

Semialgebraic Topology Over a Real Closed Field I: Paths and Components in the Set of Rational Points of an Algebraic Variety

Hans Delfs and Manfred Knebusch

Fachbereich Mathematik der Universität, Universitätsstr. 31, D-8400 Regensburg, Federal Republic of Germany

Contents

§1.	Introduction; The Strong Topology										107
	Smooth Curves										
§3.	Combinatorial Paths										116
§4.	Path Components										120
§5.	An Application to the Theory of Witt Rings										124
	ferences										

§1. Introduction: The Strong Topology

We fix once and for all a real closed base field R. By a variety X over R we mean a separated algebraic scheme X over R. For all problems attacked here we could equally well assume that X is also reduced, since we are only interested in the set X(R) of rational points of X. Notice that for every closed point x of X the residue class field $\kappa(x) = \mathcal{O}_x/m_x$ either coincides with R, i.e. x is rational, or is isomorphic to $R(\sqrt{-1})$. We call the points x of X with $\kappa(x) = R$ the real points of X and the other closed points the complex points of X.

R is a topological field, a basis of open sets being given by the open intervals

$$]a, b[=\{x \in R \mid a < x < b\}$$

of R. (Recall that R has a unique ordering compatible with addition and multiplication.) Unfortunately the topological space R is totally disconnected except in the single case that R is isomorphic to the field \mathbb{R} of real numbers.

For any variety X over R the topology of R induces in a natural way a topology on the set X(R) of real points of X, which intrinsically can be described as follows: A subbasis of open sets is given by the sets

$$f^{-1}(]a, b[) = \{p \in U(R) | f(p) \in]a, b[\}$$

with U running through the affine open subsets of X and f through the affine Ring $R[U] = \Gamma(U, \mathcal{O}_X)$. We call this topology the strong topology on X(R). If X is embedded as a locally closed subscheme into some affine standard space $\mathbb{A}^N = \mathbb{A}_R^N$, then X(R) is a subset of $\mathbb{A}^N(R) = R^N$, and the strong topology on X(R) is just the topology as a subspace of the cartesian product R^N . Again, if $R \neq \mathbb{R}$ the space X(R) is clearly totally disconnected.

If Y is a second variety over R then we denote the algebraic variety $X \times_{\text{Spec } R} Y$ simply by $X \times Y$. The space $(X \times Y)(R) = X(R) \times Y(R)$ – with the strong topology as always – is the cartesian product of the topological spaces X(R) and Y(R). Since the diagonal Δ of $X \times X$ is Zariski closed in $X \times X$, the set $\Delta(R)$ is closed in $X(R) \times X(R)$. But $\Delta(R)$ is just the diagonal of $X(R) \times X(R)$. Thus, X(R)is a Hausdorff space.

By a morphism $\varphi: X \to Y$ between varieties X and Y over R we mean of course a morphism in the category of schemes over R. Clearly, any such morphism induces a continuous map from X(R) to Y(R) which we denote by φ_R .

It is desirable to establish for the spaces X(R) a theory of homology groups $H_r(X(R), \Lambda)$ with coefficients in some abelian group Λ , cohomology groups $H^r(X(R), F)$ with suitable sheaves F, homotopy groups $\pi_k(X(R), x_0)(k \ge 1)$ resp. homotopy sets (k=0), which in the case $R=\mathbb{R}$ coincides with the classical theory. More generally one would like to do this for any semialgebraic subset M of X(R) instead of X(R) itself. We mean by such a set a subset of X(R) which Zariski-locally is a finite union of sets defined by finitely many inequalities $f(x) > 0, f(x) \ge 0$ with algebraic functions f, cf. §6 in Part II.

In the present paper this program will be carried through to some extent for the functors π_0 and H_0 . In the first part we consider the spaces X(R) and in the second part more generally "semialgebraic spaces". The main difference between the parts lies in the methods used. In Part I we remain in the cadre of general algebraic geometry, while our study in Part II will have the typical "semialgebraic" flavour. We introduce in Part I "paths" in X(R) and prove that X(R)consists of only finitely many path components. These are open and closed in the strong topology. A striking application of this result is the theorem, proved in §5, that for an algebraic scheme S over any field every signature $\sigma: W(S) \rightarrow \mathbb{Z}$ (i.e. \mathbb{Z} -valued ring homomorphism on the Witt ring W(S)) factors through a closed point of S.

In the second part of the paper we shall generalize the theory of paths and components to semialgebraic spaces. We shall also obtain refinements of some results in Part I. In particular we shall prove that the path components of X(R) are semialgebraic subsets of X(R).

G.W. Brumfiel starts in his book [2] a program containing a subprogram similar to the one outlined above. His methods are less naive and more algebraic than ours. His main objective is to understand and to describe the geometry of a semialgebraic set in terms of the ring of algebraic functions on M equipped with partial orderings induced by M. His methods are in some sense similar to the use of commutative algebra in the geometry of affine varieties over an algebraically closed base field.

Semialgebraic Topology Over a Real Closed Field I.

We believe that it pays well to divide the semialgebraic geometry over a real closed field into a "topological part", where only a substitute of the continuous mappings in the classical theory is used, and a "function theoretic part". The substitute of the continuous mappings will be our semialgebraic mappings defined in §6 and §7. This leads to the category of "semialgebraic spaces" over R introduced in §7. The "good" functions are to our opinion the "Nash functions" – properly defined. Thus one has to proceed in the function theoretic part to the category of "Nash spaces" over R. This category is more narrow than the category of semialgebraic spaces but still much coarser and hence more flexible than the category of varieties over R. Nash spaces will not occur in the present paper except perhaps in the very last sections.

We also mention the work of M.F. Coste-Roy and M. Coste on real algebraic geometry, cf. e.g. [5]. Again their program is much more ambitious than ours. They develop a general theory of "real spectra" of arbitrary commutative rings. Semialgebraic sets over real closed fields are only a rather special example in their theory.

We hope that at the end of Part II of our paper the reader will be amply convinced that the varieties and their semialgebraic subsets over a real closed field deserve well to be treated in an autonomous theory with a beauty of its own which probably cannot survive in a "general" real algebraic geometry.

§2. Smooth Curves

We now recall the main results of [10, \$1-\$8] with slight refinements from Geyer's paper [8]. In the case $R=\mathbb{R}$ these results are essentially due to Witt [17]. They yield the base of our theory of paths.

Assume that X is a smooth connected curve over R, i.e. an irreducible regular variety of dimension 1 over R. Assume further that the set X(R) of real points is not empty, hence infinite. We denote the field R(X) of rational functions on X by F. Every f in F will be identified with the induced continuous map

$$f: X(R) \to \mathbb{P}^1(R) = R \cup \infty.$$

A function $f \in F^*$ is called *definite* at a point $P \in X(R)$ if $\operatorname{ord}_p(f)$ is even. Here $\operatorname{ord}_p(f)$ means the order of f at P, positive integral if f(P)=0, negative integral if $f(P)=\infty$, and zero if $f(P) \neq 0, \infty$. If f is definite at P, then f has a well defined sign $\tau_p(f) = \pm 1$ at P. Indeed, we have a decomposition $f = \pi^{2r}g$ with some local uniformizing parameter π of X at P and some $g \in F$ with $g(P) \neq 0, \infty$, and we define $\tau_p(f)$ as the sign of $g(P), \tau_p(f) = +1$ if g(p) > 0, and $\tau_p(f) = -1$ if g(p) < 0. If $\operatorname{ord}_p(f)$ is odd we put $\tau_p(f) = 0$. We call f definite (positive definite, negative definite) on some subset A of X(R) if $\tau_p(f) \neq 0$ for all P in A (resp. $\tau_p(f) = +1$, resp. $\tau_p(f) = -1$ for all P in A).

Definition 1. Two points P and Q of X(R) are called *equivalent in* X if $\tau_P(f) = \tau_Q(f)$ for all $f \in F^*$ which are definite on X(R). The equivalence classes are called the *components* of X(R).

As has been proved in [10], X(R) consists of finitely many components $\Gamma_1, \ldots, \Gamma_r$. They are closed and open in X(R) (in the strong topology, as always). Moreover for any component Γ_i there exists a function $f \in F^*$ without poles and zeros on X(R) which is negative definite on Γ_i and positive definite on $X(R) \setminus \Gamma_i$ [10, p. 56].

According to [10] and [8] there exist rational differential forms $\omega \in \Omega^1(F/R)$ which have neither poles nor zeros on X(R). For any two such forms ω_1, ω_2 we have an equation $\omega_2 = f\omega_1$ with some function $f \in F^*$ which has neither a zero nor a pole on X(R).

Definition 2. We call ω_1 and ω_2 equivalent if f(P) > 0 for all $P \in X(R)$. An equivalence class of differentials ω without poles and zeros on X(R) is called an *orientation* of X(R). In the same way we define an orientation of a component Γ_i of X(R) focussing on differentials without poles and zeros on Γ_i .

Every component Γ_i has precisely two different orientations and X(R) has precisely 2^r different orientations.

We choose once and for all an orientation $[\omega_0]$ of X(R). We define for any function f in F^* and any point $P \in X(R)$ a number $\partial_p(f) \in \{1, -1, 0\}$ by

$$\partial_p(f) = \begin{cases} \tau_p(df/\omega_0) & \text{if } \operatorname{ord}_p f \text{ } \operatorname{odd} \\ 0 & \text{if } \operatorname{ord}_p f \text{ } \operatorname{even.} \end{cases}$$

Here the differential quotient df/ω_0 means the function $g \in F^*$ with $df = g \omega_0$. Notice that if $\operatorname{ord}_p(f)$ is even but not zero then again the equality

$$\partial_{p}(f) = \tau_{p}(df/\omega_{0})$$

holds true.

For a given function f in F^* we have $\partial_p(f) \neq 0$ only for finitely many points P of X(R). Intuitively $\partial_p(f) = +1$ if f "changes sign from -1 to +1" and $\partial_p(f) = -1$ if f "changes sign from +1 to -1" at P.

We shall use the following rules which are easily verified: Let f and g be functions in F^* . If $\tau_p(g) \neq 0$ then

$$\partial_p(fg) = \partial_p(f) \tau_p(g)$$

If $\tau_p(g) = 0$ then

$$\partial_{\mathbf{p}}(f\mathbf{g}) = \tau_{\mathbf{p}}(f) \partial_{\mathbf{p}}(\mathbf{g}).$$

Assume now that X is also complete. Then we have for every f in F^* and every component Γ of X(R)

$$\sum_{\boldsymbol{P}\in\boldsymbol{\Gamma}}\partial_{\boldsymbol{p}}(f)=0$$

cf. [10, §5]. Moreover given on each component an even number of points there exists a function f in F^* which has simple zeros precisely at these points and no other zeros and poles on X(R) [10, p. 61], [8].

Let now P_1 and P_2 be different points of a component Γ of X(R). There clearly exists a function $f \in F^*$ which has a simple zero at P_1 with $\partial_{P_1}(f) = -1$

and a simple zero at P_2 with $\partial_{P_2}(f) = +1$ and no other zeros and poles on X(R) and which is positive definite on $X(R) \setminus \Gamma$.

Definition 3. The open interval $]P_1, P_2[$ is the set of all points P in X(R) with f(P) < 0.

We call f an interval function for $]P_1, P_2[$. Clearly the set $]P_1, P_2[$ is contained in Γ and does not depend on the choice of the interval function. It moreover depends only on the given orientation of Γ , not on the orientation of the other components. We define the closed interval $[P_1, P_2]$ as the union $]P_1, P_2[\cup \{P_1\} \cup \{P_2\}, i.e.$

$$[P_1, P_2] = \{P \in X(R) \mid f(P) \leq 0\},\$$

and define similarly the "half-open" intervals $[P_1, P_2[$ and $]P_1, P_2]$. Notice that $]P_1, P_2[$ is indeed open and $[P_1, P_2]$ is closed in the strong topology. Since -f is an interval function for $]P_2, P_1[$, up to a factor which is positive definite on Γ , we see that Γ is the disjoint union of the four sets $]P_1, P_2[$, $]P_2, P_1[$, $\{P_1\}, \{P_2\}$.

Omitting an arbitrary point P_0 from the component Γ we have a total ordering of $\Gamma \setminus \{P_0\}$, defined as follows (cf. [10, §6]):

$$P \leq Q \Leftrightarrow]P_0, P[\subset]P_0, Q].$$

With respect to this ordering we have for points $P_1 < P_2$ in $\Gamma \setminus \{P_0\}$:

$$]P_1, P_2[=\{P \in \Gamma \smallsetminus \{P_0\} | P_1 < P < P_2\}.$$

In particular, we have a total ordering on any closed interval [P, Q] in Γ with smallest element P and largest element Q. Of course, this ordering does not depend on the choice of the point P_0 in $\Gamma \setminus [P, Q]$ but only on the orientation of Γ . Altogether we are thus justified to argue with the intervals defined here as with the usual intervals on an oriented circle line S^1 , as long as we consider only finitely many intervals.

The open intervals are a basis of the strong topology, cf. [10, §7]. More precisely, every set $f^{-1}(]a, b[)$ in X(R) with a < b in R and $f \in F$ is a disjoint union of finitely many "generalized open intervals" [loc.cit., Th. 7.2.i].¹ The same holds true more generally for any set $f^{-1}(]a, b[)$ with a, b different points in $\mathbb{P}^1(R)$ with standard orientation. Here we mean by a generalized open interval an open interval or a whole component Γ or a set $\Gamma \setminus \{P\}$ obtained from Γ by omission of one point P.

If Y is a Zariski open subset of X, i.e. $Y = X \setminus S$ with S a finite set of closed points of X, then the components of Y(R) are subsets of the components Γ_i of X(R) described as follows [10, 6.10]:

a) If $\Gamma_i \cap S$ contains at most one point, then $\Gamma_i \setminus S$ is a component of Y(R).

b) If $\Gamma_i \cap S$ consists of t+1 points $P_0, P_1, \dots, P_t, t \ge 1$, and if $P_1 < P_2 < \dots < P_t$ in $\Gamma_i \setminus \{P_0\}$, then the intervals $]P_0, P_1[,]P_1, P_2[, \dots,]P_{t-1}, P_t[,]P_t, P_0[$ are the components of Y(R) contained in Γ_i .

¹ In [10, Theorem 7.2] the word "finite" had inadvertently been omitted

Notice that every smooth connected curve Y may be regarded as a Zariski open subscheme of a complete smooth connected curve X uniquely determined by Y up to isomorphism. We call X the *completion* of Y. If P and Q are different points of an oriented component B of Y(R), then the orientation of B "extends" uniquely in the obvious sense to an orientation of the component Γ of X(R) containing B, and we can speak of the closed interval [P, Q] in X(R). If $B = \Gamma$ then of course $[P, Q] \subset Y(R)$. But if $B \neq \Gamma$ then B is totally ordered by the orientation and [P, Q] is contained in Y(R) if and only if P < Q. We call the intervals [P, Q] as far as they are contained in Y(R) the closed intervals of Y(R). In the same way we define the open intervals of Y(R) and the two kinds of half open intervals.

Let now $\varphi: X \to Y$ be a non constant – hence finite – morphism between complete smooth connected curves X and Y over R. Assume that X(R) and Y(R) are oriented. We fix a component Γ of X(R) and look at the restriction $\varphi_R | \Gamma: \Gamma \to Y(R)$ of $\varphi_R: X(R) \to Y(R)$. This map has been studied in [10, §8] in the special case $Y = \mathbb{P}^1$, the projective line over R, and then along the same lines by Fehlner [7] in general. We recall Fehlner's results here including proofs, since his paper is not published. We make strong use of [10, §8] without further comment.

We choose differentials ω_X , ω_Y in the orientation classes of X and Y and define a function $\Delta = \Delta(\varphi) \in R(X)^*$ by

$$\Delta := \varphi^*(\omega_{\gamma})/\omega_{\chi}.$$

This function Δ has no poles on X(R). The real zeros of Δ are precisely the points of X(R) where the finite map $\varphi: X \to Y$ is ramified.

Lemma 2.1. Let P be a point of X(R), $Q = \varphi(P)$, and f a function in $R(Y)^*$.

- i) If $\tau_{Q}(f) \neq 0$, then $\tau_{P}(f \circ \varphi) = \tau_{Q}(f)$, hence $\partial_{P}(f \circ \varphi) = 0$.
- ii) If $\tau_P(\Delta) \neq 0$, then $\partial_P(f \circ \varphi) = \partial_Q(f) \tau_P(\Delta)$.

Proof. i) Let π be a uniformizing element of R(Y) at Q and Π a uniformizing element of R(X) at P. We write $f = \pi^n h$ with $h(Q) \neq 0$, ∞ and $\pi \circ \varphi = \Pi^e \cdot k$ with $k(P) \neq 0$, ∞ .

Assume $\tau_o(f) \neq 0$, i.e. *n* even. Then

 $f \circ \varphi = \Pi^{ne}(h \circ \varphi) \, k^n$

has even order at P and

$$\tau_P(f \circ \varphi) = \operatorname{sign}(h \circ \varphi)(P) = \operatorname{sign} h(Q) = \tau_O(f).$$

ii) If f has even order at Q then $f \circ \varphi$ has even order at P and both sides of the stated equality are zero. Assume now that $\operatorname{ord}_Q(f)$ is odd. We introduce the function $g := df/\omega_r$ which has even order at Q. We have

$$\partial_{P}(f \circ \varphi) = \tau_{P}(d(f \circ \varphi)/\omega_{X})$$
$$= \tau_{P}(\varphi^{*}(df)/\omega_{X}) = \tau_{P}(\Delta \cdot \varphi^{*}(g)).$$

Now we assume that $\tau_P(\Delta) \neq 0$. Then

$$\partial_P(f \circ \varphi) = \tau_P(\Delta) \tau_P(g \circ \varphi).$$

Semialgebraic Topology Over a Real Closed Field I.

By part i) of the lemma we obtain

$$\tau_P(g \circ \varphi) = \tau_O(g) = \partial_O(f),$$

and the stated equality follows. q.e.d.

If a function $f \in R(Y)^*$ is definite on Y(R) then by the trivial part i) of the lemma the function $f \circ \varphi$ is definite on X(R). Thus clearly $\varphi(\Gamma)$ is contained in a component B of Y(R).

Assume now that $[P_1, P_2]$ is a closed interval in Γ with $\varphi([P_1, P_2]) \neq B$ and study the map

$$\psi = \varphi \mid [P_1, P_2] \colon [P_1, P_2] \to B \setminus \{Q_0\}$$

between the totally ordered sets $[P_1, P_2]$ and $B \setminus \{Q_0\}$.

Theorem 2.2 (Intermediate value theorem). $\varphi([P_1, P_2])$ contains every point Q between $\varphi(P_1)$ and $\varphi(P_2)$. In other words, if $\varphi(P_1) < \varphi(P_2)$ then $[\varphi(P_1), \varphi(P_2)] \subset \varphi([P_1, P_2])$, and if $\varphi(P_1) > \varphi(P_2)$ then $[\varphi(P_2), \varphi(P_1)] \subset \varphi([P_1, P_2])$.

Proof. Eventually reversing the orientation of B we assume $\varphi(P_1) < \varphi(P_2)$ without loss of generality. Let Q be a given point in $]\varphi(P_1), \varphi(P_2)[$. We choose an interval function f for $]Q_0, Q[$. Then $(f \circ \varphi)(P_1) < 0$ since $\varphi(P_1)$ lies in $]Q_0, Q[$, but $(f \circ \varphi)(P_2) > 0$ since $\varphi(P_2)$ does not lie in this interval. According to $[10, \S 8]$ the function $f \circ \varphi$ has a zero P (of odd order) in $]P_1, P_2[$. Since f has only two real zeros Q_0 and Q, either $\varphi(P) = Q_0$ or $\varphi(P) = Q$. By our assumption on Q_0 the first case is impossible. Thus $\varphi(P) = Q$. q.e.d.

Theorem 2.3. If $\tau_P(\Delta) = +1$ (resp. if $\tau_P(\Delta) = -1$) for all P in $]P_1, P_2[$, then $\varphi | [P_1, P_2]$ is strictly monotonely increasing (resp. decreasing) and thus maps $[P_1, P_2]$ bijectively to $[\varphi(P_1), \varphi(P_2)]$ (resp. $[\varphi(P_2), \varphi(P_1)]$).

Proof. Reversing the orientation of B we alter $\tau_P(\Delta)$ into $-\tau_P(\Delta)$. Thus we may assume that $\tau_P(\Delta) = +1$ for all $P \in]P_1, P_2[$. We choose some non constant function $f \in R(Y)$ without poles on B. Let $P'_1 < ... < P'_{t-1}$ be the zeros of odd order of the differential $d(f \circ \varphi)$ in] P_1 , P_2 [. (If there are no such zeros t = 1, otherwise t > 1.) Put $P'_0 = P_1$, $P'_t = P_2$. According to [10, §8] the function $f \circ \varphi$ is on each subinterval $[P'_{i-1}, P'_{i}]$ strictly monotone and in particular injective. Thus also φ is injective on every subinterval $[P'_{i-1}, P'_i]$. Let $Q'_i := \varphi(P'_i)$. Assume for the moment that for some fixed i we have $Q'_{i-1} < Q'_i$. Then $\varphi([P'_{i-1}, P'_i]) \supset [Q'_{i-1}, Q'_i]$ by the preceding theorem. Suppose $\varphi([P_{i-1}, P'_i])$ contains a point $Q < Q'_{i-1}, Q'_i$ $P_{i-1}' < P < P_i'$. Then $\varphi([P_{i-1}', P]) \supset [Q, Q_{i-1}']$ $= \varphi(P)$ with and also $\varphi([P, P'_i]) \supset [Q, Q'_{i-1}]$. This contradicts the injectivity of $\varphi([P'_{i-1}, P'_i])$. Thus no such point Q exists. By the same reasoning we see that $\varphi([P'_{i-1}, P'_i])$ does not contain any point $Q > P'_i$. Thus we have $\varphi([P'_{i-1}, P'_i]) = [Q'_{i-1}, Q'_i]$. In the same way we conclude in the case $Q'_{i} < Q'_{i-1}$ that $\varphi([P'_{i-1}, P'_{i}]) = [Q'_{i}, Q'_{i-1}]$.

We now introduce the differential quotients

$$h:=df/\omega_{\chi}, \quad g:=d(f\circ\varphi)/\omega_{\chi}.$$

Clearly $g = \Delta \cdot (h \circ \varphi)$. The function g has for P in the interval $]P'_{i-1}$, P'_i constant sign $\tau_P(g) = \varepsilon_i$, since the points P'_i with $1 \le i \le t-1$ are all the points of odd order

of g in] P_1 , P_2 [. We know from [10, §8] that $f \circ \varphi$ is increasing on $[P'_{i-1}, P'_i]$ if $\varepsilon_i = +1$ and decreasing if $\varepsilon_i = -1$. Since $g = \Delta \circ (h \circ \varphi)$ we see from Lemma 2.1 that h has in $\varphi([P_1, P_2])$ odd order precisely at the points Q'_i , and that $\tau_Q(h) = \varepsilon_i$ for all points Q of $\varphi(]P'_{i-1}, P'_i[$. Thus $f \circ \varphi$ is increasing (resp. decreasing) on the interval $[P'_{i-1}, P'_i]$ if and only if f is increasing (resp. decreasing) on the interval $\varphi([P'_{i-1}, P'_i])$. We conclude that φ is strictly increasing on each interval $[P'_{i-1}, P'_i]$, hence on the whole interval $[P_1, P_2]$. q.e.d.

Example. If φ is etale at every point of $]P_1, P_2[$ then Δ has no zeros on $]P_1, P_2[$, hence has constant sign on this interval, and Theorem 2.3 applies.

If Δ has on the interval $]P_1, P_2[$ zeros of odd order $P'_1 < ... < P'_{t-1}$ then by Theorem 2.3 φ is strictly monotone on each subinterval $[P'_{i-1}, P'_i]$ of $[P_1, P_2]$ {i = 1, ..., t; put $P'_0 = P_1, P'_i = P_2$ }. In particular $\varphi([P'_{i-1}, P'_i])$ is a closed interval adjacent to the interval $\varphi([P'_i, P'_{i+1}])$. Thus $\varphi([P_1, P_2])$ is again a closed interval.

More precisely φ is increasing in $[P'_{i-1}, P'_i]$ if $\tau_P(\Delta) = +1$ for P in the interior of this interval and decreasing otherwise. Since Δ has opposite sign in adjacent intervals $]P'_{i-1}, P'_i[$ and $]P'_i, P'_{i+1}[$ we see that φ has its local minima in $]P_1, P_2[$ precisely at the points P'_i with $\partial_{P'_i}(\Delta) = +1$ and its local maxima precisely at the other points P'_i ($1 \le i \le t$). Thus we have a picture of the behaviour of $\varphi|[P_1, P_2]$ as explicit as in the special case $Y = \mathbb{P}^1$ in [10, §8]. In particular we have the following converse of Theorem 2.3:

Corollary 2.4. If φ is injective on $]P_1, P_2[$, then Δ has no zeros of odd order in $]P_1, P_2[$ and thus φ is strictly monotone on $[P_1, P_2]$.

We now abandon the hypothesis that $\varphi([P_1, P_2]) \neq B$.

Theorem 2.5. There exists a subdivision

$$P_1 = P'_0 < P'_1 < \ldots < P'_t = P_2$$

of the given interval $[P_1, P_2]$ such that φ maps every subinterval $[P'_{i-1}, P'_i]$ (i = 1, ..., t) strictly monotone onto a closed interval in B. The image $\varphi([P_1, P_2])$ is either a closed interval or the whole of B. If Δ has no zeros of odd order in $]P_1, P_2[$, hence constant sign ε on this interval, then φ is increasing on every interval $[P'_{i-1}, P'_i]$ if $\varepsilon = +1$ and decreasing on every $[P'_{i-1}, P'_i]$ if $\varepsilon = -1$.

Proof. We have seen this above in the case $\varphi([P_1, P_2]) \neq B$. Assume now that $\varphi([P_1, P_2]) = B$. We choose a subset S of B consisting of three points and containing $\varphi(P_1)$ and $\varphi(P_2)$. Let

$$P_1 = R_0 < R_1 < \ldots < R_s = P_2$$

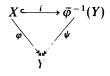
be the finitely many points P in $[P_1, P_2]$ with $\varphi(P) \in S$. Then each set $\varphi(]R_{i-1}, R_i[)$ does not meet S (i=1, ..., s). Thus $\varphi([R_{i-1}, R_i])$ misses at least one point of S, hence is different from B. Now we can divide each interval $[R_{i-1}, R_i]$ into finitely many closed subintervals such that φ is strictly monotone on each of them. The last assertion in the theorem is evident from Theorem 2.3. q.e.d.

As a first application of our theory on smooth curves we prove

Theorem 2.6. Let X and Y be varieties over R of dimension ≤ 1 and let φ be a finite morphism from X to Y. Then the continuous map $\varphi_R: X(R) \to Y(R)$ maps any closed subset A of X(R) onto a closed subset $\varphi(A)$ of Y(R).

Proof. It suffices to prove this for the restrictions $\varphi^{-1}(Z) \to Z$ of φ over the reduced irreducible components Z of Y. Thus we may assume from the beginning that the variety Y is reduced and irreducible. If Y is a point the assertion is trivial. Thus assume that Y is an integral curve. We then also may restrict φ to the reduced irreducible components of X and thus assume that X is an integral curve. Let $\pi: \tilde{X} \to X$ denote the normalization map of X. With $\tilde{\varphi} := \varphi \circ \pi$ we have $\varphi_R(A) = \tilde{\varphi}(\pi_R^{-1}(A))$, and $\pi_R^{-1}(A)$ is closed in $\tilde{X}(R)$. Thus it suffices to prove the theorem for $\tilde{\varphi}$ instead of φ , and we may assume from the beginning that X is smooth and connected.

Let \overline{Y} be a projective completion of Y and let \overline{X} be the smooth completion of X. Then φ extends uniquely to a morphism $\overline{\varphi}: \overline{X} \rightarrow \overline{Y}$ which is again finite. We have a commutative diagram



with ψ the restriction of $\bar{\varphi}$ over Y and *i* the inclusion map, which is an open immersion. Since φ and ψ are finite also *i* is finite. This implies $X = \bar{\varphi}^{-1}(Y)$. Let \bar{A} denote the closure of A in $\bar{X}(R)$ in the strong topology. Then

$$\varphi(A) = \overline{\varphi}(\overline{A}) \cap Y(R).$$

Thus it suffices to prove that $\overline{\varphi}(\overline{A})$ is closed in $\overline{Y}(R)$. We learn that we may assume without loss of generality in addition that the curves X and Y are projective.

Let Q be a given point of $Y(R) \setminus \varphi(A)$. We have to find a neighbourhood W of Q in Y(R) which does not meet $\varphi(A)$. Now Y is a closed reduced subvariety of some projective space \mathbb{P}^n over R. We choose a projective system of coordinates x_0, \ldots, x_n of \mathbb{P}^n such that $Q = (1: 0: \ldots: 0)$, and we introduce on Y the rational function

$$g = (x_1^2 + \ldots + x_n^2)/(x_0^2 + \ldots + x_n^2).$$

g is defined in a Zariski neighbourhood U of Y(R), and has on Y(R) the only zero Q. We consider the regular function $h:=g \circ \varphi$ on the Zariski open set $\varphi^{-1}(U) \supset X(R)$ in X. The real algebraic function $h_R: X(R) \rightarrow R$ has as set of zeros the finite fibre

$$\varphi_R^{-1}(Q) = \{P_1, \ldots, P_r\},\$$

which is disjoint from A. We choose open intervals J_1, \ldots, J_r in X(R) containing P_1, \ldots, P_r respectively and all disjoint from A. Then

$$N := X(R) \smallsetminus (J_1 \cup \ldots \cup J_r)$$

is a disjoint union of closed intervals and whole components, and N contains A. According to [10, §8] – or the more general Theorem 2.5 – the function $h | N: N \to R$ attains on N a minimum $\varepsilon > 0$. Thus $g(Q') \ge \varepsilon$ for all Q' in $\varphi(N)$. The neighbourhood $\{Q' \in Y(R) | g(Q') < \varepsilon\}$ of Q does not meet $\varphi(N)$, hence does not meet the subset $\varphi(A)$ of $\varphi(N)$. q.e.d.

§3. Combinatorial Paths

Let now X be an arbitrary variety over the real closed field R.

Definition 1. An elementary path in X(R) is a totally ordered subset γ of X(R) which either consists only of one point ("degenerate" elementary path) or has the following two properties:

a) The Zariski closure of γ in X is an irreducible curve $D \subset X$ (with the reduced subscheme structure of course).

b) Let $\pi: \tilde{D} \to D$ denote the normalization map of D. Then after choice of a suitable orientation on $\tilde{D}(R)$ there exists a closed interval $[\tilde{P}, \tilde{Q}]$ in $\tilde{D}(R)$ which under π is mapped bijectively and order preserving onto γ .

Notice that $[\tilde{P}, \tilde{Q}]$, is uniquely determined by γ , namely if $S \subset D(R)$ is the finite set of points $P \in D(R)$ with $\pi_R^{-1}(P)$ containing more than one point then $[\tilde{P}, \tilde{Q}]$ is clearly the closure of $\pi_R^{-1}(\gamma \setminus S)$ in $\tilde{D}(R)$. We call the ordered set $[\tilde{P}, \tilde{Q}]$ the normalization $\tilde{\gamma}$ of the elementary path γ .

Notice also that the ordered set γ has a first element $P := \pi(\tilde{P})$, called the starting point of γ , and a last element $Q := \pi(\tilde{Q})$, called the *end point* of γ . If $\gamma = \{P\}$ then we of course call P both the starting and end point of γ .

Applying Theorem 2.6 to $\pi: \tilde{D} \rightarrow D$ we see that every elementary path γ is a closed subset of X(R).

If P and Q are points of a nondegenerate elementary path γ with P < Q, then we denote by [P, Q] the set of all points P' of γ with $P \le P' \le Q$, equipped with the total ordering from γ . Clearly [P, Q] is again an elementary path, called a *closed interval of* γ or an elementary subpath of γ .

Definition 2. A combinatorial path γ in X(R) is a finite sequence $(\gamma_1, ..., \gamma_i)$ of elementary paths γ_i in X(R) such that the end point of γ_i coincides with the starting point of γ_{i+1} $(1 \le i \le t-1)$.

Of course we define the starting point P of γ as the starting point of γ_1 and the end point Q of γ as the end point of γ_i , and we often call P and Q "the end points" of γ . We also say that γ is a path connecting P with Q.

The composition $\gamma * \delta$ of a path $\gamma = (\gamma_1, ..., \gamma_t)$ ending at Q with a path $\delta = (\delta_1, ..., \delta_t)$ starting at Q is defined by

$$\gamma * \delta = (\gamma_1, \ldots, \gamma_t, \delta_1, \ldots, \delta_r).$$

If γ is an elementary path, then let γ^{-1} denote the set γ equipped with the opposite total ordering. Clearly γ^{-1} is again an elementary path. For a combinatorial path $\gamma = (\gamma_1, ..., \gamma_t)$ the *inverse path* γ^{-1} is defined as $(\gamma_t^{-1}, ..., \gamma_1^{-1})$. If γ starts at P and ends at Q then γ^{-1} starts at Q and ends at P.

Semialgebraic Topology Over a Real Closed Field I.

Definition 3. We call two points P and Q of X(R) path connectable if there exists a combinatorial path γ in X(R) starting at P and ending at Q. Since we have the possibility to compose and to invert paths, "connectable" is an equivalence relation on the set X(R). We call the equivalence classes the path components of X(R) and we denote the set of all path components of X(R) by $\pi_0(X(R))$.

It is clear from §2 that in the case that X is a smooth curve the path components of X(R) are just the components $\Gamma_1, \ldots, \Gamma_r$ of X(R) introduced there.

In Part II of the paper we shall define beside combinatorial paths also "semialgebraic paths". But up to then we are free to use the shorter word "path" instead of "combinatorial path".

We now explain that in some sense the image of a path in X(R) under any morphism $\varphi: X \to Y$ is again a path.

Theorem 3.1. Let $\varphi: X \to Y$ be a morphism between varieties over R and γ be an elementary path in X(R) starting at P and ending at Q.

i) If the restriction $\varphi | \gamma$ is injective then φ maps γ order preserving onto an elementary path δ of Y(R).

ii) If $\varphi(\gamma)$ contains more than one point then there exists a subdivision

$$P_0 = P < P_1 < P_2 < \ldots < P_t = Q$$

of γ such that φ is injective on every subpath $\gamma_i := [P_{i-1}, P_i]$ and thus maps γ_i order preserving onto an elementary path δ_i starting at $\varphi(P_{i-1})$ and ending at $\varphi(P_i)$.

Proof. We may restrict φ to the Zariski closure D of γ and thus assume from the beginning that X is an irreducible reduced curve. Let $\pi: \tilde{X} \to X$ be the normalization of X and $\tilde{\gamma}$ the normalization of γ . It suffices to prove both statements for $\tilde{\gamma}$ and $\varphi \circ \pi$ instead of γ and φ . Thus we assume since now that X is a smooth connected curve. We assume further that $\varphi(\gamma)$ is not a point, i.e. that $\varphi(X)$ is not a point. Then the Zariski closure of $\varphi(X)$ in Y – with reduced subscheme structure, as always – is an irreducible curve. We may replace Y by this curve. Thus we assume since now that Y is an irreducible reduced curve. We further replace Y by a projective completion of X, hence assume that Y is projective. Then φ can be extended to a morphism $\overline{X} \to Y$ with \overline{X} the smooth completion of X. Thus we also assume without loss of generality that X is complete.

We choose an orientation of X(R) such that γ is an interval [P, Q]. We have a canonical commutative triangle



with π the normalization of Y. We now prove part i) of the theorem. Since φ is injective on γ also $\tilde{\varphi}$ is injective on γ . Thus by the result of §2 (Cor. 2.4 and Th. 2.5) $\tilde{\varphi}$ maps γ onto the interval $\delta := [\tilde{\varphi}(P), \tilde{\varphi}(Q)]$ in a strictly monotonely increasing way after suitable choice of an orientation of $\tilde{Y}(R)$. Moreover π is injective on δ . Thus by the definition of elementary paths π maps δ order preserving onto an elementary path δ in Y(R).

We now prove part ii) of the theorem. According to Theorem 2.5 we have a subdivision

$$P_0 = P < P_1 < \ldots < P_m = Q$$

such that $\tilde{\varphi}$ maps each interval $[P_{i-1}, P_i]$ order preserving onto an elementary path $\tilde{\delta}_i$ in $\tilde{Y}(R)$. Thus it suffices to prove the assertion ii) for the map π and the paths $\tilde{\delta}_i$ instead of φ and γ , and we may assume since now that φ is the normalization map $\pi: \tilde{Y} \to Y$. We choose an orientation on $\tilde{Y}(R)$ such that γ is an interval [P, Q] in $\tilde{Y}(R)$. Let $P_1 < P_2 < ... < P_{m-1}$ be the finitely many points in]P, Q[which lie over singular points of $Y(m \ge 1; m=1)$ if there are no such points). Put $P_0 := P$ and $P_m := Q$. Then π is injective on every open interval $]P_{i-1}, P_i[$. If $\pi(P_{i-1}) \neq \pi(P_i)$ then π is injective on $[P_{i-1}, P_i]$. Then we do not subdivide this interval further. If $\pi(P_{i-1}) = \pi(P_i)$ then we choose some point S_i in $]P_{i-1}, P_i[$. Now π is injective on both subintervals $[P_{i-1}, S_i]$ and $[S_i, P_i]$. We thus obtain a subdivision of $\gamma = [P, Q]$ which fulfills the requirements of assertion ii). q.e.d.

If $\varphi: X \to Y$ is a morphism and $\gamma = (\gamma_1, ..., \gamma_i)$ is a path in X(R) then by Theorem 3.1 there exists a "refinement" of γ to a path $\gamma' = (\gamma'_1, ..., \gamma'_s)$ by suitable subdivision of the γ_i such that φ maps every γ'_i to an elementary path δ_i in an order preserving way. Thus we can associate with γ a path $\varphi(\gamma) = (\delta_1, ..., \delta_s)$. This path $\varphi(\gamma)$ is determined by γ uniquely up to refinements in the obvious sense. We shall usually speak of "the" image $\varphi(\gamma)$ of a path γ under a morphism φ for short.

As an immediate consequence of the existence of the image of a path we obtain

Corollary 3.2. Let $\varphi: X \to Y$ be a morphism between varieties over R. Then φ maps every path component Γ of X(R) into a path component B of Y(R). Thus φ induces a well defined map $\varphi_*: \pi_0(X) \to \pi_0(Y)$.

We now turn to the problem to lift paths. In the following theorem it will be important for applications that we do *not* assume that the morphism φ is etale over the end point of the path γ .

Theorem 3.3. Let γ be an elementary path in Y(R) with end point Q_1 and let $\varphi: X \to Y$ be a finite morphism. Let Q_0 be some point of γ different from Q_1 and let P_0 be a point in X(R) with $\varphi(P_0) = Q_0$. Let B denote the path component of P_0 in X(R). We assume that φ is etale at all points of the set $B \cap \varphi^{-1}(\gamma \setminus \{Q_1\})$. Then there exists a unique elementary path β in X(R) which contains P_0 and is mapped by φ bijectively and order preserving onto γ .

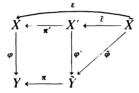
Proof. The path γ is nondegenerate. We may replace Y by the (reduced) Zariski closure D of γ and X by the scheme $\varphi^{-1}(D)$. We may then replace $\varphi^{-1}(D)$ by the reduced scheme $\varphi^{-1}(D)_{red}$ since this does not change $\varphi^{-1}(D)$ on a Zariski neighbourhood of $B \cap \varphi^{-1}(\gamma \setminus \{Q_1\})$. Thus we may assume from the beginning that X is reduced and Y is an irreducible reduced curve.

Let $\pi: \tilde{Y} \to Y$ denote the normalization of Y and let



be the scheme theoretic fibre product of \tilde{Y} and X over Y. Let $\tilde{\gamma}$ denote the normalization of γ and let \tilde{Q}_0, \tilde{Q}_1 denote the preimages of Q_0, Q_1 in $\tilde{\gamma}$. Let P'_0 denote the point of X'(R) lying over P_0 and \tilde{Q}_0 , and let finally B' denote the path component of P'_0 in X'(R). The morphism π' maps B' into B. Thus φ' is etale on the set $B' \cap \varphi'^{-1}(\tilde{\gamma} \setminus \{\tilde{Q}_1\})$. In particular this set does not contain singular points. It is easy to see that any lifting β of γ through P_0 yields a unique lifting β of $\tilde{\gamma}$ through P'_0 and vice versa. The irreducible components of X' yield a disjoint decomposition of $B' \cap \varphi'^{-1}(\tilde{\gamma} \setminus \{\tilde{Q}_1\})$ into relatively Zariski-closed subsets. Therefore it is obvious that any lifting β of $\tilde{\gamma}$ through P'_0 will run in the irreducible component X'' of X' containing P'_0 . The set $\pi'(X'')$ is an irreducible component of X. We may replace X by this component, equipped with the reduced structure, and X' by the preimage of this component under π' . This preimage is irreducible (hence coincides with X''). Thus we assume without loss of generality that X and X' are already irreducible reduced curves.

Let $\tilde{\varepsilon}: \tilde{X} \to X'$ be the normalization of X'. We obtain a natural commutative diagram



All maps are finite, and ε is the normalization of X. The morphism $\tilde{\varepsilon}$ is an isomorphism over a Zariski neighbourhood of $B' \cap \varphi'^{-1}(\tilde{\gamma} \setminus \{\tilde{Q}_1\})$. In particular there is a unique point \tilde{P}_0 in $\tilde{X}(R)$ over P'_0 . Let \tilde{B} denote the path component of \tilde{P}_0 in $\tilde{X}(R)$. Clearly $\tilde{\varphi}$ is etale on the set $\tilde{B} \cap \tilde{\varphi}^{-1}(\tilde{\gamma} \setminus \{\tilde{Q}_1\})$. We further see that the elementary paths β through P_0 which lift γ correspond uniquely with the elementary paths $\tilde{\beta}$ through \tilde{P}_0 which lift $\tilde{\gamma}$. Thus in the proof of our theorem we may replace φ , γ , P_0 by $\tilde{\varphi}$, $\tilde{\gamma}$, \tilde{P}_0 , and we assume since now that X and Y are smooth connected curves. We may extend φ to a finite morphism $\bar{\varphi}: \bar{X} \to \bar{Y}$ between the smooth completions \bar{X} , \bar{Y} of X and Y. Notice that $\bar{\varphi}$ maps $\bar{X} \setminus X$ onto $\bar{Y} \setminus Y$ since φ is finite. Thus we assume in addition that X and Y are complete.

After a suitable orientation of Y(R) the path γ is an interval $[S, Q_1]$. We choose some orientation on the component B of X(R), and we know from §2 that $\varphi^{-1}([S, Q_1]) \cap B = M$ is either a disjoint union of finitely many closed intervals M_1, \ldots, M_r or the whole of B.

We now apply the results of §2 to the map $\varphi | M: M \to [S, Q_1]$. There exists some point T' of M where φ attains its minimum on M. Certainly $\varphi(T') \neq Q_1$, thus φ is etale at T'. In particular the function $\Delta(\varphi)$, introduced in §2, does not vanish at T'. Thus T' cannot be an interior point of M. In particular $M \neq B$. Let M_1 be the closed interval among the M_i which contains P_0 , and let now T denote a point of M_1 where φ attains its minimum on M_1 . Then as above T is one of the end points of M_1 . Eventually reversing the orientation of B we assume that T is the initial point of M_1 . Thus $M_1 = [T, P_1]$ with some other point P_1 of B. The value Q_1 is certainly not a local minimum of $\varphi: [T, P_1] \to [S, Q_1]$, since φ is nowhere locally constant on $[T, P_1]$ (cf. §2). We see as above that φ attains nowhere a local minimum on $]T, P_1[$ and according to §2 we have to consider two cases:

Case 1. φ is strictly increasing on $[T, P_1]$.

Case 2. There exists some point A in]T, P_1 [such that φ is strictly increasing on [T, A] and strictly decreasing on [A, P_1].

In both cases we must have $\varphi(T) = S$. Indeed, if $\varphi(T) > S$ there would exist an interval $[T_1, T]$ which by φ would be mapped into $[S, Q_1]$. This is impossible by the definition of M_1 . By the same argument we see that in the first case $\varphi(P_1) = Q_1$ and in the second case $\varphi(P_1) = S$. In the first case we are through: The interval $[T, P_1]$ is mapped strictly monotonely increasing onto $[S, Q_1]$ and is the only such elementary path containing P_0 .

Assume now that we are in the second case. Suppose that $\varphi(A) < Q_1$. Then φ is etale at A and $\Delta(\varphi)(A) \neq 0$. But this is impossible since φ attains at A the maximum. Thus $\varphi(A) = Q_1$. We see that [T, A] is mapped strictly monotonely increasing onto $[S, Q_1]$ and that $[A, P_1]$ is mapped strictly monotonely decreasing onto $[S, Q_1]$. Clearly $P_0 \neq A$. If P_0 lies in [T, A[then $\beta = [T, A]$ is the only solution of our lifting problem. Otherwise P_0 lies in $]A, P_1]$ and the path β inverse to $[A, P_1]$ is the only solution of our lifting problem. Theorem 3.2 is proved.

Remark. In the same way it is possible to prove the existence and uniqueness of a lifting of the elementary path γ through P_0 if we only assume that φ is etale in $B \cap \varphi^{-1}(\hat{\gamma})$ with $\hat{\gamma}$ the set obtained from γ by omission of both end points, as long as Q_0 is different from the end points. Indeed, we only have to modify the end of the proof of Theorem 3.2 slightly. We meet a more complicated situation than the two cases there but we can use the same method of proof.

We state a modest consequence of Theorem 3.2.

Corollary 3.3. Let $\varphi: X \to Y$ be a finite morphism between varieties over R. Let B be a path component of X(R) on which φ is everywhere etale. Then $\Gamma := \varphi(B)$ is a path component of Y(R) and all the fibres of the restriction $B \to \Gamma$ of φ contain the same number of points.

§4. Path Components

Theorem 4.1. For any variety X over R the set $\pi_0(X(R))$ of path components is finite.

Proof. We proceed by induction on the dimension n of X. The case n=0 is trivial. Assume since now that $n \ge 1$. Since X is the union of finitely many affine open subsets, we may assume that X itself is affine. All the finitely many reduced irreducible components of X have at most dimension n. Thus we may assume in addition that X is reduced and irreducible, hence $X = \operatorname{Spec} A$ with A an n-dimensional affine algebra over R which has no zero divisors. By Noether's normalization Lemma A contains a polynomial algebra $R[t_1, \ldots, t_n]$ with indeterminates t_1, \ldots, t_n such that A is finite over $R[t_1, \ldots, t_n]$. The inclusion $R[t_1, \ldots, t_n] \hookrightarrow A$ corresponds with a finite surjective morphism $\varphi: X \to A^n$ from X to the affine standard space A^n over R. Let Δ denote the discriminant hypersurface of φ in A^n . This is a closed proper subvariety of A^n such that φ is etale over $A^n \smallsetminus \Delta$. Let Z denote the subscheme $\varphi^{-1}(\Delta)$ of X.

We consider first the case that Z(R) is empty. Then φ is etale on X(R). Clearly R^n is path connected, since any two points of R^n can be joined by an interval on a real line in R^n . According to Corollary 3.3 all fibres of the map $\varphi_R: X(R) \to R^n$ have the same finite cardinality d, and by the same corollary X(R) has at most d path components.

We now consider the case that Z(R) is not empty. Z has at most dimension n-1. Thus by induction hypothesis $\pi_0(Z(R))$ has a finite cardinality N. Let now P_0 be some point in $X(R) \setminus Z(R)$. We want to show that P_0 can be connected to some point in Z(R) by an (elementary) path. Then we know that X(R) has at most N path components. We choose some straight line G in \mathbb{R}^n through the point $Q_0 := \varphi(P_0)$ which meets the non empty set $\Delta(R)$. Then $G \cap \Delta(R)$ consists of finitely many points. Thus there exists an elementary path γ in G with initial point Q_0 and end point Q_1 in $\Delta(R)$, such that no point of γ different from Q_1 lies in $\Delta(R)$. By Theorem 3.2 γ can be lifted (in a unique way) to an elementary path β in X(R) which starts at P_0 . The end point P_1 of β lies over Q_1 and hence in Z(R). This finishes the proof of the theorem.

We now want to prove that the path components of X(R) are open in X(R). For this we need a generalization of Theorem 2.6 to arbitrary dimensions.

Theorem 4.2. For any finite morphism $\pi: X \to Y$ between varieties X, Y over R the map $\pi_R: X(R) \to Y(R)$ is closed.

Proof. We may assume without loss of generality that Y and hence also X is affine. Up to isomorphisms we meet the following situation:

(1) $Y = \operatorname{Spec} A \subset \mathbb{A}^n$, $A = \mathbb{R}[X_1, \dots, X_n]/J$ with some ideal J of polynomials in n variables X_1, \dots, X_n ,

(2) $X = \operatorname{Spec} A[T_1, \dots, T_m]/I \subset \mathbb{A}^{n+m}$ with further variables T_1, \dots, T_m , and π the restriction to X of the projection $pr: \mathbb{A}^{n+m} \to \mathbb{A}^n$ to the first n coordinates. We denote the image of T_i in $A[T_1, \dots, T_m]/I$ by t_i and introduce the varieties

$$Y_i = \operatorname{Spec} A[t_1, \ldots, t_i] \subset \mathbb{A}_R^{n+i}.$$

 $(i=0,\ldots,m, Y_0=Y, Y_m=X)$. We have a factorization

$$\pi = \pi_1 \circ \ldots \circ \pi_m$$

with π_i denoting the canonical projection from Y_i to Y_{i-1} . All these morphisms $\pi_i: Y_i \to Y_{i-1}$ are again finite, and it suffices to prove that every map $(\pi_i)_R$ is closed. Thus we may assume without loss of generality that m = 1, hence

$$X = \operatorname{Spec} A[T]/I \subset \mathbb{A}^{n+1}$$

with one variable T. Since X is finite over Y there certainly exists a normed polynomial $G(T) \in A[T]$ lying in I. Then X is a closed subscheme of

$$X_1 := \operatorname{Spec} A[T]/(G(T)) \subset \mathbb{A}^{n+1}$$

and the natural projection $\psi: X_1 \to Y$ extends π . Since X(R) is closed in $X_1(R)$ it suffices to prove that ψ_R is closed. We further can lift G(T) to an element $F(X_1, \ldots, X_n, T) = T^d + a_1(X_1, \ldots, X_n) T^{d-1} + \ldots + a_d(X_1, \ldots, X_n)$ of the polynomial ring $R[X_1, \ldots, X_n, T]$. Then X_1 is a closed subscheme of the hypersurface

$$Z := \operatorname{Spec} R[X_1, \dots, X_n, T] / (F(X_1, \dots, X_n, T))$$

in \mathbb{A}^{n+1} , and the natural projection

 $\varphi: Z \to \mathbb{A}^n$

to the first *n* coordinates extends ψ . Since $X_1(R)$ is closed in Z(R) and Y(R) is closed in \mathbb{R}^n , it suffices to prove that φ_R is closed.

We proceed in a similar way as in the proof of Theorem 2.6. Let M be a closed subset of Z(R) and let $b = (b_1, \ldots, b_n)$ be a point of R^n not contained in $\varphi(M)$. We choose some $\sigma > 0$ in R. Then the preimage of the cube

$$Q = \{(y_1, ..., y_n) \in \mathbb{R}^n | |y_i - b_i| \le \sigma, i = 1, ..., n\}$$

under φ_R is bounded, i.e. contained in a rectangle

$$P = \{(y, x_{n+1}) \in \mathbb{R}^n \times \mathbb{R} \mid y \in Q, |x_{n+1}| \le \rho\}$$

by a well known elementary estimate. Let $(b, a_1), (b, a_2), ..., (b, a_r)$ be the points in the fibre $\varphi_R^{-1}(b)$. We choose some $\delta > 0$ in R such that the open cubes (j=1,...,r)

$$N_{j} = \{(y, x_{n+1}) \in \mathbb{R}^{n} \times \mathbb{R} \mid |y_{i} - b_{i}| < \delta \quad \text{for } i = 1, \dots, n \mid x_{n+1} - a_{j}| < \delta \}$$

around the points (b, a_j) are all disjoint from the closed set M, and such that the closures \overline{N}_i are contained in the interior \mathring{P} of P.

Now $P \setminus (N_1 \cup ... \cup N_r)$ is the union of finitely many closed rectangles $T_1, ..., T_s$ of R^{n+1} . The polynomial

$$h(x_1, ..., x_{n+1}) = \sum_{i=1}^n (x_i - b_i)^2$$

is nowhere zero on $Z(R) \cap (T_1 \cup ... \cup T_s)$. Using Tarski's principle (cf. e.g. [3]) we see that $|h(x)| \ge \varepsilon_i^2$ for every x in $Z(R) \cap T_i$ with some $\varepsilon_i > 0$, since over \mathbb{R} the

polynomial $\sum_{i=1}^{n} X_i^2$ attains its minimum on the intersection of any given hypersurface in \mathbb{R}^{n+1} with any closed rectangle of \mathbb{R}^{n+1} . Thus $|h(x)| \ge \varepsilon^2$ for all x in M with $\varepsilon = \operatorname{Min}(\varepsilon_1, \ldots, \varepsilon_r, \sigma)$. This means that the ball of radius ε around b in \mathbb{R}^n does not meet the set $\varphi(M)$. q.e.d.

Remark. In Part II of the paper we shall eliminate Tarki's principle in the proof of Theorem 4.2. Using a different method we shall also prove the following more general result: For any proper mapping $\varphi: X \to Y$ between varieties over R any closed semialgebraic subset of X(R) is mapped by φ_R onto a closed semialgebraic subset of Y(R).

We now are able to prove for any variety X over R

Theorem 4.3. Every path component of X(R) is open in X(R), hence also closed.

Proof. We proceed in a similar way as in the proof of Theorem 4.1 by induction on the dimension n of X. The case n=0 is trivial, and we assume now n>0. As above we retreat to the case that X is affine and irreducible. As above we have a finite morphism $\varphi: X \to A^n$, and again we denote by Δ the subvariety of A^n consisting of the points, where φ is not etale in some point of the fibre, and by Z the preimage $\varphi^{-1}(\Delta)$. Let B be a given path component of X(R) and P_1 be a given point of B. We have to find a neighbourhood of P_1 in X(R) which is contained in B. Let Q_0 denote the image $\varphi(P_1)$ and let P_1, P_2, \dots, P_r be all the points in the real fibre $\varphi_R^{-1}(Q_0)$. We choose for each P_i an open neighbourhood W_i so small that the following four properties hold true:

a) $W_i \cap W_j = \emptyset$ for $i \neq j$.

b) If P_1 lies in Z(R) then $W_1 \cap Z(R)$ is contained in the path component A of P_1 in Z(R). {Notice that by induction hypothesis A is open in Z(R).}

c) If P_i does not lie in Z(R) then the intersection $W_i \cap Z(R)$ is empty (i=1,...,r).

d) There exists a function f in the affine ring R[X] of X which is negative everywhere on W_1 and positive everywhere on $W_2 \cup ... \cup W_r$.

It is easy to fulfill also the last requirement d) by use of the Chinese remainder theorem applied to the ring R[X] and the maximal ideals of this ring corresponding to P_1, \ldots, P_r .

Since the map $\varphi_R: X(R) \to R^n$ is closed (Theorem 4.2) there exists an euclidean open ball U around Q_0 of some radius δ such that $\varphi_R^{-1}(U)$ is contained in $W_1 \cup \ldots \cup W_r$. {Choose U in the complement of $\varphi \left(X(R) \smallsetminus \bigcup_{i=1}^r W_i \right)$ in R^n .} Then $V_i := W_i \cap \varphi_R^{-1}(U)$ is again an open neighbourhood of P_i and $\varphi_R^{-1}(U) = V_1 \cup \ldots \cup V_r$.

We now show that $V_1 \subset B$, which will finish the proof. Let P be a given point of V_1 , different from P_1 , and $Q := \varphi(P)$. If P lies in Z(R) then by properties b) and c) $P \in A$. Thus $P \in B$, since A is contained in B.

Assume now that P does not lie in Z(R). We consider first the case that Q_0 does not lie in $\Delta(R)$. Then U does not meet the set $\Delta(R)$ by property c). Let γ be the straight elementary path from Q to Q_0 . This path is contained in U. Let β denote the unique lifting of γ which starts at P. This elementary path β runs in $V_1 \cup \ldots \cup V_r$. The function f has no zeros in $V_1 \cup \ldots \cup V_r$ and is negative at the

starting point P of β . Thus f is negative everywhere on β . This means that β runs in V_1 . The end point of β lies over Q_0 and thus must be the point P_1 . Thus P lies in the path component B of P_1 .

We finally consider the case that P does not lie in Z(R) but Q_0 lies in $\Delta(R)$. Let again γ be the straight elementary path from Q to Q_0 in U. Let Q'_0 be the first point of γ contained in $\Delta(R)$, and let γ' denote the subpath $[Q, Q'_0]$ of γ . By Theorem 3.3 there exists a unique elementary path β' in X(R) which starts at P and lifts γ' . As above we see that β' runs in V_1 . The end point P'_1 of β' lies in $V_1 \cap Z(R)$, hence in A by property a) above. Since $A \subset B$ we see that P lies in B. Now Theorem 4.3 is completely proved.

If R coincides with the field \mathbb{R} of real numbers then every path component Γ of $X(\mathbb{R})$ as defined here is also path connected in the sense of topology, where a path is just a continuous map from [0, 1] to $X(\mathbb{R})$. Thus Γ is also connected in the topological sense, i.e. Γ is not the union of two non empty subsets which are open in Γ . It now follows from Theorem 4.3 that the topological components, the topological path components, and the path components of $X(\mathbb{R})$ as defined above are the same subsets of $X(\mathbb{R})$.

We return to an arbitrary real closed base field R.

Corollary 4.4. Every point P of X(R) has a fundamental system of path connected open neighbourhoods.

Proof. We may assume that X is a closed subvariety of some affine standard space A^N , hence $X(R) \subset R^N$, and that P is the point (0, ..., 0) in R^N . Let some $\rho > 0$ in R be given. Consider the function

$$f(x_1, ..., x_N) = \rho^2 - \sum_{i=1}^N x_i^2$$

on X(R). Let X' denote the Zariski open subset of X where f does not vanish, and let Γ denote the path component of P in X'(R). By Theorem 4.2 the set Γ is open in X'(R) hence also open in X(R). The algebraic function f has no zeros on X'(R) hence constant sign on Γ . Since $f(P) = \rho^2 > 0$ the function f is positive everywhere on Γ . This means that Γ is contained in the open ball $B_{\rho}(P)$ of radius ρ around P. Thus $B_{\rho}(P) \cap X(R)$ contains the path connected open neighbourhood Γ of P. q.e.d.

In Part II of the paper we shall see that every path component of X(R) is also a semialgebraic subset of X(R). We then know by the proof of Corollary 4.4 that every point P of X(R) has a fundamental system of path connected *semialgebraic* open neighbourhoods.

§5. An Application to the Theory of Witt Rings

We now assume that X is a divisorial scheme of finite type over a field k. This is a very general class of schemes containing all quasiprojective and all regular varieties over k, cf. [9, 2.2] and [12, Chap. III§1]. We are interested in the signatures of X, i.e. the ring homomorphisms from the Witt ring W(X) of bilinear spaces over X to the ring of integers \mathbb{Z} . We refer the reader to the lectures [12] and the paper [11] for the general theory and the meaning of signatures.

It is known that every signature $\sigma: W(X) \to \mathbb{Z}$ factors through some point x of X, i.e. there exists a commutative diagram



with $W(X) \to W(\kappa(x))$ the natural map from W(X) to the Witt ring of the residue class field $\kappa(x) = \mathcal{O}_x/m_x$ of the point x [12, V.1]. Recall that the signatures of $\kappa(x)$ correspond uniquely with the total orderings of the field $\kappa(x)$. The main goal of this section is to prove the following refinement of this result.

Theorem 5.1. Every signature σ of X factors through a closed point x of X.

We shall prove this first in the special case that k is a real closed field R. If a closed point x of X is complex then $\kappa(x) = R(\sqrt{-1})$ has no signatures, but if x is real then $W(\kappa(x) = W(R) \cong \mathbb{Z}$ and $\kappa(x)$ has a unique signature. Thus every real point $P \in X(R)$ yields one signature

$$\tau_{X,P} \colon W(X) \to W(\kappa(P)) \xrightarrow{\sim} \mathbb{Z}$$

of X. The claim of Theorem 5.1 is that the $\tau_{X,P}$ are all the signatures of X.

Every signature of X factors through some point of X and hence a fortiori factors through the reduced scheme X_{red} . Since the signatures correspond uniquely with the minimal prime ideals of W(X) this implies that the kernel of the natural map from W(X) to $W(X_{red})$ consists of nilpotent elements. Thus every signature of X factors through X_{red} in a unique way, and we may replace X by X_{red} in the whole study. Thus we assume since now that X is a reduced divisorial variety over R.

Lemma 5.2. If P and Q are path connectable points of X(R) then $\tau_{X,P} = \tau_{X,Q}$.

Proof. It suffices to prove this in the case that there exists an elementary non degenerate path γ starting at P and ending at Q. Let D be the Zariski closure of γ in X. Then $\tau_{X,P}$ has the factorization

$$\tau_{X,P} \colon W(X) \longrightarrow W(D) \xrightarrow{\operatorname{tp}} \mathbb{Z}$$

with $W(X) \to W(D)$ the natural map. The same holds true for $\tau_{X,Q}$. Thus it suffices to prove $\tau_{D,P} = \tau_{D,Q}$. Replacing X by D we assume since now that X is an irreducible curve.

Let $\pi: \tilde{X} \to X$ denote the normalization of X and let \tilde{P} , \tilde{Q} denote the preimages of P and Q in the normalization $\tilde{\gamma}$ of γ . We have the factorization

$$\tau_{X,P} \colon W(X) \xrightarrow{\pi^*} W(\tilde{X}) \xrightarrow{\tau_{\tilde{X}} \to \tilde{P}} \mathbb{Z}$$

of $\tau_{X,P}$ and an analogous factorization of $\tau_{X,Q}$. It suffices to prove $\tau_{\tilde{X},\tilde{P}} = \tau_{\tilde{X},\tilde{Q}}$. Thus we may replace X by \tilde{X} and assume that X is a smooth connected curve. But then the assertion is evident from the explicit description of W(X) in [10, §9].

Definition. Two real points P and Q of the variety X over R are called Witt equivalent in X if the signatures $\tau_{X,P}$ and $\tau_{X,Q}$ are equal. The classes of this equivalence relation on X(R) are called the Witt components of X(R).

By Lemma 5.2 above every Witt component is a union of path components. We know from §4 that there exist only finitely many path components in X(R). A fortiori X(R) consists of only finitely many Witt components B_1, \ldots, B_s . Every Witt component B_i corresponds uniquely with a signature τ_i of X, defined by $\tau_i = \tau_{X,P}$ with P an arbitrary point of B_i .

The claim of Theorem 5.1 in the case k=R is that τ_1, \ldots, τ_s are already all signatures of X. Suppose there exists a signature σ of X different from τ_1, \ldots, τ_s . We choose elements z_1, \ldots, z_s of W(X) such that $\sigma(z_i) \neq \tau_i(z_i)$ for $i=1, \ldots s$. Our goal is now to find a point P in X(R) such that $\sigma(z_i) = \tau_{X, P}(z_i)$ for $i=1, \ldots s$. This will be the desired contradiction.

There exists a point $y \in X$, not necessarily closed, such that σ factors through a signature $\tilde{\sigma}$ of $\kappa(y)$,



Let $Y = \overline{\{y\}}$ denote the irreducible closed subvariety of X with generic point y and $E := \kappa(y)$ denote the function field of Y over R. We choose a non empty open affine subset Z of Y which contains only regular points. The affine ring A = R[Z] of Z is a finitely generated algebra over R without zero divisors and with the quotient field E. For an element z of W(X) we denote by z | A the natural image of z in W(Z) = W(A) ("restriction" of z to A) and by z | E the image of z in W(E). We represent $z_i | E$ by a diagonal quadratic form $\langle f_{i1}, \ldots, f_{inv} \rangle$ with coefficients f_{ij} in A. Then

$$\sigma(z_i) = \tilde{\sigma}(z_i \mid E) = \sum_{j=1}^{n_i} \tilde{\sigma}(f_{ij}).$$

Here we use the common abbreviation $\tilde{\sigma}(f) := \tilde{\sigma}(\langle f \rangle) = \pm 1$. According to Artin and Lang there exists a place $\lambda: E \to R \cup \infty$ over R with the following properties:

a) λ is finite on A.

b) $\lambda(f_{ij}) \neq 0$, sign $\lambda(f_{ij}) = \tilde{\sigma}(f_{ij})$.

([1], [16, Theorem 8], [15, Prop. 6.4]; recall that $\tilde{\sigma}$ corresponds with an ordering of *E*.) The kernel of the homomorphism $\lambda | A$ from *A* to *R* is a maximal ideal m of *A* with A/m = R. Thus $\lambda | A$ is the evaluation homomorphism of *A* at a point $P \in Z(R)$, $\lambda(f) = f(P)$ for all *f* in *A*. We obtain (i = 1, ..., s)

$$\sigma(z_i) = \sum_{j=1}^{n_i} \operatorname{sign} f_{ij}(P).$$

We claim that $\tau_{X,P}$ has the same values on the z_i . Notice that $\mathcal{O}_{Y,P}$ is a regular local ring with quotient field E. The elements $z_i | \mathcal{O}_{Y,P}$ and $\langle f_{i1}, \ldots, f_{in_i} \rangle$ of $W(\mathcal{O}_{Y,P})$ have the same image in W(E). {Notice that the f_{ij} are units in $\mathcal{O}_{Y,P}$, hence $\langle f_{i1}, \ldots, f_{in_i} \rangle$ may indeed be regarded as an element of $W(\mathcal{O}_{Y,P})$.} By the specialization theorem [14, Cor. 2.3] the images of these elements of $W(\mathcal{O}_{Y,P})$ in $W(\kappa(P)) = \mathbb{Z}$ are also the same. This means that

$$\tau_{\boldsymbol{\chi},\boldsymbol{P}}(\boldsymbol{z}_i) = \sum_{j=1}^{n_i} \operatorname{sign} f_{ij}(\boldsymbol{P}).$$

Thus indeed $\tau_{X, P}(z_i) = \sigma(z_i)$ for i = 1, ..., s. This contradiction finishes the proof of Theorem 5.1 in the case k = R.

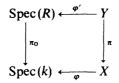
Remark. The proof was essentially the same as in [12, Chap. V] for the special case $R = \mathbb{R}$. At the end we have slightly simplified the argument avoiding here the theorem of Craven-Rosenberg-Ware cited in [12]. This simplification has been suggested to us by J.L. Colliot-Thélène.

As in the special case $R = \mathbb{R}$ [12, p. 249] we have the following natural

Question. Is every Witt component of X(R) already a path component?

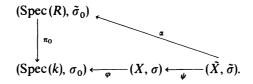
It is clear from [10, II §9] that the answer is "Yes" for X a smooth curve. G. Dietel has recently proved that the answer is also "Yes" for singular curves [6]. In the case $R = \mathbb{R}$ the answer is known to be "Yes" if X is a complete smooth surface or an abelian variety [4].

We now prove Theorem 5.1 for an arbitrary base field k. We apply the rudiments of the theory of real closures of schemes with respect to signatures in [11] and [13] and use the terminology developed there. Let σ_0 denote the signature of Spec(k) obtained from σ by restriction with respect to the structure morphism $\varphi: X \to \text{Spec}(k)$. Let R denote the real closure of k with respect to σ_0 in the classical sense and let $\tilde{\sigma}_0$ be the unique signature of R. We have a cartesian square

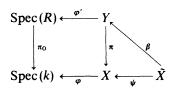


with $Y = X \bigotimes_k R$, π_0 the profinite covering induced by the inclusion map from k to R, and also π a profinite covering.

Let $\psi: (\tilde{X}, \tilde{\sigma}) \to (X, \sigma)$ be a real closure of the pair (X, σ) {"strict real closure" in the language of [11]}. According to the theory of real closures there exists a morphism α from $(\tilde{X}, \tilde{\sigma})$ to the pair (Spec(R), $\tilde{\sigma}_0$) such that the following diagram commutes:



We then obtain from α a morphism $\beta: \tilde{X} \to Y$ such that the diagram



commutes. We introduce the signature

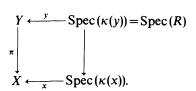
$$\tau := \tilde{\sigma} \circ \beta^* \colon W(Y) \to W(\tilde{X}) \to \mathbb{Z}$$

on Y. Then

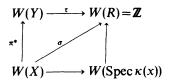
$$\tau \circ \pi^* = \tilde{\sigma} \circ \beta^* \circ \pi^* = \tilde{\sigma} \circ \psi^* = \sigma.$$

Thus we have extended the given signature σ on X to a signature τ on Y.

Now Y is of finite type and divisorial over R. Thus – as proved above – there exists a closed point y of Y such that τ factors through y. Let x denote the point $\pi(y)$ of X. Since $\kappa(x)$ is a subfield of R the point x is closed in X. We have a commutative diagram



Under the functor "Witt ring" this diagram turns into the horizontal and vertical part of the following diagram.



The square and the upper triangle both commute. Thus also the lower triangle commutes. This is the desired factorization of σ through the closed point x. Theorem 5.1 is proved.

References

- 1. Artin, E.: Über die Zerlegung definiter Funktionen in Quadrate, Abh. Math. Sem. Univ. Hamburg 5, 100-115 (1927)
- 2. Brumfiel, G.W.: Partially ordered rings and semi-algebraic geometry. London Math. Soc. Lecture Note Series 37, Cambridge U.P. 1979
- 3. Cohen, P.J.: Decesion procedures for real and *p*-adic fields. Comm. pure appl. math. 22, 131-151 (1969)
- 4. Colliot-Thélène, J.L., Sansuc, J.J.: Fibrés quadratiques et composantes connexes réelles. Math. Ann. 224, 105-134 (1979)

128

- 5. Coste-Roy, M.F., Coste, M.: Topologies for real algebraic geometry, in: A. Kock ed., Topos theoretic methods in geometry. Various Publications Series no. 30, Aarhus Universitet (1979)
- 6. Dietel, G.: Thesis on the Witt rings of curves over real closed fields, Regensburg, in preparation
- 7. Fehlner, J.: Analysis der Morphismen von Kurven über reell abgeschlossenen Körpern, Staatsexamensarbeit Regensburg 1978
- Geyer, W.D.: Reelle algebraische Funktionen mit vorgegebenen Null- und Polstellen. Manuscr. math. 22, 87-103 (1977)
- Grothendieck, A., Illusie, L.: Théorie des intersections et théorème de Riemann-Roch. Seminaire de Géometrie algébrique du Bois Marie 1966/67. Lecture Notes Math. 225. Berlin-Heidelberg-New York: Springer 1971
- Knebusch, M.: On algebraic curves over real closed fields I, II. Math. Z. 150, 49-70; 151, 189-205 (1976)
- 11. Knebusch, M.: Real closures of commutative rings I, II. J. reine angew. Math. 274/275, 61-89 (1975); 286/287, 278-313 (1976)
- 12. Knebusch, M.: Symmetric bilinear forms over algebraic varieties. In: Conference on quadratic forms (Kingston 1976), pp. 103-283. Queen's papers in Pure Appl. Math. 46 (1977)
- 13. Knebusch, M.: Real closures of algebraic varieties, ibid., pp. 548-568
- 14. Knebusch, M.: Spezialization of quadratic and symmetric bilinear forms, and a norm theorem. Acta arithm. 24, 279-299 (1973)
- Knebusch, M.: On the uniqueness of real closures and the existence of real places. Comment. math. Helv. 47, 260-269 (1972)
- 16. Lang, S.: The theory of real places. Ann. Math. 57, 378-391 (1953)
- 17. Witt, E.: Zerlegung reeller algebraischer Funktionen in Quadrate, Schiefkörper über reellem Funktionenkörper, J. reine angew. Math. 171, 4-11 (1934)

Received August 25, 1980