. If in $R[Y]$ every psd function is sos, then the same holds in $R[Y - S]$.*

Proof. This would be obvious if $R[Y - S]$ were a localization of $R[Y]$; this, however, need not be the case if Y is non-rational, and so one has to be a little more careful.

We first show that there exists $s \in R[Y]$ which vanishes in each point of S, and such that every zero of s in $Y - S$ is real. Indeed, let X be the smooth compactification of Y, and consider the divisor $D := \sum_{P \in S} P$ on X. Choose a point $\infty \in X \smallsetminus Y$. By 2.7 and 2.8, there are $P_1, \dots, P_r \in X(R)$ and $k \geq 0$ such that $-D \sim \sum_i P_i - k\infty$. Let $s \in R(X)^*$ be a function with $\operatorname{div}(s) = D + \sum_i P_i - k\infty$. Clearly, $s \in R[Y]$, and s has the required properties.

Now for $g \in R[Y - S]$, there is $n \geq 0$ such that $s^{2n} g =: f$ has no poles on Y, i.e., lies in $R[Y]$. If g is psd on $Y - S$, then f is psd on Y, and so we can write $f = \sum_j f_j^2$ with $f_j \in R[Y]$, by hypothesis. Thus $g = \sum_j (\frac{f_j}{s^n})^2$. If P is any real point in $Y - S$, then $v_P(\frac{f_j}{s^n}) \geq 0$ by Lemma 0.2. Since s has no complex zeros on $Y - S$, the $\frac{f_j}{s^n}$ lie in $R[Y - S]$, and hence g is sos in $R[Y - S]$. □

2.17. Proposition. *Let Y be a smooth affine curve over R which is rational. Then every psd element of $R[Y]$ is a sum of squares.*

Proof. We can assume $Y(R) \neq \emptyset$, since otherwise every element of $R[Y]$ is a sum of squares. By Lemma 2.16 it is enough to consider the case where Y is either the affine line $\mathbb{A}^1$ or the circle, i.e., the plane curve with equation $x^2 + y^2 = 1$. Both cases are well known and elementary. A proof for the second can be found in [5], Thm. 3.7, but it is also an immediate consequence of the first. □

2.18. Example. The last proposition does not extend to singular rational curves. On the cuspidal curve $y^2 = x^3$ the function x is psd, but locally at the origin it is not sos (1.9.3). A different kind of example is given by the nodal curve $y^2 = x^3 + x^2$: The function $x + 1$ is psd but not sos, although it is a sum of squares locally at every point of the curve.

3. Non-rational curves I: The non-compact case

Let R be a real closed base field, and let X be a smooth projective irreducible curve of genus $g \geq 1$ over R. Let T be a finite non-empty set of closed points of X, and let $Y = X \setminus T$. The points in T will also be called the *points of Y at infinity*. *Throughout this section we assume that every point of Y at infinity is real*. Under this assumption it is generally quite easy to construct psd functions in $R[Y]$ which are not sos. We will actually show that there are many such functions, in a suitable sense.

We start with an easy example which already conveys the reason why one should expect to find psd, non-sos functions:

3.1. Example. Let Y be the affine elliptic curve with equation $y^2 = x^3 + x$. Let ∞ be its point at infinity (which is real). The function x is visibly psd on Y, but an immediate argument shows that x is not sos. To understand this example more conceptually, note that x has a pole at ∞ of order two. If x were sos, say $x = \sum f_i^2$ with $f_i \in R[Y]$, then the pole order of each f_i at ∞ could be at most one. Since the curve Y is not rational, this would imply that the f_i are constant, a contradiction.

This argument can be generalized to give the following result:

3.2. Theorem. *Let Y be a smooth connected affine curve of genus $g \geq 1$ which has only real points at infinity. Then there exists $f \in R[Y]$ which is strictly positive on $Y(R)$ but is not a sum of squares in $R[Y]$.*

Proof. Write $Y = X \setminus T$ as above. Fix a point $\infty \in T$ and let $U = X \setminus \{\infty\}$. Let Σ be the semigroup of all integers $m \geq 0$ for which there exists $f \in R[U]$ with $v_\infty(f) = -m$. By Riemann-Roch, there are only finitely many integers $m \geq 0$ not contained in Σ. Hence there exists m with $m \notin \Sigma$ but $2m \in \Sigma$ (otherwise we would have $1 \in \Sigma$, contradicting $g \geq 1$). Let $f \in R[U]$ with $v_\infty(f) = -2m$; hence also $v_\infty(c \pm f) = -2m$ for all $c \in R$. I claim that $c \pm f$ is never a sum of squares in $R[Y]$. Indeed, assume $c \pm f = \sum_i f_i^2$ with $f_i \in R[Y]$. Then by Lemma 0.2, the f_i lie actually in $R[U]$ since the points in $T \setminus \{\infty\}$ are real; and $v_\infty(f_i) = -m$ for at least one index i, again by Lemma 0.2, since ∞ is real. By our choice of m, this can't happen.

On the other hand, the value $f(P)$ has a unique sign for $P \in U(R)$ approaching ∞, since $v_\infty(f)$ is even. Therefore f is bounded on $U(R)$ from either above or below. Hence there exist $c \in R$ and a sign $\epsilon \in \{\pm 1\}$ such that $c + \epsilon f$ is strictly positive on $U(R)$. □

3.3. Remark. Beyond proving the mere existence of psd, non-sos functions, we want to show that there are "many" such functions. For this it will be convenient to use the notion of a preorder, which we now recall:

Let A be a ring. A *preorder* (*cône* in [3]) in A is a subset P of A which is closed under addition and multiplication and contains all (sums of) squares. Any intersection of preorders is again a preorder, so it makes sense to speak of the preorder generated by a family of elements of A, or to say that a given preorder is finitely generated. Given finitely many elements $f_1, \dots, f_r$ in A, the preorder generated by those consists of all sums

$$\sum_{i \in \{0,1\}^r} s_i \, f_1^{i_1} \cdots f_r^{i_r}$$

where the s_i are sums of squares in A. The smallest preorder in A is ΣA^2.

Now let V be an affine R-variety. If P is a finitely generated preorder in the coordinate ring $R[V]$, we associate with P the basic closed semi-algebraic set

$$\mathcal{S}(P) \;:=\; \{x \in V(R) \colon\; f(x) \geq 0 \text{ for every } f \in P\}.$$

Note that if P is generated by $f_1, \dots, f_r$, then $\mathcal{S}(P)$ is the subset of $V(R)$ described by the simultaneous inequalities $f_1 \geq 0, \dots, f_r \geq 0$.

Conversely, if S is a subset of $V(R)$, we can associate with S the preorder

$$\mathcal{P}(S) \;:=\; \{f \in R[V] \colon\; f \geq 0 \text{ on } S\}.$$

Such a preorder is called saturated. The saturation of a finitely generated preorder P is defined to be $\widehat{P} := \mathcal{P}(\mathcal{S}(P))$. For example, the saturation of $P = \Sigma R[V]^2$ is $R[V]_+$. (There is good reason to study the saturation of arbitrary preorders, but for this one has to use the real spectrum, and we won't go into it here.)

A basic question in real algebraic geometry is: Given $f, f_1, \dots, f_r \in R[V]$ such that $f \geq 0$ on the closed semi-algebraic set $\{f_1 \geq 0, \dots, f_r \geq 0\}$, when is f contained in the preorder generated by $f_1, \dots, f_r$? In other words, given a finitely generated preorder P and $f \in \widehat{P}$, when is $f \in P$? In particular, when is P saturated?

These questions are closely related to the psd = sos question. Indeed, if W is the affine R-scheme defined by $R[W] = R[V][t_1, \dots, t_r]/(t_1^2 - f_1, \dots, t_r^2 - f_r)$, then it is elementary to see that f is psd on W, and that f lies in P if and only if f is a sum of squares in $R[W]$.

As before, assume now that X is an irreducible smooth projective curve over R and that $Y = X \smallsetminus T$, where T is a finite non-empty subset of $X(R)$.

3.4. Theorem. *If X has genus $g \geq 1$, then the preorder of all psd functions in $R[Y]$ is not finitely generated.*

If the base field is archimedean, we can prove the following version, which is much stronger. It shows that there is a large class of saturated preorders which are not finitely generated (cf. the discussion in Remark 3.3):

3.5. Theorem. *Let $R = \mathbb{R}$, and let X, T and $Y = X \smallsetminus T$ be as before. Let P be a finitely generated preorder in $\mathbb{R}[Y]$, and assume that T is contained in the closure of the semi-algebraic set $\mathcal{S}(P)$, the closure being taken in $X(\mathbb{R})$. Then there exists a psd function $f \in \mathbb{R}[Y]$ which is not contained in P.*

The proof of 3.5, resp. 3.4, is based on the following technical lemma, resp. on its subsequent corollary:

3.6. Lemma. *Let X be a smooth projective irreducible curve over $\mathbb{R}$ of genus $g \geq 1$, let $\infty \in X(\mathbb{R})$ and $Y = X \setminus \{\infty\}$. Moreover let U be a non-empty open subset of $X(\mathbb{R})$.*

a) *If X is not an M-curve (cf. 2.4), or if U is not contained in the same oval as ∞, there exists $f \in \mathbb{R}[Y]$ such that*
   (α) *f is psd, but not a square;*
   (β) *all zeros of f are real, and they all lie in U.*

b) *If X is an M-curve and $U \cup \{\infty\}$ is contained in a single oval C, such f does not exist. But for any preassigned point $N \in X(\mathbb{R})$ with $N \notin C$, there is $f \in \mathbb{R}[Y]$ satisfying (α) and*
   (β') *f has only real zeros, of which one is a double zero at N and all others lie in U.*

Over an arbitrary real closed field R, the lemma fails for general U (Remark 3.8 below). However, from its proof we will at least get the case $U = X(R) \setminus \{\text{a finite set}\}$ without additional effort:

3.7. Corollary. *Let R be an arbitrary real closed field and X a smooth projective irreducible non-rational curve over R. Let $\infty \in X(R)$ and $Y = X \setminus \{\infty\}$. Given finitely many points $Q_1, \ldots, Q_r \in Y(R)$, there is a function $f \in R[Y]$ which is psd but not a square, has only real zeros, and satisfies $f(Q_i) \neq 0$ $(i = 1, \ldots, r)$.*

Proof of 3.6 and 3.7. Suppose we are in the situation of Lemma 3.6. Let s be the number of ovals of $X(\mathbb{R})$. We first assume that X is not an M-curve, i.e., $s < g+1$. Let L and L_0 be defined as in Lemma 2.3. Since X is not an M-curve, we find $0 \neq \alpha \in L_0$, by 2.3c). According to Lemma 2.12 there are $n \geq 1$ and $P_1, \ldots, P_n \in U$ such that α is the class of the divisor $D := P_1 + \cdots + P_n - n \cdot \infty$. Let $f \in \mathbb{R}(X)^*$ be a function with $\operatorname{div}(f) = 2D$. By diagram (3) (before Lemma 2.3), one of $\pm f$ is psd. Replacing f by $-f$ if necessary, it is clear that (α) and (β) are satisfied.

Exactly the same argument works over an arbitrary real closed base field R if $U = X(R) \setminus \{Q_1, \ldots, Q_r\}$, replacing the reference to 2.12 by 2.7 and 2.14.

Now assume $s = g + 1$ in the situation of 3.6. Let $\varphi = \varphi^{(1)} \colon X \to J$ be the canonical morphism associated to the base point ∞, cf. Sect. 2. If $U \cup \{\infty\}$ is not contained in a single oval, then the (open) subgroup of $J(\mathbb{R})$ generated by $\varphi(U)$ is not contained in $J(\mathbb{R})_0$ (2.6). Therefore it contains some $0 \neq \alpha \in L$, (cf. 2.3d); one can now proceed as before. Again, the argument extends to general R if $U = X(R) \setminus \{Q_1, \ldots, Q_r\}$, since then $\varphi(U)$ generates all of $J(R)$ (2.14).

Consider now the situation of 3.6b) (so $R = \mathbb{R}$ again). Then $[N - \infty] \notin J(\mathbb{R})_0$, so there is $0 \neq \beta \in L$ lying in the same connected component as $[N - \infty]$. Now apply the above argument to $\alpha := \beta - [N - \infty]$: Since $\alpha \in J(\mathbb{R})_0$, there are $P_1, \ldots, P_n \in U$ with $\alpha = \sum_i [P_i - \infty]$. Hence there is $f \in \mathbb{R}(X)^*$ with $\operatorname{div}(f) = 2N + 2\sum_i P_i - 2(n+1)\infty$. Again by (3), one of $\pm f$ is psd; and $\pm f$ are not squares since $\beta \neq 0$.

Note that if X is an M-curve and $U \subset C$, there is no f satisfying (α) and (β), since $\varphi(U) \subset J(\mathbb{R})_0$ in this case, and $L_0 = \{1\}$. □

3.8. Remark. If R is non-archimedean and $U \neq \emptyset$ is a very small open subset of $X(R)$, the (open) subgroup of $J(R)$ generated by $\varphi(U)$ will be very small, and in

particular, it need not contain any 2-torsion point of $J(R)$. Therefore Lemma 3.6 is definitely false in general for such R.

We now give the proof of Theorem 3.5. Say P is generated by $f_1, \dots, f_r \in \mathbb{R}[Y]$. We can assume that the f_ν are not constant. Write $S := \mathcal{S}(P)$ for the basic closed set associated with P. Since S is not a finite set, there is a non-empty open subset U of $Y(\mathbb{R})$ contained in S. By removing finitely many points from U we can in addition assume that none of the f_ν vanishes anywhere in U.

Fix a point $\infty \in T$, and let $f \in \mathbb{R}[Y]$ be a function as in Lemma 3.6 (part a) or b), depending on the situation). We proceed to show that $f \notin P$ (if N is suitably chosen in case b), see below).

First assume we are in case a), so all zeros of f lie in $U \subset S$. Assume $f \in P$. Then $f = g_1 + \cdots + g_s$ where each g_j is a product $a^2 f_1^{i_1} \cdots f_r^{i_r}$ with $i_\nu \in \{0, 1\}$ and $a \in \mathbb{R}[Y]$. Now Lemma 0.2 shows $v_M(g_j) \geq v_M(f)$ for all points $M \in X(\mathbb{R})$ in the closure of S. In particular, this remark applies to the zeros of f and the points in T. So $\operatorname{div}(g_j) \geq \operatorname{div}(f)$ for all j, which shows that $g_j = f$ up to a positive scalar factor. Since the f_ν do not vanish in the zeros of f, the g_j must be squares. This contradicts that f is not a square.

Now assume we are in the exceptional case b) of 3.6. We can then assume $\operatorname{div}(f) = 2N + 2D$, where $D = P_1 + \cdots + P_m - (m+1)\infty$ with $P_1, \dots, P_m \in S$, and $N \in X(\mathbb{R})$ is a point of which we can require $N \notin E := T \cup \{M\colon\ 2M \sim 2\infty\}$. Assuming $f \in P$, we have again $f = g_1 + \cdots + g_s$ where each g_j is a product $a^2 f_1^{i_1} \cdots f_r^{i_r}$ as above. The same argument as before gives us now $\operatorname{div}(g_j) \geq 2D$ for all j. Since $f_\nu(P_\mu) \neq 0$, the summands g_j must therefore either be squares a^2, or else products $a^2 f_\nu$, in which case the pole order of f_ν at ∞ must be 2. Actually the latter case cannot occur: We would necessarily have $\operatorname{div}(a^2) = 2\sum_\mu (P_\mu - \infty) = 2D + 2\infty$ (since $f_\nu(P_\mu) \neq 0$), which would imply $2N - 2\infty = \operatorname{div}(f/a^2)$, contradicting $N \notin E$. Therefore the summands g_j must be squares again. But then $\operatorname{div}(g_j) \geq \operatorname{div}(f)$ (and thus equality) as before, contradicting that f is not a square. □

The proof of Theorem 3.4 (over arbitrary R) works exactly along the lines of the first part of the above proof, using Corollary 3.7 instead of Lemma 3.6. □

3.9. Corollary. *($R = \mathbb{R}$) Assume $|T| = 1$, i.e., $Y = X \smallsetminus \{\infty\}$ with $\infty \in X(\mathbb{R})$, and assume again $g \geq 1$. If P is any finitely generated preorder in $\mathbb{R}[Y]$ for which $\mathcal{S}(P)$ is not compact, then P does not contain all psd functions in $\mathbb{R}[Y]$, and in particular, is not saturated.*

4. Non-rational curves II: The compact case

We now study the case of "compact" curves. More precisely, let X be a smooth irreducible projective curve over a real closed field R, and let ∞ be a *complex* closed point of X. We study sums of squares on the affine curve $Y := X \smallsetminus \{\infty\}$. Note that $Y(R)$ is semi-algebraically compact since $Y(R) = X(R)$. We will always use g to denote the genus of X. Again our basic question is: Is every psd function in $R[Y]$ a sum of squares in $R[Y]$?

The main result (Theorem 4.10) says that this is indeed the case for elliptic curves, provided that the base field R is archimedean. For non-archimedean R there are, however, counter-examples (Theorem 4.11). One wonders whether over archimedean R the positive result might hold for arbitrary genus.

Assume that R is archimedean, i.e., $R \subset \mathbb{R}$. Then Schmüdgen's theorem [30] implies that every *strictly* positive $f \in R[Y]$ is a sum of squares. (See [33] for

Wörmann's beautiful proof of Schmüdgen's theorem in purely algebraic terms.) But it is not clear whether this fact can be used for proving that all psd functions are sos. Instead we try to make a reduction to the case of psd functions with only real zeros. We first present some arguments which are valid for arbitrary genus, and then show how the argument can be completed in the special case $g = 1$.

Let $K = R(X)$ always be the function field of X, resp. Y, and write $K_C = K(i)$, $i = \sqrt{-1}$. We will assume $g \geq 1$, the case $g = 0$ being clear (2.17).

4.1. Proposition. *Let $f \in R[Y]$ be a psd function. Then there is a psd function $f' \in R[Y]$ which has at most $g-1$ complex zeros (counted with multiplicity), and such that $f - f'$ is a sum of squares in $R[Y]$.*

Proof. Let $2n$, resp. r, be the number of real, resp. complex, zeros of f. In other words, we have $\mathrm{div}_X(f) = 2D + D' - (n+r)\infty$ where D, D' are effective divisors with $\mathrm{supp}(D) \subset X(R)$ and $\deg(D) = n$, and $\mathrm{supp}(D') \cap X(R) = \emptyset$, $\infty \notin \mathrm{supp}(D')$ and $\deg(D') = 2r$.

Assume that there exists a sum of squares $h \neq 0$ in $R[Y]$ with $\mathrm{div}_X(h) \geq 2D-(n+r)\infty$. The function $\frac{h}{f} \in K^*$ is psd and has no real poles, so it is bounded on $X(R) = Y(R)$. Hence it assumes its maximal value $\mu > 0$ in some point $P \in X(R)$. Put $f' := f - \frac{1}{\mu}h$. Clearly $v_P\big(\mu - \frac{h}{f}\big) \geq 2$, which means $v_P(f') \geq 2 + v_P(f)$. Since obviously $v_Q(f') \geq v_Q(f)$ for every real point Q, the function $f' \in R[Y]$ has at least one double real zero more than f. Since on the other hand the pole order has not increased, we conclude that f' has at least one complex zero less than f.

To see how far this will take us, consider divisors on the complexification X_C of X. Denote the two closed points of X_C which lie over the closed point ∞ of X by ∞_0 and $\infty_1 = \overline{\infty_0}$. Assume that p, q are integers with $p+q = n+r$ such that there exists $a \in K_C^*$ with $\mathrm{div}_{X_C}(a) \geq D - p\infty_0 - q\infty_1$. By Riemann-Roch, such a will certainly exist if $r \geq g$. The function $h := a\bar{a}$ is a sum of two squares in $R[Y]$ and satisfies $\mathrm{div}_X(h) \geq 2D - (n+r)\infty$. Using the above argument, we can therefore reduce the number of complex zeros of f by at least one.

In summary, we have proved that by successively subtracting suitable sums of squares from f we can arrive at a psd function f' with at most $g-1$ complex zeros. □

Now we consider psd functions $f \in R[Y]$ which have only real zeros. For such f we show that the question whether f is sos depends only on the square class of f in K^*, and that there are exactly 2^{g+1} such square classes.

4.2. Lemma. *Let $0 \neq f$, $f' \in R[Y]$ be such that f'/f is a square in K^*. If f is sos, and if $v_Q(f') \geq v_Q(f)$ for every complex point $Q \neq \infty$, then also f' is sos.*

Proof. Let $h \in K^*$ with $f' = fh^2$, and let $f = \sum_j f_j^2$ be an sos representation in $R[Y]$. Thus $f' = \sum_j (f_j h)^2$, and it suffices to show $f_j h \in R[Y]$ for all j. From the hypothesis it follows that h, and therefore $f_j h$, has no complex poles other than ∞. On the other hand, if P is a real point, then clearly $v_P(f_j h) \geq 0$, by Lemma 0.2. □

4.3. Proposition. *Let W be the subgroup of K^*/K^{*2} consisting of the square classes of all psd functions $f \in R[Y]$ which have only real zeros. Then the order of W is equal to 2^{g+1}.*

Proof. The map which sends fK^{*2} to $v_\infty(f) \bmod 2$ is a homomorphism $\rho\colon W \to \mathbb{Z}/2$. Let W_0 be its kernel. Recall the definitions of H, L etc. from 2.3. I claim that $W_0 = H_+$ (2.5); i.e., given $f \in K^*$, the square class fK^{*2} lies in W_0 iff f is psd and $\operatorname{div}(f)$ is divisible by 2. Clearly $W_0 \subset H_+$. To prove equality, let $\theta \in L = \ker(\bar{\phi})$ be given. We show that the square class in H_+ which maps to θ under $\sigma\colon H_+ \xrightarrow{\sim} L$ can be represented by a psd function in $R[Y]$. Indeed, there are $m \geq 1$ and real points $P_1, \dots, P_{2m}$ on X with $\theta = [\sum_j P_j - m\infty]$ (2.8). Thus there is $f \in K^*$ with $\operatorname{div}(f) = 2\big(\sum_j P_j - m\infty\big)$. One of $\pm f$ is psd since $\bar{\phi}(\theta) = 0$, so $\pm fK^{*2} \in H_+$ for a suitable choice of sign $\pm$; and $\sigma(\pm fK^{*2}) = \theta$. This completes the proof of $W_0 = H_+$.

Thus W_0 has order 2^g (2.5), and it remains to show that ρ is not trivial. Fix a base point $P_0 \in X(R)$ and an odd number $q \geq n$, with n as in Theorem 2.7. The class $\alpha := [\infty - 2P_0]$ lies in $J(R)_0$ by Lemma 2.6, so there is $\beta \in J(R)_0$ with $2\beta = q\alpha$. By 2.7 we can write $\beta = \sum_{i=1}^{q} [P_i - P_0]$ with $P_1, \dots, P_q \in X(R)$. Thus there exists $f \in R[Y]$ with $\operatorname{div}(f) = 2\sum_{i=1}^{q} P_i - q\infty$. Clearly the function f is lsd.

Choose $\theta \in {}_2J(R)$ such that $\bar{\phi}(\theta) = \psi(-1, f)$, cf. diagram (3). Moreover choose points $Q_1, \dots, Q_{2m} \in X(R)$ such that θ is the class of the divisor $\sum_{i=1}^{2m} Q_i - m\infty$ (2.8). Let $f' \in R[Y]$ be a function with $\operatorname{div}(f') = 2\big(\sum_{i=1}^{2m} Q_i - m\infty\big)$. The function f' is lsd, and $\psi(-1, f') = \psi(-1, f)$ by construction. So one of $\pm ff'$ is psd. Moreover $ff' \in R[Y]$ is a function which has only real zeros and for which $v_\infty(ff') = -(q + 2m)$ is odd. □

4.4. Corollary. *Let $f_0, f_1, \dots, f_g \in R[Y]$ be $g + 1$ psd functions with only real zeros whose classes in K^*/K^{*2} are linearly independent (over $\mathbb{Z}/2$). If each f_i is sos, then every psd function $f \in R[Y]$ with only real zeros is sos.*

This follows immediately from the proposition, using Lemma 4.2. □

To check whether every psd function on Y with only real zeros is sos is therefore, in principle, a finite task. We will now carry it out explicitly in the case of elliptic curves.

4.5. So assume that $g = 1$. It follows from Riemann-Roch that Y is isomorphic to a plane affine curve with equation

$$y^2 + q(x) = 0, \tag{4}$$

where (x, y) are plane affine coordinates and $q(x)$ is a quartic separable polynomial. Since $Y(R)$ is semi-algebraically compact, we can assume that $q(x)$ is monic; moreover $q(x)$ has at least two real roots since $Y(R) \neq \emptyset$. So we can write

$$q(x) = (x - a)(x - b) \cdot q_1(x)$$

where q_1 is a monic quadratic polynomial which has positive values outside the open interval $]a, b[$. Let $f_0 = x - a$ and $f_1 = b - x$, and write $q = f_i \cdot h_i$ with $h_i \in R[x]$ for $i = 0, 1$.

4.6. Lemma. *The following two conditions are equivalent:*

(i) *Every psd element of $R[Y]$ is a sum of squares;*
(ii) *for $i = 0, 1$ there are psd polynomials $P_i(x)$, $Q_i(x) \in R[x]$ satisfying $P_i f_i - Q_i h_i = 1$.*

Proof. Since $g = 1$, condition (i) is actually equivalent to the apparently weaker condition that every psd element with only real zeros is a sum of squares. This follows from Proposition 4.1. On the other hand, Corollary 4.4 shows that the latter is equivalent to f_0 and f_1 being sums of squares: Indeed, f_0 and f_1 are psd functions on Y with only real zeros, and they are linearly independent in K^*/K^{*2}.

We show now that f_i is sos if and only if condition (ii) holds for i ($i = 0, 1$). Fix one of $i = 0, 1$, and write $f = f_i$ and $h = h_i$, so $y^2 + fh = 0$. First suppose that there are psd polynomials $P(x)$, $Q(x)$ in $R[x]$ with $Pf - Qh = 1$. Multiplying this identity with f gives $Pf^2 + Qy^2 = f$, which shows that f is sos.

Conversely assume that f is sos. Let $M \in Y(R)$ denote the (unique) zero of f on Y. We need the following

4.7. Sublemma. *The ideal I in $R[Y]$ consisting of the functions which vanish at M is the $R[x]$-submodule of $R[Y]$ generated by f and y.*

Proof. I is generated by f and y as an ideal: For this it suffices that f and y generate I locally at any closed point N of Y. If $N \neq M$, then f generates I at N since $f(N) \neq 0$, while y generates I at M since $v_M(y) = 1$. The assertion of 4.7 follows now from $R[Y] = R[x] + yR[x]$ and $y^2 \in fR[x]$. □

Back to the proof of Lemma 4.6. Suppose $f = \sum_j a_j^2$ with $a_j \in R[Y]$. Then every a_j must vanish in M, so by 4.7 we can write $a_j = P_j(x) \cdot f + Q_j(x) \cdot y$ with polynomials P_j, $Q_j \in R[x]$. Expanding the sum of squares we find that $f = \sum_j P_j(x)^2 f^2 + \sum_j Q_j(x)^2 y^2$ (and $\sum_j P_j(x)Q_j(x) = 0$). Thus we have found psd polynomials $P(x)$, $Q(x)$ in $R[x]$ with $f = P(x)f^2 - Q(x)fh$, from which identity (ii) follows. Lemma 4.6 is proved. □

We will now show that the equivalent conditions of Lemma 4.6 are indeed satisfied, provided that the base field R is archimedean. Indeed, let f be one of $f_0 = x - a$ and $f_1 = b - x$, and let $h = h(x)$ be defined by $y^2 + fh = 0$, as before. The existence of psd polynomials P, Q in $R[x]$ with $Pf - Qh = 1$ is a particular case of the following more general result (put $g := -h$ there):

4.8. Proposition. *Let R be an archimedean real closed field, and let f, $g \in R[x]$ be two polynomials satisfying the following conditions (a)-(c):*

(a) $(f, g) = 1$;
(b) *the set* $\{x \in R\colon\ f(x) \geq 0,\ g(x) \geq 0\}$ *is bounded;*
(c) *the set* $\{x \in R\colon\ f(x) \leq 0,\ g(x) \leq 0\}$ *is empty.*

Then there exist psd polynomials P, $Q \in R[x]$ such that $Pf + Qg = 1$.

Observe that conditions (a)-(c) are trivially necessary for the existence of P and Q, at least if f and g are not constant.

Proof. We can assume $R = \mathbb{R}$. Let p, q be any two polynomials with $pf + qg = 1$. We are looking for a polynomial F such that $P := p + Fg$ and $Q := q - Ff$ are psd (note $Pf + Qg = 1$). In other words, we want the two inequalities

$$F(x)g(x) \geq -p(x), \tag{5a}$$

$$F(x)f(x) \leq q(x) \tag{5b}$$

to hold for all $x \in \mathbb{R}$.

Claim: There is a continuous function $h\colon \mathbb{R} \to \mathbb{R}$ *which satisfies* (5a) *and* (5b) *(in place of* F*), even with strict inequalities.*

Let $x_0 \in \mathbb{R}$. If $f(x_0)g(x_0) < 0$, the inequalities (5) are either both upper or both lower bounds for $F(x_0)$. If $f(x_0) > 0$ and $g(x_0) > 0$, they say $F(x_0) \in [-\frac{p(x_0)}{g(x_0)}, \frac{q(x_0)}{f(x_0)}]$, which is an interval of positive length $1/f(x_0)g(x_0)$.

Assume $f(x_0) = 0$. Then $g(x_0) > 0$, and so $q(x_0) > 0$. Therefore it is easy to see that for every continuous solution F to (5a) there exists $\epsilon > 0$ such that F also satisfies (5b) for $|x - x_0| < \epsilon$. A similar statement holds if $g(x_0) = 0$, with the roles of (5a) and (5b) being reversed. From this discussion the claim follows easily.

By hypothesis (b) we can find $c > 0$ such that $f(x)g(x) < 0$ for $|x| \geq c$. Then for $x \geq c$ the two inequalities (5) are either both lower or both upper bounds for $F(x)$. Similarly for $x \leq -c$.

Now let h be a function as in the claim. Using Weierstraß approximation, we find a polynomial F_1 which approximates h very closely on $[-c, c]$. Then $F_1(x)$ satisfies the strict versions of inequalities (5) for all $x \in [-c, c]$. We want to find a second polynomial F_2 such that $|F_2|$ is very small on $[-c, c]$ (so small that $F := F_1 + F_2$ still satisfies (5) there), and such that $F := F_1 + F_2$ satisfies (5) on $|x| \geq c$. Indeed, F will then satisfy (5) globally, and the proposition will be proved.

The existence of F_2 follows easily from the following lemma, whose proof therefore completes the proof of Proposition 4.8:

4.9. Lemma. *Let* $r(x)$ *be a rational function without poles for* $|x| \geq 1$*, and with* $r(\pm 1) = 0$*. Given* $\epsilon > 0$*, there exists a polynomial* G *with* $|G(x)| < \epsilon$ *for* $|x| \leq 1$ *and with* $G(x) \geq |r(x)|$ *for* $|x| \geq 1$*.*

Proof. We have $r(x) = (x^2 - 1)\, q(x)$ where $q(x)$ has no poles for $|x| \geq 1$. There is a polynomial whose values for $|x| \geq 1$ are larger than those of $q(x)$. So we can assume that $q(x)$ is a polynomial. Evidently we can also assume that $q(x)$ is even. Let $c = \max_{i \geq 0} |q^{(i)}(1)|$, and choose n so large that $2n \geq \deg(q)$ and $n\epsilon > c$. Then $G(x) = c \cdot (x^2 - 1)x^{2n}$ has the required properties. □

Summing up, we get the following theorem, which for elliptic curves (over archimedean R) gives a complete answer to the question whether all psd elements are sums of squares:

4.10. Theorem. *Assume that the real closed field* R *is archimedean. Let* Y *be a smooth affine elliptic curve over* R*.*

a) *If* Y *has at least one complex point at infinity, then every positive semidefinite function in* $R[Y]$ *is a sum of squares.*
b) *If all points of* Y *at infinity are real, then there exist positive definite functions in* $R[Y]$ *which are not sums of squares.*

Proof. b) is contained in Theorem 3.2 and is only recorded here for completeness. To prove a), it suffices by Proposition 2.16 to prove the case where Y has only one complex and no real point at infinity. This is the situation that was considered above. □

There remains the question whether the assumption of R being archimedean was truly essential. We show that this is indeed so. Notice that this gives also a counter-example to Proposition 4.8 over non-archimedean R:

4.11. Theorem. *Assume that the real closed field R is not archimedean. Then, for a suitably chosen equation* (4) *of an affine elliptic curve Y over R (with $q(x)$ a monic quartic separable polynomial), there exists a psd function $f \in R[Y]$ which is not a sum of squares.*

Proof. Let $\varepsilon \in R$ be a positive element which is smaller than any positive rational number. Let Y be the curve with equation $y^2 + xh(x) = 0$, where

$$h(x) \;=\; \big(x - \frac{1}{\varepsilon^2}\big)\,(x^2 + \varepsilon x + \varepsilon^2).$$

The polynomial h satisfies $h(0) = -1$ and has only one real zero, at $x = \frac{1}{\varepsilon^2} > 0$. Therefore $f := x$ is psd on $Y(R)$. I claim that x is not sos in $R[Y]$. If it were, there would be psd polynomials P, $Q \in R[x]$ with $Px - Qh = 1$ (cf. proof of 4.6), i.e., the polynomial $x\,.\,(1 + Qh)$ would be psd. So Q would satisfy

$$Q(x) \geq -\frac{1}{h(x)} \text{ for } x < 0 \quad \text{and} \quad 0 \leq Q(x) \leq -\frac{1}{h(x)} \text{ for } 0 < x < \frac{1}{\varepsilon^2}.$$

In particular $Q(0) = 1$. It is easily checked that $h'(x) < 0$ on $[0, 1]$, which implies $0 \leq Q(x) \leq 1$ for $x \in [0, 1]$, and so $\|Q\| = 1$. (By $\|\cdot\|$ we denote the supremum norm on $[0, 1]$.) On the other hand, $h(\varepsilon) = 3\varepsilon^3 - 3 < -2$, and therefore $Q(\varepsilon) < \frac{1}{2}$. It follows that there exists $0 < x < \varepsilon$ with $Q'(x) < -\frac{1}{2\varepsilon}$, which is a negative number of infinitely large absolute value.

However, such a polynomial Q cannot exist. Indeed, it is easy to see that for every n there must be a real number $c_n > 0$ such that $\|q'\| \leq c_n\|q\|$ for every $q \in \mathbb{R}[x]$ of degree $\leq n$. In fact, Markov's inequality ([4], p. 233) says that one can take $c_n = 2n^2$ (and this is best possible). By Tarski's principle, the same holds over any real closed field, and with the same c_n. The polynomial Q from above would satisfy $\|Q\| = 1$ and $\|Q'\| \geq \frac{1}{2\epsilon}$, which is larger than any integer, contradiction. □

4.12. Remark. Even if R is not archimedean, there are still many elliptic curves (4) on which every psd polynomial is a sum of squares, as is obvious from the above discussion.

4.13. Remark. On singular "compact" curves it may well happen that psd functions are not sums of squares, even if the curve is rational. For an example let X be the projective closure of the affine curve $y^2 = x^3$, and let Y be the complement of the two (complex) zeros of $1 + x^2$ on X. Then $Y(\mathbb{R}) = X(\mathbb{R})$ is compact, and $f = \frac{x+x^2}{1+x^2}$ is a regular function on Y which is psd but not sos. Indeed, from $f = x + \frac{x^2-x^3}{1+x^2}$ one sees that f is not even sos in the local ring at the singularity.

5. Extending psd functions

In this section we prove that a psd regular function on a smooth affine curve can always be extended to a psd regular function on any ambient smooth affine variety (Theorem 5.6).

Let a be any ring, let $f \in A$, and let $\alpha \in \operatorname{Sper} A$. Recall that $\beta \in \operatorname{Sper} A$ is called a generalization of α if $\alpha \in \overline{\{\beta\}}$. We say that f is *locally psd around* α if the following two equivalent conditions hold:

(i) $f(\beta) \geq 0$ for every generalization β of α in $\operatorname{Sper} A$;
(ii) there is a neighborhood U of α in $\operatorname{Sper} A$ such that $f \geq 0$ on U.

(The asserted equivalence follows easily from a compactness argument.)

First let A be a regular local domain of dimension d, with maximal ideal $\mathfrak{m}$ and residue field $\kappa = A/\mathfrak{m}$. Let $0 \neq f \in A$ and $n = \mu(f)$, i.e., $f \in \mathfrak{m}^n \setminus \mathfrak{m}^{n+1}$. We study the connection between positivity properties of f and of the leading form $\ell(f)$. Fixing a regular system $x_1, \dots, x_d$ of parameters for A, $\ell(f)$ is a non-zero homogeneous polynomial of degree n in $\kappa[\xi_1, \dots, \xi_d]$, where $\xi_i = \ell(x_i)$ (cf. Sect. 1). Given an ordering α of κ, we say that $\ell(f)$ is *psd at* α if $\ell(f)$ is psd when considered as a homogeneous polynomial over the real closure κ_α of κ at α. We say that $\ell(f)$ is *pd at* α if $\ell(f)$ is psd at α and has no non-trivial zero in κ_α^d.

5.1. Lemma. *Let A be a regular local ring with maximal ideal $\mathfrak{m}$ and residue field $\kappa = A/\mathfrak{m}$. Let $\alpha \in \operatorname{Sper} \kappa$, and let $0 \neq f \in A$.*

a) *If f is locally psd around α, then the leading form $\ell(f)$ is psd at α.*
b) *Conversely, if $\ell(f)$ is pd at α, then f is locally psd around α. In fact, $f(\beta) > 0$ for every proper generalization β of α.*

If $\ell(f)$ is only psd (but not pd) at α, then f need not be locally psd around α. For example, take $f = y^2 - x^3$ around the origin.

Proof. We fix a regular system of parameters $x_1, \dots, x_d$ and consider $\lambda := \ell(f)$ as a homogeneous polynomial of degree $n = \mu(f)$ in $\kappa[\xi_1, \dots, \xi_d]$, where $\xi_i = \ell(x_i)$.

First assume that f is locally psd around α, and suppose there are $a_1, \dots, a_d \in \kappa_\alpha$ with $\lambda(a_1, \dots, a_d) < 0$. Define the homomorphism $\varphi\colon A \to \kappa_\alpha[\![t]\!]$ to be the composition of $A \to \widehat{A}$ with the κ-homomorphism

$$\widehat{A} = \kappa[\![x_1, \dots, x_d]\!] \to \kappa_\alpha[\![t]\!], \quad x_i \mapsto a_i t.$$

Then $\varphi(f) = \lambda(a_1, \dots, a_d)\, t^n +$ (terms of higher order). So at least one of the two orderings of $\kappa_\alpha(\!(t)\!)$ makes $\varphi(f)$ negative, say ω. Clearly $\beta := \varphi^*(\omega)$ is a generalization of α and has $f(\beta) < 0$, which contradicts the hypothesis.

Conversely, suppose that λ is pd at α. We have to show the following. Let $\rho\colon A \to R$ be a homomorphism into a real closed field R such that $\rho(u) > 0$ for every $u \in A^*$ with $u(\alpha) > 0$. If moreover $\ker(\rho) \neq \mathfrak{m}$, then $\rho(f) > 0$.

There is a convex subring $B \neq R$ of R with $\rho(A) \subset B$ and $\rho^{-1}(\mathfrak{m}_B) = \mathfrak{m}$, where $\mathfrak{m}_B$ is the maximal ideal of B ([19], p. 132). Let v be the valuation of R associated with B, with value group Γ. We have $v(\rho x_i) > 0$, and $v(\rho x_i) \neq \infty$ for at least one i, since $\ker(\rho) \neq \mathfrak{m}$. We can arrange the x_i such that $v(\rho x_1) = \cdots = v(\rho x_e) = \gamma$ and $v(\rho x_i) > \gamma$ for $i = e+1, \dots, d$, where $1 \leq e \leq d$ and $0 < \gamma \in \Gamma$. Clearly $v(\rho a) \geq j\gamma$ for every $a \in \mathfrak{m}^j$, $j \geq 0$.

The form λ has even degree $n = 2m = \mu(f)$. By choosing representatives in A for the coefficients of λ, we find a homogeneous polynomial $L = L(\xi_1, \dots, \xi_d)$ of degree n with coefficients in A whose reduction mod $\mathfrak{m}$ is λ. Define elements f_0, f_1, $f_2 \in A$ by

$$f_0 = L(x_1, \dots, x_e, 0, \dots, 0), \quad f_1 = L(x_1, \dots, x_d) - f_0$$

and

$$f_2 = f - L(x_1, \dots, x_d) = f - (f_0 + f_1).$$

Since λ is pd at α, it is also pd considered as a form over the residue field of B (which is real closed and contains (κ, α) as an ordered subfield). This implies that $\rho f_0 = (\rho x_1)^n \cdot u$ with u a positive unit of B. In particular, $\rho f_0 > 0$ and $v(\rho f_0) = n\gamma$. Moreover clearly $v(\rho f_1) > n\gamma$, since each monomial occurring in f_1

involves one of $x_{e+1}, \dots, x_d$. And $v(\rho f_2) \geq (n+1)\gamma$ since $f_2 \in \mathfrak{m}^{n+1}$. Therefore $v(\rho f_0) < v\big(\rho(f - f_0)\big)$, which implies $\rho f > 0$ as asserted. □

Now let A be any ring and I an ideal of A. We address the following question: When can every psd element of A/I be lifted to a psd element of A? In other words, when is the map $A_+ \to (A/I)_+$ surjective?

5.2. Remarks.

1. Every sum of squares in A/I can be lifted to a sum of squares in A, and in particular, lies in the image of $A_+ \to (A/I)_+$. Therefore $A_+ \to (A/I)_+$ is surjective whenever $(A/I)_+ = \Sigma(A/I)^2$.

2. In general, $A_+ \to (A/I)_+$ fails to be surjective. For example, if $A = \mathbb{R}[x, y]$ and $I = (y^2 - x^3)$, then $f = x$ is psd on the curve $y^2 = x^3$, but cannot be lifted modulo I to a psd polynomial in $\mathbb{R}[x, y]$, not even locally around the origin. This is obvious from Lemma 5.1a).

We introduce some terminology. Given a subset M of A, write $Z(M) = Z_A(M)$ for the closed subset $\{\alpha\colon\ f(\alpha) = 0 \text{ for every } f \in M\}$ of $\operatorname{Sper} A$. So $Z(M)$ can be identified with $\operatorname{Sper}(A/I)$ where I is the ideal generated by M. An element $f \in A$ will be called *psd around* $Z(M)$ if $f(\alpha) \geq 0$ for every $\alpha \in \operatorname{Sper} A$ with $\overline{\{\alpha\}} \cap Z(M) \neq \emptyset$.

5.3. Lemma. *Let A be a ring and I an ideal of A. Let $f \in A$ be such that $\bar{f} = f + I$ is psd in A/I. Assume that for every $\alpha \in Z(I + (f))$ there is some $h_\alpha \in I \cap A_+$ such that $f + h_\alpha$ is locally psd around α. Then there is $h \in I$ with $f + h \in A_+$.*

This sufficient condition for $\bar{f}$ to have a psd lift is clearly not necessary. For example, take $A = \mathbb{R}[x, y]$, $I = (x)$ and $f = x + y^2$. There is no *psd* polynomial $h \in I$ which would make $f + h$ psd around the origin. Still it is obvious that f can be changed mod I into a psd polynomial (e.g., into y^2).

Proof. It suffices to prove that for any $\alpha \in \operatorname{Sper} A$ there is $h_\alpha \in I \cap A_+$ with $(f + h_\alpha)(\alpha) \geq 0$. Indeed, the subsets $\{f + h_\alpha \geq 0\}$ of $\operatorname{Sper} A$ being constructible, the compactness of $\operatorname{Sper} A$ under the constructible topology implies that there are finitely many points $\alpha_1, \dots, \alpha_n \in \operatorname{Sper} A$ such that $\bigcup_{i=1}^n \{f + h_{\alpha_i} \geq 0\} = \operatorname{Sper} A$. Putting $h := \sum_{i=1}^n h_{\alpha_i}$, we then have $h \in I$ and $f + h \geq 0$ on $\operatorname{Sper} A$.

If α has a specialization β in $Z(I)$, we are done by the hypothesis (note that we can take $h_\alpha = 0$ if $f(\beta) \neq 0$). So we may assume $\overline{\{\alpha\}} \cap Z(I) = \emptyset$. This means that the ideal $I + \operatorname{supp}(\alpha)$ is not contained in any α-convex ideal of A ([19], p. 130). Therefore there is $g \in I$ with $g(\alpha) \geq 1$. Put $h_\alpha := (1 + f^2) \cdot g^2$. Then $h_\alpha \in I \cap A_+$, and $h_\alpha(\alpha) > |f(\alpha)|$, so this h_α does what we want. □

5.4. Corollary. *For every $f \in A$ which is psd around $Z(I)$, there is $h \in I$ with $f + h \in A_+$.* □

5.5. Corollary. *For every $f \in A$ which is strictly positive on $Z(I)$, there is $h \in I$ such that $f + h$ is strictly positive on* $\operatorname{Sper} A$.

Proof. The construction in the proof of Lemma 5.3 gives for every $\alpha \in \operatorname{Sper} A$ an element $h_\alpha \in I \cap A_+$ for which $(f + h_\alpha)(\alpha) > 0$. Conclusion as in the proof of Lemma 5.3. □

A weak version of Corollary 5.4 has been proved earlier by Gondard and Ribenboim. Namely, in the case where $A = R[x_1, \dots, x_n]$ is the polynomial ring over a

real closed field, they proved 5.4 under the additional assumption that the subset $\{f < 0\}$ of R^n is bounded ([13], Thm. 2).

We'll show now that every psd function on a smooth curve can be extended to a psd function on any ambient smooth variety. More generally:

5.6. Theorem. *Let A be a regular noetherian ring and $\mathfrak{p}$ a prime ideal of A such that $A/\mathfrak{p}$ is regular of dimension at most one. Then $A_+ \to (A/\mathfrak{p})_+$ is surjective.*

Proof. The proof is obvious if $\mathfrak{p}$ is a maximal ideal, cf. the first remark in 5.2. So we assume $\dim(A/\mathfrak{p}) = 1$. Let $f \in A$, $f \notin \mathfrak{p}$, such that $\bar{f} = f + \mathfrak{p}$ is psd in $A/\mathfrak{p}$. Let $\mathfrak{q}_1, \dots, \mathfrak{q}_s$ be those maximal ideals of A which contain $\mathfrak{p} + (f)$ and have a formally real residue field. The strategy of the proof will be as follows: For each $j = 1, \dots, s$ we'll find an integer $m_j \geq 0$ and an element $h_j \in \mathfrak{p}$ such that, for every $h \in A$ with $h \equiv h_j \bmod \mathfrak{q}_j^{m_j}$, the element $f + h$ is psd around $Z(\mathfrak{q}_j)$. Then there will be $h \in \mathfrak{p}$ with $h \equiv h_j \bmod \mathfrak{q}_j^{m_j}$ for $j = 1, \dots, s$ (Chinese Remainder Theorem; the usual proof shows that h can indeed be chosen in $\mathfrak{p}$). By construction, therefore, $f + h$ is psd around $Z(\mathfrak{p} + (f))$, and hence around $Z(\mathfrak{p})$. An application of Corollary 5.4 completes the proof.

So let $\mathfrak{q}$ be a maximal ideal containing $\mathfrak{p} + (f)$ whose residue field $k = A/\mathfrak{q}$ is formally real. Let d be the height of $\mathfrak{q}$. We can find a sequence $x_1, \dots, x_d$ of elements of $\mathfrak{q}$ whose images in $A_\mathfrak{q}$ form a regular system of parameters for $A_\mathfrak{q}$, and such that $\mathfrak{p}A_\mathfrak{q}$ is generated by $x_2, \dots, x_d$ ([21], Thm. 36, 4), p. 121). Hence $\bar{x}_1$ is a local uniformizer at $\mathfrak{q}/\mathfrak{p}$ in the Dedekind domain $A/\mathfrak{p}$, and $x_2, \dots, x_d \in \mathfrak{p}$.

Since f is psd on $A/\mathfrak{p}$ and k is real, the leading term of $\bar{f}$ in the discrete valuation ring $A_\mathfrak{q}/\mathfrak{p}A_\mathfrak{q}$ must have the form $\bar{s} \cdot \bar{x}_1^{2n}$ with $n \geq 1$ and $\bar{s} \in k^*$ a sum of squares. Consider the isomorphism

$$A/\mathfrak{q}^{2n+1} \xrightarrow{\sim} A_\mathfrak{q}/(\mathfrak{q}A_\mathfrak{q})^{2n+1} \cong k[x_1, \dots, x_d]/(x_1, \dots, x_d)^{2n+1}. \tag{6}$$

(The left-hand map is indeed an isomorphism since the ideal $\mathfrak{q}$ is maximal.) Under (6), the class of f in $A/\mathfrak{q}^{2n+1}$ corresponds therefore to the class of a polynomial $\bar{s} \cdot x_1^{2n} + p(x_1, \dots, x_d)$, where each monomial occurring in p is divisible by one of $x_2, \dots, x_d$. Thus there is an element $g \in \mathfrak{p}$ with $f + g \equiv \bar{s} \cdot x_1^{2n}$ modulo $\mathfrak{q}^{2n+1}$.

Put $h := g + x_2^{2n} + \cdots + x_d^{2n}$, an element of $\mathfrak{p}$. If $h' \in A$ with $h' \equiv h \bmod \mathfrak{q}^{2n+1}$, then $f + h'$ is modulo $\mathfrak{q}^{2n+1}$ congruent to $\bar{s} \cdot x_1^{2n} + x_2^{2n} + \cdots + x_d^{2n}$. So the leading form of $f + h'$ in $A_\mathfrak{q}$ is visibly positive definite with respect to every ordering of k. By Lemma 5.1b), $f + h'$ is therefore psd around $Z(\mathfrak{q})$. □

6. Surfaces and higher-dimensional varieties

Let R be a real closed field and V an affine algebraic variety over R of dimension $n \geq 2$. We ask the same question as for curves: Is every psd regular function on V a sum of squares of regular functions? For example, if V is the affine n-space $\mathbb{A}^n$, the answer is no. This was proved by Hilbert in 1888 [15].

Given any V as above, suppose we want to prove the analogue of Hilbert's theorem, i.e., show that there are psd elements in $R[V]$ which are not sums of squares. How could we possibly proceed?

The easiest line of argumentation would be to use local obstructions, provided there are any. As we have seen in Sect. 1, such obstructions do indeed exist if $n = \dim(V) \geq 3$. This allows us to find a psd, non-sos function on any connected such V with $V(R) \neq \emptyset$ (see 6.2).

In the case of smooth surfaces, however, this method is doomed to failure since we do not know a single example of a two-dimensional regular local ring A for which $A_+ \neq \Sigma A^2$. (See the discussion in Sect. 1.) Instead we will try to reduce the question to suitable smooth curves on V and use the extension theorem, Theorem 5.6, for psd functions. See Theorem 6.4 below.

We start with $\dim(V) \geq 3$. For the notion of a preorder P and its associated basic closed set $\mathcal{S}(P)$, see Sect. 3. As indicated above, our reasoning will essentially be of local nature:

6.1. Proposition. *Let V be an affine R-variety, and let P be a finitely generated preorder in $R[V]$ such that the semi-algebraic set $\mathcal{S}(P)$ has dimension ≥ 3. Then there exists a psd polynomial in $R[V]$ which is not contained in P.*

Proof. Put $A := R[V]$. Let $f_1, \ldots, f_r \in A$ be generators of P, let $S = \mathcal{S}(P) = \{f_1 \geq 0, \ldots, f_r \geq 0\}$ and assume $\dim(S) \geq 3$. Let Z be an irreducible component of the Zariski closure of S in V with $\dim(Z) \geq 3$, and let $\mathfrak{p} \subset A$ be the prime ideal of Z. After relabelling, let $f_1, \ldots, f_s$ be those f_i with $f_i \notin \mathfrak{p}$ (here $0 \leq s \leq r$). There is a point $Q \in S \cap Z(R)$ which is regular on Z and for which $f_1(Q) > 0, \ldots, f_s(Q) > 0$. Indeed, $S \cap Z(R)$ would otherwise be contained in $Z_{\mathrm{sing}} \cup \{f_1 \cdots f_s = 0\}$, which is a proper Zariski closed subset of Z, contradicting that $S \cap Z(R)$ is Zariski dense in Z.

The local ring of Q on Z is $B := A_{\mathfrak{q}}/\mathfrak{p}A_{\mathfrak{q}}$ ($\mathfrak{q} :=$ maximal ideal of Q), and is a regular local ring of dimension $d := \dim(Z) \geq 3$. There is a sequence $x_1, \ldots, x_d$ of elements in A such that the images $\bar{x}_1, \ldots, \bar{x}_d$ of the x_j in B form a regular sequence of parameters for B. Let $h(t_1, \ldots, t_d)$ be a psd form over R which is not a sum of squares (see remark before 1.2), and let $f := h(x_1, \ldots, x_d) \in A$. Then $f \in A_+$ as in the proof of 1.2. I claim that $f \notin P$. Indeed, assume there is an expression $f = \sum_i s_i \cdot f_1^{i_1} \cdots f_r^{i_r}$ with $s_i \in \Sigma A^2$, where i is ranging over $\{0,1\}^r$. Read this identity in $\widehat{B}$. For $i \leq s$, f_i is a square in $\widehat{B}$ since $f_i(Q) > 0$. The remaining f_i are zero in $\widehat{B}$. Therefore the image $\bar{f}$ of f in $\widehat{B}$ is a sum of squares. But the leading form $\ell(\bar{f})$ in B or in $\widehat{B}$ is h, and hence is not a sum of squares, a contradiction to Lemma 1.1. □

Note that no smoothness assumption was needed in Proposition 6.1. Some hypothesis on $\mathcal{S}(P)$ is, however, necessary to guarantee that $\mathcal{S}(P)$ is not too small. For example, for any point $M \in V(R)$ the preorder $\mathcal{P}(M) = \{f \in R[V] \colon\ f(M) \geq 0\}$ is finitely generated.

If we ask only for the existence of psd, non-sos functions, we get the positive answer even for any connected variety of dimension ≥ 3 with an R-point. More generally, we can prove a version valid over any base field:

6.2. Theorem. *Let k be a field and A a connected k-algebra of finite type. Suppose that $\dim A \geq 3$ and $\operatorname{Sper} A \neq \emptyset$. Then there is a psd element in A which is not a sum of squares.*

Proof. We say that a ring A is *real reduced* if the identity $a_1^2 + \cdots + a_n^2 = 0$ in A implies $a_1 = \cdots = a_n = 0$. It is equivalent that A is reduced and every minimal prime ideal of A has a formally real residue field. (See [19], III §2, where such rings have been called real.)

In the situation of the theorem, if A is real reduced, then it satisfies the hypotheses of Corollary 1.3 (and so we are done). Indeed, there is $s \in A$ such that

A_s is a regular domain with formally real quotient field and $\dim(A_s) \geq 3$; by the Artin-Lang theorem ([3], p. 76), A_s has (plenty of) maximal ideals with formally real residue field.

In the remaining cases we can apply the following general fact:

6.3. Lemma. *Let A be a connected noetherian ring with* $\operatorname{Sper} A \neq \emptyset$*, and suppose that A is not real reduced. Then $A_+ \neq \Sigma A^2$; in fact, there is $f \in A$ with $f \equiv 0$ on* $\operatorname{Sper} A$ *but $f \notin \Sigma A^2$.*

First note that $I^2 \neq I$ holds for any ideal $(0) \neq I \neq (1)$ of A. Indeed, if $I^2 = I$, then the Nakayama lemma says that there is $a \in I$ with $(1-a)I = 0$ ([21], p. 8). So $a^2 = a$, and A connected implies $a = 0$ or $a = 1$, i.e., $I = (0)$ or $I = (1)$.

Now let $I = \sqrt[\mathrm{re}]{(0)}$, the so-called real nilradical of A. By definition, I is the intersection of all prime ideals whose residue field is formally real ([19], *loc.cit.*). By hypothesis, $(0) \neq I \neq (1)$, and so $I^2 \neq I$. Any $f \in I \setminus I^2$ has the property required in the lemma, since A/I is real reduced. □

We will finally try to study the case of smooth surfaces over R. If V is one, then by a curve on V we mean an effective divisor on V. By a compactification $\overline{V}$ of V we mean a complete R-scheme $\overline{V}$ of finite type which contains V as an open dense subscheme. An irreducible curve over R is called *real* if it has infinitely many R-rational points, or equivalently, if its function field is formally real.

6.4. Theorem. *Let V be a smooth affine surface over R which has a smooth compactification $\overline{V}$ for which every irreducible curve contained in $\overline{V} \setminus V$ is real. Then the preorder of all psd elements in $R[V]$ is not finitely generated. In particular, $R[V]$ contains psd elements which are not sums of squares.*

Proof. Let $C_1, \ldots, C_r$ be the irreducible components of $\overline{V} \setminus V$. (The C_i are curves since V is affine.) We will show that there exists a smooth irreducible non-rational curve C on $\overline{V}$ such that $C \cap C_i$ consists of real points only, for every $i = 1, \ldots, r$. This will suffice: Suppose that the preorder of psd elements in $R[V]$ is generated by $f_1, \ldots, f_n$. Let $C' = C \cap V$. By Theorem 3.4, there exists $\bar{g} \in R[C']$ which is psd on C' but not contained in the preorder of $R[C']$ generated by the restrictions $f_i|_{C'}$ of the f_i to C'. By 5.6 we can find a psd function $f \in R[V]$ with $f|_{C'} = \bar{g}$. It is clear that f cannot lie in the preorder of $R[V]$ generated by the f_i, contradiction.

6.5. Lemma. *Let D_0 be a very ample effective divisor on $\overline{V}$. There is an integer N such that for every $n > N$, there exists $D \in |nD_0|$ such that D intersects each C_i in real smooth points of C_i, and all intersection points are transversal.*

Using this lemma we can complete the proof of Theorem 6.4, as follows. Since $\overline{V}$ is projective, the lemma gives us a very ample effective divisor D on $\overline{V}$ which intersects each C_i transversally in real smooth points of C_i. The elements in $|D|$ with the same property form an open subset of $|D|$ (with respect to the semi-algebraic topology). So a Bertini argument ([14], ch. V, Lemma 1.2) implies that we can even find such D which is irreducible and smooth. Moreover, if we replace D by nD for suitable $n \geq 1$ (and apply Lemma 6.5 plus Bertini again), we can make the genus g_D of the smooth irreducible curve D arbitrarily large (and in particular, make D non-rational), by the adjunction formula $2g_D - 2 = D \,.\, (D + K)$ and since $D^2 > 0$.

So it remains to prove Lemma 6.5. First of all, after replacing D_0 with a linearly equivalent effective divisor D we can assume that $C_i \not\subset \operatorname{supp}(D)$ and $\operatorname{supp}(D) \cap C_i$ is contained in the smooth locus of C_i, for each i. Fix one of the C_i, and denote by $D \cap C_i$ the intersection divisor; this is an effective divisor on $(C_i)_{\mathrm{reg}}$ whose degree is $D \,.\, C_i > 0$. Now there is $N_i > 0$ such that for every $n > N_i$ there exists a rational function $f_i \neq 0$ on C_i such that the divisor $n(D \cap C_i) + \operatorname{div}_{C_i}(f_i)$ on C_i has the form $P_1 + \cdots + P_s$ where $P_1, \ldots, P_s$ are distinct smooth real points on C_i. Indeed, this follows from applying Corollary 2.10 to the normalization of C_i and using also Remark 2.14. Given $n > N := \max_i N_i$, let (for each i) f_i be a rational function on C_i with the just stated property. There is a rational function $0 \neq f$ on $\overline{V}$ without zeros or poles along any of the C_i, which restricts to f_i on C_i for each i. The divisor $nD + \operatorname{div}_{\overline{V}}(f)$ has the required properties. □

6.6. Remarks.

1. In the situation of 6.4, there even exists $f \in R[V]$ which is strictly positive on $V(R)$ but not a sum of squares. Indeed, by Theorem 3.2 one finds $\bar{g} \in R[C']$ (cf. the first part of the proof of 6.4) which is strictly positive on $C'(R)$ but not sos, and can then apply Corollary 5.5.

2. Let $R = \mathbb{R}$. Our reasoning through restriction to suitable curves works fine for many surfaces, thereby generalizing Hilbert's theorem. Still there are many surfaces left for which it doesn't tell us anything. In particular, this is so for all surfaces V for which $V(\mathbb{R})$ is compact (and also for every affine Zariski-open subset of such V). The reason is that we have no example of a psd, non-sos function on a smooth curve whose set of real points is compact, cf. Sect. 4.

6.7. Remark. At least over $R = \mathbb{R}$ we may generalize Theorem 6.4 to other preorders, using Theorem 3.5. We leave it to the reader to formulate suitable versions. For example, if P is a finitely generated preorder in $\mathbb{R}[x, y]$ such that $\mathcal{S}(P)$ contains some non-empty open cone in $\mathbb{R}^2$, then P cannot contain all psd polynomials.

6.8. Remark. The preceding remark gives already an idea of how results on curves can be useful for the study of psd polynomials. Recall that it is not a trivial matter to exhibit psd polynomials which are not sos; the first example which figured in the published literature was given by Motzkin in 1965 [24], almost 80 years after Hilbert had proved the existence of such examples. Although meanwhile many other classes of such polynomials have been found, there is still some interest in new constructions; see Reznick's survey [27]. Some of the results and methods of this paper can be used to produce a great variety of new examples. Indeed, in Sect. 3 we exhibited various ways in which one can construct psd, non-sos functions on smooth plane curves; to a good extent, these constructions can be made explicit in concrete cases. By the extension theorem, Theorem 5.6, all these functions extend to psd, non-sos polynomials in $R[x, y]$, even in many possible ways. Again, the proof of 5.6 is constructive enough to allow explicit applications.

We would finally like to remark that the idea of constructing examples of psd, non-sos polynomials $f(x, y)$ by extending suitable regular functions from plane curves to the affine plane was already applied successfully by Stengle in 1979 [31].

7. Some open problems

It will have become obvious that there is a large number of unsolved questions connected with the subject of this paper. Here we want to isolate three basic problems. They seem to be central for a better understanding of sums of squares on affine varieties in general, and we hope to stimulate some future work by posing them explicitly.

Problem 1. *Study regular local domains A of dimension two (containing $\frac{1}{2}$), with quotient field $K = \mathrm{Quot}(A)$, and decide whether $A \cap \Sigma K^2 = \Sigma A^2$ holds or not. In other words, is every psd element of such a ring a sum of squares?*

See Remarks 1.7 for a very short list of cases where this problem has a positive answer.

Problem 2. *Given a smooth affine algebraic curve Y over $\mathbb{R}$ for which $Y(\mathbb{R})$ is compact, is it true that every psd polynomial function $f \in \mathbb{R}[Y]$ on Y is a sum of squares?*

One possible approach could be to try to bound the complexity of sos representations of strictly positive functions (if one expects a positive answer). Every strictly positive $f \in \mathbb{R}[Y]$ is sos by Schmüdgen's theorem. Alternatively, one could try to proceed in the spirit of Sect. 4 of this paper. For example, it may be shown that the answer is positive if it is so for all psd f with at most one complex zero (assuming here that Y has only one point at infinity).

Problem 3. *Study affine smooth algebraic surfaces V over $\mathbb{R}$ with $V(\mathbb{R})$ compact, and decide whether every psd $f \in \mathbb{R}[V]$ is a sum of squares or not.*

It seems that not a single example of such a surface V (with $V(\mathbb{R}) \neq \emptyset$) is known for which one could either prove or disprove that every psd function is a sum of squares! To look at an example, let $f = f(x, y, z)$ be a positive definite form in $\mathbb{R}[x, y, z]$ and let V be the complement of the curve $f = 0$ in $\mathbb{P}^2$. Then V is a smooth affine surface, and $V(\mathbb{R}) = \mathbb{P}^2(\mathbb{R})$ is compact. It is easy to see that each psd element of $\mathbb{R}[V]$ is sos if and only if for each psd form $g(x, y, z)$ there exists $n \geq 0$ such that the form $f^{2n} \cdot g$ is a sum of squares. Whether or not a form f with this property exists is not known.

If the answer to Problem 1 is "no", and if one has counterexamples which are local rings of surfaces over $\mathbb{R}$, this would imply a negative answer to Problem 3 in general.

If the answer to Problem 2 is "no", this would imply a negative answer to Problem 3 in general, by an application of Theorem 5.6.

Added in proof (October 1998): Recent results provide answers to some of the questions above. In particular, the answer to Problem 2 is "yes". As to Problem 3, there do exist smooth affine surfaces V over $\mathbb{R}$ with $V(\mathbb{R})$ compact and non-empty, for which psd = sos holds in $\mathbb{R}[V]$. Details will be published elsewhere.

References

[1] E. Artin: Über die Zerlegung definiter Funktionen in Quadrate. Abh. Math. Sem. Univ. Hamburg **5**, 100-115 (1927). See: Coll. Papers, Addison-Wesley, Reading, MA, 1965, pp. 273-288. MR **31:**1159

[2] R. Baeza: *Quadratic Forms over Semilocal Rings*. Lect. Notes Math. **655**, Springer, Berlin, 1978. MR **58:**10972

[3] J. Bochnak, M. Coste, M.-F. Roy: *Géométrie Algébrique Réelle*. Erg. Math. Grenzgeb. (3) **12**, Springer, Berlin, 1987. MR **90b:**14030

[4] P. Borwein, T. Erdélyi: *Polynomials and Polynomial Inequalities*. Grad. Texts Math. **161**, Springer, New York, 1995. MR **97e:**41001

[5] M. D. Choi, Z. D. Dai, T. Y. Lam, B. Reznick: The Pythagoras number of some affine algebras and local algebras. J. reine angew. Math. **336**, 45-82 (1982). MR **84f:**12012

[6] M. D. Choi, T. Y. Lam: An old question of Hilbert. In: *Quadratic Forms* (Kingston 1976), G. Orzech (ed.), Queen's Papers Pure Appl. Math. **46**, Kingston, ON, 1977, pp. 385-405. MR **58:**16503

[7] M. D. Choi, T. Y. Lam, B. Reznick, A. Rosenberg: Sums of squares in some integral domains. J. Algebra **65**, 234-256 (1980). MR **81h:**10028

[8] J.-L. Colliot-Thélène, C. Scheiderer: Zero-cycles and cohomology on real algebraic varieties. Topology **35**, 533-559 (1996). MR **97a:**14009

[9] M. Coste, M.-F. Roy: La topologie du spectre réel. In: *Ordered Fields and Real Algebraic Geometry* (San Francisco 1981), D. W. Dubois, T. Recio (eds.), Contemp. Math. **8**, Providence, RI, 1982, pp. 27-59. MR **83m:**14017

[10] Ch. N. Delzell: A constructive, continuous solution to Hilbert's 17^{th} problem, and other results in semi-algebraic geometry. Ph. D. thesis, Stanford University, June 1980. Cf. also "Bad points for positive semidefinite polynomials: preliminary report", Abstracts of papers presented to the AMS **18**, # 926-12-174 (1997).

[11] Ch. N. Delzell: Kreisel's unwinding of Artin's proof. In: *Kreiseliana: About and Around Georg Kreisel,* P. Odifreddi (ed.), A. K. Peters, Wellesley, MA, 1996, pp. 113-246. CMP 97:08

[12] D. Gondard: Le 17ème problème de Hilbert et ses développements récents. Sém. Structures Algébriques Ordonnées, Univ. Paris VII, Vol. II, 21-49 (1990).

[13] D. Gondard, P. Ribenboim: Fonctions définies positives sur les variétés réelles. Bull. Sci. Math. (2) **98**, 39-47 (1974). MR **55:**5601

[14] R. Hartshorne: *Algebraic Geometry*. Grad. Texts Math. **52**, Springer, New York, 1977. MR **57:**3116

[15] D. Hilbert: Über die Darstellung definiter Formen als Summe von Formenquadraten. Math. Ann. **32**, 342-350 (1888). See: Ges. Abh., Bd. II, Springer, Berlin, 1933, pp. 154-161.

[16] D. Hilbert: Über ternäre definite Formen. Acta math. **17**, 169-197 (1893). See: Ges. Abh., Bd. II, Springer, Berlin, 1933, pp. 345-366.

[17] D. Hilbert: Mathematische Probleme. Arch. Math. Phys. (3) **1**, 44-63 and 213-237 (1901). See: Ges. Abh., Bd. III, Springer, Berlin, 1933, pp. 290-329.

[18] D. Hilbert: Hermann Minkowski. Gedächtnisrede, 1. Mai 1909. Math. Ann. **68**, 445-471 (1910). See: Ges. Abh., Bd. III, Springer, Berlin, 1933, pp. 339-364.

[19] M. Knebusch, C. Scheiderer: *Einführung in die reelle Algebra*. Vieweg, Braunschweig, 1989. MR **90m:**12005

[20] H. Lindel: Projektive Moduln über Polynomringen $A[T_1, \ldots, T_m]$ mit einem regulären Grundring A. Manuscr. math. **23**, 143-154 (1978). MR **57:**12597

[21] H. Matsumura: *Commutative Algebra*. Second edition. Benjamin, Reading, Mass., 1980. MR **82i:**13003

[22] J. S. Milne: Jacobian Varieties. In: *Arithmetic Geometry*, G. Cornell, J. H. Silverman (eds.), Springer, New York 1986, Chapter VII, pp. 167-212. MR **89b:**14029

[23] H. Minkowski: Untersuchungen über quadratische Formen. Bestimmung der Anzahl verschiedener Formen, welche ein gegebenes Genus enthält. Inauguraldissertation, Königsberg 1885; see Ges. Abh., Bd. I, Teubner, Leipzig, 1911, pp. 157-202.

[24] T. S. Motzkin: The arithmetic-geometric inequality. In: *Inequalities*, Proc. Symp. Wright-Patterson AFB 1965, O. Shisha (ed.), Academic Press, New York, 1967, pp. 205-224. MR **36:**6569

[25] A. Pfister: *Quadratic Forms with Applications to Algebraic Geometry and Topology*. London Math. Soc. Lecture Note Ser. **217**, Cambridge University Press, Cambridge, 1995. MR **97c:**11046

[26] V. Powers: Hilbert's 17th problem and the champagne problem. Am. Math. Monthly **103**, 879-887 (1996). MR **97m:**12008

[27] B. Reznick: Some concrete aspects of Hilbert's 17th problem. Preprint, see Sém. Structures Algébriques Ordonnées, Univ. Paris VII, 1996. Revised version to appear in Proc. RAGOS, Contemp. Math.
[28] C. Scheiderer: *Real and Étale Cohomology*. Lect. Notes Math. **1588**, Springer, Berlin, 1994. MR **96c:**14018
[29] C. Scheiderer: Classification of hermitian forms and semisimple groups over fields of virtual cohomological dimension one. Manuscr. math. **89**, 373-394 (1996). MR **97g:**20056
[30] K. Schmüdgen: The K-moment problem for compact semi-algebraic sets. Math. Ann. **289**, 203-206 (1991). MR **92b:**44011
[31] G. Stengle: Integral solution of Hilbert's seventeenth problem. Math. Ann. **246**, 33-39 (1979). MR **81c:**12035
[32] E. Witt: Zerlegung reeller algebraischer Funktionen in Quadrate. Schiefkörper über reellem Funktionenkörper. J. reine angew. Math. **171**, 4-11 (1934).
[33] Th. Wörmann: Positive polynomials on compact sets. To appear Manuscr. math.

Fachbereich Mathematik, Universität Duisburg, 47048 Duisburg, Germany
E-mail address: claus.@math.uni-duisburg.de